



저작자표시-동일조건변경허락 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



동일조건변경허락. 귀하가 이 저작물을 개작, 변형 또는 가공했을 경우에는, 이 저작물과 동일한 이용허락조건하에서만 배포할 수 있습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

August 2014
Master's Degree Thesis

Invariant Color Features
Detector and Descriptor using
Fast Explicit Diffusion in
Nonlinear Scale Spaces

Graduate School of Chosun University

Department of Computer Engineering

Adrianto Tedjokusumo

Invariant Color Features Detector and Descriptor using Fast Explicit Diffusion in Nonlinear Scale Spaces

비선형 비례 공간에서 명시적 고속 확산을 이용한
불변색상 특징 추출 및 기술

August 25, 2014

Graduate School of Chosun University

Department of Computer Engineering

Adrianto Tedjokusumo

Invariant Color Features Detector and Descriptor using Fast Explicit Diffusion in Nonlinear Scale Spaces

Advisor: Prof. Sang-Woong Lee

A thesis submitted in partial fulfillment of the
requirements for a Master's degree

May 2014

Graduate School of Chosun University

Department of Computer Engineering

Adrianto Tedjokusumo

테드조쿠수모 아드리안토의 석사학위논문을 인준함

위원장	조선대학교 교수	문인규	(인)
위 원	조선대학교 교수	양희덕	(인)
위 원	조선대학교 교수	이상웅	(인)

2014 년 5 월

조선대학교 대학원

TABLE OF CONTENT

TABLE OF CONTENT	i
LIST OF FIGURES	ii
LIST OF TABLES	v
ABSTRACT	vi
I. INTRODUCTION	1
A. Introduction to Features Detector and Descriptor	1
B. Problem Statement.....	3
C. Research Objectives	4
D. Thesis Layout	5
II. RELATED WORKS.....	6
A. Grayscale-based Features Detector and Descriptor Algorithm	6
B. Color-based Features Detector and Descriptor Algorithm	11
III. INVARIANT COLOR FEATURES DETECTOR AND DESCRIPTOR...	13
A. Structure of System	13
B. Feature Detection Algorithm	15
C. Feature Description Algorithm	24
IV. PERFORMANCE EVALUATION	26
A. Simulation Environment.....	26
B. Experimental Results and Performance Comparison	27
V. CONCLUSIONS.....	61
BIBLIOGRAPHY	62

LIST OF FIGURES

Figure 1. An example of three colors areas in R, G, and B and its grayscale conversion using luminance method.....	3
Figure 2. Two non-uniformly illuminated sample images in different point of view and their corresponding pixels.	4
Figure 3. The impact of color boosting algorithm which cause generalization of color in some regions of pixels.....	14
Figure 4. The conductivity coefficient g_1 in the Perona and Malik equation as a function of the parameter k . Notice that for increasing values of k only higher gradients are considered. This image considers as grayscale images of range 0-255 [7].....	17
Figure 5. Image detection results of the image datasets (first row) (a) blur; (b) JPEG compression; (c) Illumination; (d) Viewpoint; (e) Zoom + rotation. In row sequence from second row to down: ORB, BRISK, SIFT, SURF, KAZE, A-KAZE, HLS-AKAZE.....	32
Figure 6. Image matching results of the first row image datasets (a) blur; (b) JPEG compression; (c) Illumination; (d) Viewpoint; (e) Zoom + rotation. In row sequence from second row to down: ORB, BRISK, SIFT, SURF, KAZE, A-KAZE, HLS-AKAZE.....	36
Figure 7. Detector repeatability score for an overlap area error 60%. Best view in color. (a) blur; (b) JPEG compression; (c) Illumination; (d) Viewpoint; (e) Zoom + rotation.....	38
Figure 8. Recall vs 1-precision graphs for nearest neighbor matching strategy of image 1 vs image 4. Best view in color. (a) blur; (b) JPEG compression; (c) Illumination; (d) Viewpoint; (e) Zoom + rotation.....	42

Figure 9. Benchmark image matching performance of image 1 vs image 4. Best view in color. (a) Number of detected keypoint image 1; (b) Number of detected keypoint image 2; (c) Total Matches of correspondence images; (d) Inliers Ratio; (e) Feature detection and description time; (f) Matching descriptor time. 45

Figure 10. Two color images with their less-informative grayscale images. (a) wall_indoor images in color; (b) its grayscale images with the markers which show the less-informative region; (c) wall_street images in color; (d) its grayscale images with the markers which show the less-informative region. 48

Figure 11. Image detection results of the wall_indoor images. (a) A-KAZE method; (b) HLS-AKAZE Grad method; (c) HLS-AKAZE SD method. ... 49

Figure 12. Image detection results of the wall_street images. (a) A-KAZE method; (b) HLS-AKAZE Grad method; (c) HLS-AKAZE SD method. ... 50

Figure 13. Image matching results of the wall_indoor images. (a) A-KAZE method; (b) HLS-AKAZE Grad method; (c) HLS-AKAZE SD method. ... 51

Figure 14. Image matching results of the wall_street images. (a) A-KAZE method; (b) HLS-AKAZE Grad method; (c) HLS-AKAZE SD method. ... 52

Figure 15. Image detection results of the less-informative grayscale image datasets (a) wall_indoor; (b) wall_street using other grayscale-based feature detector methods. In row sequence from top to down: ORB, BRISK, SIFT, SURF, KAZE. 54

Figure 16. Image detection results of the less-informative grayscale image datasets (a) wall_indoor; (b) wall_street using other grayscale-based feature detector methods. In row sequence from top to down: ORB, BRISK, SIFT, SURF, KAZE. 55

Figure 17. (a) Detector repeatability score for an overlap area of 50%; (b) Recall vs 1-precision graphs for wall_indoor images; (c) Recall vs 1-precision graphs for wall_street images..... 56

Figure 18. Benchmark image matching performance of both less-informative grayscale images. Best view in color. (a) Number of detected keypoint image 1; (b) Number of detected keypoint image 2; (c) Total Matches of correspondence images; (d) Inliers Ratio; (e) Feature detection and description time; (f) Matching descriptor time. 60

LIST OF TABLES

Table 1. Simulation parameters	27
--------------------------------------	----

ABSTRACT

Invariant Color Features Detector and Descriptor using Fast Explicit Diffusion in Nonlinear Scale Spaces

Adrianto Tedjokusumo

Advisor: Prof. Sang-Woong Lee

Department of Computer Engineering

Graduate School of Chosun University

In recent years, several works have aimed to find the most effective and accurate feature extraction method. Considering the computational time and simplicity, most of state-of-the-art methods operate on grayscale images by applying color conversion mechanism. In the other hand, conversion to grayscale images causes some important information lost which potentially reduce the performance. In this paper, the authors propose a multi-scale 2D invariant color detection and description algorithm in nonlinear scale spaces. The algorithm exploits and utilizes color information of image in Hue, Lightness, and Saturation (HLS) space. Therefore, nonlinear scale spaces are built separately for each color channel (HLS) and adaptive integrated determinant Hessian responses are calculated in finding keypoints of the image.

We implement Fast Explicit Diffusion (FED) scheme in the nonlinear scale spaces which causes locally adaptive blurring to the image data. It can reduce image noise but retain object boundaries in obtaining superior localization accuracy and distinctiveness, where linear scale spaces system such as Gaussian blurring smooth

image to the same degree both details and noise. In addition, the authors propose a Color Modified-Local Difference Binary (CM-LDB) descriptor which exploits HLS color information concurrently with its gradient information from nonlinear scale spaces. Besides the improvements of precision and accuracy in general image matching, our proposed system (HLS-AKAZE) shows significant performance improvements compared with other grayscale feature extraction methods in the certain image which has low / unclear information in its grayscale image. Moreover, this proposed system is mostly invariant in the rotated, scaled, and illumination-changed images.

I. INTRODUCTION

A. Introduction to Features Detector and Descriptor

In many computer vision applications, such as image panorama stitching, image recognition, especially 3D Visual SLAM (Simultaneous Localization and Mapping) [1], one of the major processes is image matching. Image matching of every captured image is essentially needed in building a map by joining or combining those images corresponding to image features. The main concept of image matching is finding the similar important features or keypoints of two analyzed images. Therefore, features detector and descriptor must be initially calculated.

Significant progress on these has been made in the past decade through the development of local invariant features. Those features allow an application to find local image structures which occur repeatedly and to encode them in a representation that is invariant to a range of image transformations, such as translation, rotation, scaling, and affine deformation. There are many proposed local invariant features which are categorized as state-of-the-art methods; ORB, BRISK, SIFT, SURF, KAZE, and A-KAZE.

The basic purpose of local invariant features is to provide a representation that allows to efficiently match local structures between images. A sparse set of local measurements that capture the essence of the underlying input image and encode their interesting structure should be obtained for that purpose. There are two important criteria to meet this goal. First, the feature extraction process should be repeatable and precise which means the same features extracted on two images

showing the same object. Second, the features should be distinctive, namely different image structures can be told apart from each other.

The standard of feature extraction pipeline is organized as follows. First, calculate and find a set of distinctive keypoints. Second, define a region around each keypoint in a scale- or affine-invariant manner. Next, extract and normalized the region content. Fourth, compute a descriptor from the normalized region. Finally, match the local descriptors.

In finding a set of distinctive keypoints as the detectors, we should consider the robustness of viewpoint changes, presence of noise, and localization under varying image conditions. It can be determined by finding the local extrema points on a particular subset of points or regions. By exhibiting strong first or second order derivatives of the image such as Hessian and Harris detector, local extrema can be calculated.

Besides, the detector can be reliably extracted under scale changes, namely scale-invariant by computing the image derivatives at many scales. There are two scale spaces systems used in multi-scale feature extraction; linear and nonlinear scale spaces. Further information regarding these is explained in the following section. After a scale-invariant region has been detected, its content needs to be normalized for rotation invariance. This typically done by finding the region's dominant orientation followed by rotating the region content according to this angle in order to bring the region into canonical orientation.

Once a set of interest regions has been extracted from an image, their content needs to be encoded in a descriptor that is suitable for discriminative matching. Pixel intensity, image gradient, and image orientation mainly used as the descriptor value.

B. Problem Statement

Almost all features extraction methods including A-KAZE as the latest one only use the intensity information from image. The previous methods' authors use the grayscale image to simplify the algorithm and reduce computational requirements. However, converting HLS image containing three channels to grayscale image in one channel of 8-bit representation has a number of side-effects in certain condition which cause lack of precision. The gray-value version of color image does not preserve chromatic saliency which only defines the luminance values, not the hue and saturation values. The grayscale values can be calculated in many methods [2]; Intensity, Gleam, Luminance, Luma, Lightness, etc. which are the generalization of RGB channels by forming a weighted sum of the R, G, and B components. Therefore, there are some certain cases where the RGB color values of distinct objects are different, but in the intensity value seems homogenous. Figure 1 shows the simple example of three colors areas in R, G, and B and its grayscale image conversion using luminance method by MATLAB (*rgb2gray* function). The same result is also shown using OPENCV *cvtColor* function. Very prominent difference in colors is represented in the same grayscale values of the entire image. That phenomenon potentially causes many important keypoints in terms of color cannot be detected which reduces matching result performance.

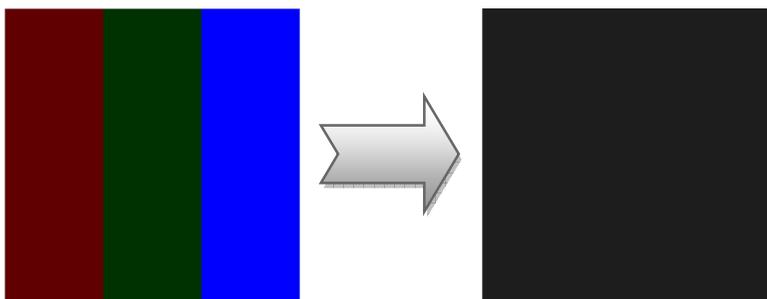


Figure 1. An example of three colors areas in R, G, and B and its grayscale conversion using luminance method.

Non-uniform illumination in the certain object of the observed image is also considered as problem in feature detection and descriptor algorithm. Figure 2 shows two captured images of the same object in different point of view which is not-uniformly illuminated. The pointed pixel position of one image is supposed to match the corresponding pixel on another image. The grayscale values of both pixels shows a significant difference where one of them is high in intensity and otherwise (concurrently with the neighborhood pixels around them). Therefore, most of grayscale feature extraction methods either eliminate this kind of keypoint or conduct image matching error where its descriptor will have different value on another image. In terms of color, it can be categorized in the same hue value and processed in better analysis.



Figure 2. Two non-uniformly illuminated sample images in different point of view and their corresponding pixels.

C. Research Objectives

Our main objective in this thesis is improving the matching performance of two corresponding images in the varying conditions, especially in the images containing different-color-same-grayscale regions. This goal can be obtained by combining impact of chromatic with the luminance information at both the

detection and description levels. To exploit the color information of images, it requires to extend multi-scale feature extraction theory from scalars to vectors, which means from luminance to HLS color. In finding the local extrema in determining selected keypoints, adaptive integrated determinant Hessian response is implemented. This system is using calculated weight of each channel Hessian response based on local standard deviation and local gradient information. The weighted values determine the importance of each color channel. Nonlinear scale spaces system is selected as our multi-scale feature extraction method of each color space channel which adopt adaptive conductivity diffusion algorithm in each scale to reduce image noise but enhanced or kept important edges using Fast Explicit Diffusion (FED) schemes. By means of FED schemes, nonlinear scale spaces can be built much faster than with any other kind of discretization scheme which also more accurate and easier to implement than others. In the descriptor level, Color Modified-Local Difference Binary (CM-LDB) descriptor is introduced as our scale-, rotation-, and illumination-invariant descriptor which exploits HLS color information concurrently with its gradient information from nonlinear scale spaces.

D. Thesis Layout

The rest of the thesis is organized as follows. First, in Chapter II, we give a brief explanation about the existing features detector and descriptor based on grayscale value including linear and nonlinear scale spaces, followed by the existing features extraction based on color information. In Chapter III, we discuss our proposed features detector and descriptor algorithm, which combining color features extraction concept and nonlinear scale spaces system. In chapter IV, benchmark performance of our system is evaluated and discussed. Comparison in performance with other existing methods is also presented in this chapter. Finally, the conclusions of the thesis are given in section V.

II. RELATED WORKS

A. Grayscale-based Features Detector and Descriptor Algorithm

Grayscale-based feature extraction method only utilizes luminance pixel values of the observed image. Grayscale information is believed can simplify the algorithm and reduce the computational complexity. In this section, we can classify this algorithm based on the scale spaces system; grayscale-based feature extraction in linear and nonlinear scale spaces. SIFT, SURF, BRISK, and ORB are built in linear scale spaces. Otherwise, KAZE and A-KAZE are using nonlinear scale spaces. All of these features extraction methods will be explained briefly in this section.

1. SIFT

Scale-invariant feature transform (SIFT) is a multi-scale feature extraction in linear scale spaces which was published by David Lowe in 1999. Lowe's SIFT approach is widely accepted as one of the highest quality options currently available with its promising distinctiveness. However, SIFT has a high expense on computational cost. SIFT approach takes an image and transforms in into a "large collection of local features vectors" [3]. Each of these features vectors is invariant to any scaling, rotation, and translation of the image. To aid the extraction of these features, the SIFT algorithm applies a four main stages filtering approach which generally become the base of other methods; scale spaces extrema detection, keypoint localization, orientation assignment, and keypoint descriptor.

In scale spaces extrema detection, the image is convolved with Gaussian filters at different scales and the Difference-of-Gaussian (DoG) is calculated using the image in Gaussian pyramid as the approximation of Laplacian-of-Gaussian (LoG). LoG itself acts as a blob detector which detects blobs in various sizes due to

change in scale, but this method has a high time consumption cost. Afterwards, we find local extrema points over scale and space on local neighborhood pixels and they will be determined as selected keypoints.

Keypoint localization is used to eliminate some unstable keypoints. This stage performs a detailed fit using interpolation of the nearby data for accurate location, scale, and ratio of principal curvatures based on the quadratic Taylor expansion of DoG. In addition, this algorithm will reject the low contrast keypoints which are sensitive to noise and eliminate poorly localized keypoints along an edge.

Orientation assignment is the stage to assign one or more orientations to each keypoint based on local image gradient directions. By this step, SIFT descriptor can be represented relative to this orientation which achieves the invariance of rotation. The gradient magnitude and orientation are precomputed using pixel differences of the Gaussian-smoothed image in the corresponding scale. Orientation histogram bins are created for each keypoint. Each sample in the neighboring window added to a histogram bin is weighted by its gradient magnitude and by a Gaussian weighted circular window.

Keypoint descriptor typically uses a set of orientation histograms of 4x4 pixel neighborhoods of the corresponding keypoint with 8 bins each. The descriptor becomes a vector of all the values of these histograms which totally has 128 elements.

2. SURF

Speeded Up Robust Features (SURF) [4] is a robust multi-scale local feature extraction which was published by Herbert Bay et al. in 2006. The SURF algorithm is based on the same principles and steps, but it utilizes in different scheme. SURF is claimed by its authors several times faster and more robust against different

image transformations than SIFT. SURF is showing a robust performance at handling images with blurring and rotation, but not superior in handling viewpoint and illumination change. Instead of SIFT method which approximate Laplacian of Gaussian with DoG for finding scale space, SURF approximate LoG with Box Filter. Convolution with Box Filter as the approximation of Haar wavelets can be easily calculated with the help of integral images which also can be done in parallel for different scales. SURF detector using basic Hessian matrix approximation where detects blob-like structures at locations where the determinant is maximum. Accurate localization of multi-scale SURF requires interpolation and non-maximum suppression in $3 \times 3 \times 3$ neighborhood.

For orientation assignment, SURF uses wavelet responses in horizontal and vertical direction within a circular neighborhood of certain radius around interest point. The dominant orientation is estimated by calculating the sum of all responses within a sliding orientation window of angle 60 degrees. SURF descriptor is using the same concept with orientation assignment by summing the wavelet responses in horizontal and vertical direction. In order to bring in information about the polarity of the intensity changes, the sum of the absolute values of the responses is also extracted. Therefore, each sub-region has a four-dimensional descriptor. Concatenate this for 4×4 sub-regions of corresponding keypoint, this results in a descriptor vector of length 64.

3. BRISK

Binary Robust Invariant Scalable Keypoints (BRISK) [5] is a multi-scale fast feature extraction algorithm which is comparable to SURF in the robustness of performance. BRISK adopt Features from Accelerated Segment Test (FAST) algorithm as feature detector and Binary Robust Independent Elementary Features (BRIEF) algorithm as feature descriptor. FAST is a corner detection algorithm to

find interest points in an image by investigate the intensity of contiguous pixels surround the corresponding observed pixel in a circle pattern. Briefly, the pixel which is majorly brighter or darker than its contiguous pixels is selected as an interest point. The location and scale of each keypoint are obtained in the continuous domain via quadratic function fitting.

BRISK represents its descriptors as a binary string by concatenating the results of simple brightness comparison tests. The idea is employ BRIEF algorithm in a far more qualitative manner. It also identifies the characteristic direction of each keypoint to achieve rotation invariance. This descriptor makes use of a circle concentric pattern which resembles DAISY descriptor, used for sampling the neighborhood of the keypoint. The basic idea of each bit value of the BRIEF concept is the comparison of the intensity of selected sampling pattern. BRISK contains only a bit-string of length 512, which improve the matching performance in time compared to SIFT and SURF.

4. ORB

ORB [6] feature extraction algorithm is the combination of Oriented Fast and Rotated BRIEF, which is almost the same concept with BRISK. The contribution of this method is finding an addition of a fast and accurate orientation component to FAST and increase the efficiency computation of oriented BRIEF features. ORB is a very fast multi-scale feature extraction which is also comparable in performance with SURF and also rotation invariant and resistant to noise. In the detection, the addition of orientation in FAST is achieved by finding the intensity centroid. Intensity centroid assumes that a corner's intensity is offset from its center, and this vector maybe used to impute an orientation. In the descriptor, the proposed Rotation-Aware BRIEF is an efficient method to steer BRIEF according to the orientation of keypoints. This descriptor uses the patch orientation and the

calculation of the corresponding rotation matrix from the feature set location of sample binary tests.

5. KAZE

KAZE [7] is a multi-scale two-dimensional feature detection and description algorithm in nonlinear scale spaces published by Pablo Fernandes Alcantarilla et al. in 2012. KAZE is more expensive to compute than SURF due to the construction of the nonlinear scale spaces, but shows a step forward performance both in detection and description against previous state-of-the-art methods. Previous approaches detect and describe features at different scale levels by building or approximating the Gaussian scale space of an image. In the authors perspective, Gaussian blurring does not respect the natural boundaries of objects and smoothes to the same degree both details and noise, reducing localization accuracy and distinctiveness.

In contrast, KAZE detect and describe 2D features in a nonlinear scale space by means of nonlinear diffusion filtering. In this way, KAZE can make blurring locally adaptive to the image data, reducing noise but retaining object boundaries, obtaining superior localization accuracy and distinctiveness. The nonlinear scale space is built using efficient Additive Operator Splitting (AOS) techniques and variable conductance diffusion. In the other hand, KAZE descriptor adopts SURF descriptor principal in finding the dominant orientation and also use 64 descriptor vector contains the derivatives responses in horizontal and vertical direction.

6. A-KAZE

Accelerated-KAZE (A-KAZE) [8] is the speed-up version of KAZE published by the same authors. This feature extraction wants to obtain low-computationally

demanding features taking advantage of the benefits of nonlinear diffusion filtering. For this purpose, A-KAZE introduces a recent mathematical framework called Fast Explicit Diffusion (FED) to feature detection and description problems. By means of FED schemes, nonlinear scale spaces can be built much faster than with any other kind of discretization scheme. Furthermore, FED schemes are extremely easy to implement and are more accurate than KAZE's AOS schemes. A-KAZE embeds FED schemes in a pyramidal framework with a fine to coarse strategy to speed-up dramatically feature detection in nonlinear scale spaces.

To preserve low computational demand and storage requirement, A-KAZE also introduces a highly efficient Modified-Local Difference Binary (M-LDB) descriptor. While the original LDB descriptor presented in [9] is neither rotation nor scale invariant as BRIEF is, the authors build a robust binary descriptor that is rotation and scale invariant and exploits gradient information from the nonlinear scale space, increasing distinctiveness. The authors claim that A-KAZE features are faster to compute than SURF, SIFT and KAZE and also exhibit much better performance in detection and description than previous methods, including ORB and BRISK.

B. Color-based Features Detector and Descriptor Algorithm

Grayscale-based feature extraction methods are focusing on the shape saliency of the local neighborhood. These methods are luminance based which has disadvantage that the distinctiveness of the local color information is completely ignored in determining salient image features. In [10], color-based feature detector exploits the possibilities of salient point detection in color images by filtering the image using color saliency boosting algorithm. The proposed idea of this color saliency boosting function is to achieve uniformly distributed color derivatives distribution which is usually dominated by luminance changes. Another way to

exploit the color information of images is to extend multi-scale feature extraction theory from scalars to vectors. This means the extension from luminance to color is an extension from scalar to vector signals. The basic approach to achieve that is by computing the derivatives of each channel separately and then combining the partial results. The extension that is proposed by the authors is incorporate color saliency in Laplacian-of-Gaussian, Harris-Laplace, and Hessian detector.

In the descriptor, the authors use the Color Attention method which means the descriptor framework is separately processed color and shape before they are combined. A class-specific color attention map is constructed. This map is used to modulate the sampling of shape features; in regions with high attention more shape features are sampled than in regions with low attention. The final representation is obtained by concatenating all the class-specific histograms.

III. INVARIANT COLOR FEATURES DETECTOR AND DESCRIPTOR

A. Structure of System

In this thesis, the authors propose an invariant color features detector and descriptor which is adopting A-KAZE system approach. Recently [10], it has been shown that the combined impact of color at both the detection and description levels for object recognition can be successfully used to improve the performance of image classification results. In this system, it is using a color saliency boosting algorithm as the initial step of feature detection to expose the saliency of color edges. This system is built in linear scale spaces which adopts Gaussian blurring that smooth image to the same degree both details and noise. It means the significance of the initial filtering only improve the color details in the real scale image. The impact of filtering will not be seen in the other scale of image pyramid, causing the system as not a robust scale invariant detector.

We also verify this method by using nonlinear scale spaces system. The accuracy and precision of matching performance is low even in the colorful images. The color boosting algorithm has an impact which removes some important information and causes the generalization of color in some regions of pixels, as we can see in Figure 3. In the other hand, we adopt their concept to extend multi-scale feature extraction theory from scalars to vectors which means the extension from luminance to color. Instead of extend to RGB channels, our approach uses HLS color space to extend the luminance of A-KAZE detector because hue and saturation value is important to be additional information which are robust of the illumination change.



Figure 3. The impact of color boosting algorithm which cause generalization of color in some regions of pixels.

Each color channel (HLS) is processed individually in the nonlinear scale spaces. Nonlinear scale spaces are the diffusion filtering process which describes the evolution of each intensity color channel of an image through increasing scale levels as the divergence of a certain flow function that control the diffusion process. Nonlinear scale spaces are built using variable conductance diffusion and efficient Fast Explicit Diffusion (FED) techniques. Conductivity function in diffusion equation causes the diffusion adaptive to local image structure. It means it can reduce image noise but enhanced or kept important edges. Since there is no analytical solutions for nonlinear diffusion equation, discretizing the equation can be the approximated solution. One solution is using FED scheme which combines the advantages of explicit and semi-implicit schemes. By means of FED schemes, nonlinear scale spaces can be built much faster than with any other kind of discretization scheme which also more accurate and easier to implement than Adaptive Operator Splitting (AOS) introduced by KAZE. This system embeds FED schemes in a pyramidal framework with a fine to coarse strategy.

In determining the keypoints of an image in the nonlinear scale spaces to be the selected feature detectors, we need to find extrema points on local neighborhood pixels by using the adaptive integrated determinant of Hessian response value. This

algorithm computes the derivatives of each channel separately and combines the partial results as the determinant of Hessian response. In combining the response, we proposed an adaptive weight of response calculation in two ways; Standard Deviation Weight Response and Local Gradient Weight Response.

In feature descriptor, the authors introduced a Color Modified-Local Difference Binary (CM-LDB) descriptor which exploits HLS color information concurrently with its gradient information from nonlinear scale spaces. Based on the previous M-LDB concept from A-KAZE, we extend the luminance-based descriptor into color-based descriptor using the same concept previously described. Based on nonlinear scale spaces, CM-LDB descriptor is scale invariant. This system is also robust to rotation variant achieved by adopting the Orientation SURF algorithm. Besides, the color gradient information extracted from the calculated feature detection method is also invariant to the change of illumination.

B. Feature Detection Algorithm

1. Nonlinear Diffusion Filtering

Nonlinear diffusion approaches describe the evolution of the luminance of an image through increasing scale levels as the divergence of a certain flow function that controls the diffusion process. These approaches are normally described by nonlinear partial differential equations (PDEs), due to the nonlinear nature of the involved differential equations that diffuse the luminance of the image through the nonlinear scale space. Equation 1 shows the classic nonlinear diffusion formulation:

$$\frac{\partial L}{\partial t} = \text{div} (c(x, y, t) \cdot \nabla L) \quad (1)$$

where div and ∇ are respectively the divergence and gradient operators. Conductivity function (c) in the diffusion equation makes the diffusion adaptive to the local image structure. The function c depends on the local image differential structure, and this function can be either a scalar or a tensor. The time t is the scale parameter, and larger values lead to simpler image representations. In this thesis, we will focus on the case of variable conductance diffusion, where the image gradient magnitude controls the diffusion at each scale level.

a. Perona and Malik Diffusion Equation

Nonlinear diffusion filtering was introduced in the computer vision literature in [11]. Perona and Malik proposed to make the function c dependent on the gradient magnitude in order to reduce the diffusion at the location of edges, encouraging smoothing within a region instead of smoothing across boundaries. In this way, the function c is defined as:

$$c(x, y, t) = g(|\nabla L_\sigma(x, y, t)|) \quad (2)$$

where the luminance function ∇L_σ is the gradient of a Gaussian smoothed version of the original image L . Perona and Malik described two different formulations for the conductivity function g :

$$g_1 = \exp\left(-\frac{|\nabla L_\sigma|^2}{k^2}\right) \quad , \quad g_2 = \frac{1}{1 + \frac{|\nabla L_\sigma|^2}{k^2}} \quad (3)$$

where the parameter k is the contrast factor that controls the level of diffusion. The function g_1 promotes high-contrast edges, whereas g_2 promotes wide regions over smaller ones. Weickert [12] proposed a slightly different diffusion function for rapidly decreasing diffusivities, where smoothing on both sides of an edge is much stronger than smoothing across it. That selective smoothing prefers intraregional smoothing to interregional blurring. This function, which we denote here as g_3 , is defined as follows:

$$g_3 = \begin{cases} 1 & , |\nabla L_\sigma|^2 = 0 \\ 1 - \exp\left(-\frac{3.315}{(|\nabla L_\sigma|/k)^8}\right) & , |\nabla L_\sigma|^2 > 0 \end{cases} \quad (4)$$

The contrast parameter k can be either fixed by hand or automatically by means of some estimation of the image gradient. The contrast factor determines which edges have to be enhanced and which have to be canceled. In this paper we take an empirical value for k as the 70% percentile of the gradient histogram of a smoothed version of the original image. This empirical procedure gives in general good results in our experiments. However, it is possible that for some images a more detailed analysis of the contrast parameter can give better results. Figure 4 depicts the conductivity coefficient g_1 in the Perona and Malik equation for different values of the parameter k . In general, for higher k values only larger gradients are taken into account.

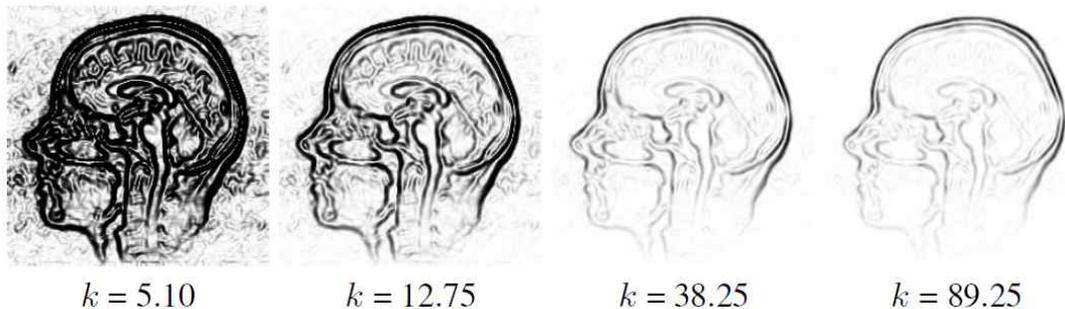


Figure 4. The conductivity coefficient g_1 in the Perona and Malik equation as a function of the parameter k . Notice that for increasing values of k only higher gradients are considered. This image considers as grayscale images of range 0-255 [7].

b. Fast Explicit Diffusion (FED)

FED combines the advantages of explicit and semi-implicit schemes while avoiding their shortcomings. FED schemes are motivated from a decomposition of box filters in terms of explicit schemes [13]. Iterated box filters approximate

Gaussian kernels with good quality and are easy to implement. The main idea is to perform M cycles of n explicit diffusion steps with varying step sizes τ_j that originate from the factorization of the box filter:

$$\tau_j = \frac{\tau_{max}}{2 \cos^2(\pi \frac{2j+1}{4n+2})} \quad (5)$$

where τ_{max} is the maximal step size that does not violate the stability condition of the explicit scheme. The corresponding stopping time θ_n of one FED cycle is obtained as:

$$\theta_n = \sum_{j=0}^n \tau_j = \tau_{max} \frac{n^2+n}{3} \quad (6)$$

Some of the step sizes τ_j from Eq. 10 may violate stability conditions. However, due to the similarities between FED and box filtering (always stable), we obtain also a stable scheme at the end of a FED cycle. The discretization of Eq. 6 using an explicit scheme can be expressed in vector-matrix notation as:

$$\frac{L^{i+1}-L^i}{\tau} = A(L^i)L^i \quad (7)$$

where $A(L^i)$ is a matrix that encodes the conductivities for the image and τ is a constant time step size such that $\tau_j < \tau_{max}$ in order to respect stability conditions.. In the explicit scheme, the solution L^{i+1} is computed in a direct way from the solution at the previous evolution level L^i and image conductivities $A(L^i)$:

$$L^{i+1} = (I + \tau_j A(L^i))L^{i+1,j}, \quad j = 0, \dots, n - 1 \quad (8)$$

where I is the identity matrix. Considering the a priori estimate $L^{i+1,0} = L^i$, a FED cycle with n variable step sizes τ_j is obtained as:

$$L^{i+1} = (I + \tau_j A(L^i))L^{i+1,j}, \quad j = 0, \dots, n - 1 \quad (9)$$

It is important to note here that the nonlinearities from the matrix $A(L^i)$ are kept constant during the whole FED cycle. Once a FED cycle is done, we compute the new values of the matrix $A(L^i)$.

2. Nonlinear Scale Spaces with Fast Explicit Diffusion

Our system use FED scheme to speed up the construction of nonlinear scale spaces considering anisotropic diffusion. We embed the FED scheme into a fine to coarse pyramidal framework. At first, we need to define a set of evolution times from which we can build the nonlinear scale spaces. The scale spaces are discretized in a series of O octaves and S sublevels. The octave and sublevel indexes are mapped to their corresponding scale σ (pixels) through the following formula:

$$\sigma_i(o, s) = 2^{o+s/S}, o \in [0 \dots O - 1], s \in [0 \dots S - 1], i \in [0 \dots M] \quad (10)$$

M is the total number of filtered images which equals to the total of evolution. The total number of filtered image can be concluded as $M = O \times S$. Nonlinear diffusion filtering operates in time units. Therefore, we need to convert the set of discrete in pixel units σ_i to time units. The mapping formula is adopting Gaussian scale space concept, where the convolution of an image with a Gaussian of standard deviation σ (in pixels) is equivalent to filtering the image for time $t = \sigma^2/2$. Therefore, we apply this conversion in order to obtain a set of evolution times and transform the scale spaces $\sigma_i(o, s)$ to time units using the following mapping formula:

$$t_i = \frac{1}{2} \sigma_i^2, i = \{0 \dots M\} \quad (11)$$

In FED schemes, there are inner and outer FED cycles. There are $M-1$ outer cycles and for each cycle we computer the minimum number of inner steps n . In the case of 2D images, the maximal step size which does not violate the stability conditions is $\tau_{max} = 0.25$, considering a grid size of 1 pixel for image derivatives.

In each outer cycles, at first we compute the diffusivity matrix $A(L^i)$ using the selected conductivity function. In our experiment, we use conductivity function g_2 which promotes wide area regions which are more suitable for blob-like features such as the detected result by determinant of Hessian. Second, we compute FED outer cycle which covers a cycle time $T = t_{i-1} - t_i$. Then, based on that cycle time, we compute the number of FED inner steps n which is also the explicit diffusion steps. It can be calculated by the following formula:

$$n = \sqrt{\frac{3T}{\tau_{max}} + 0.25} \quad (12)$$

Afterwards, we compute step sizes τ_j for each inner steps using Eq. 5. The FED cycle time θ_n as defined in Eq.6 covers only a discrete set of values. In order to allow arbitrary cycle times T , we need to calculate the minimum cycle length n with $\theta_n \geq T$ and then multiply the time steps τ_j by scale factor $q = T / \theta_n$. By set the prior estimate $L^{i+1,0} = L^i$, we calculate the next evolution level image of FED cycle using Eq. 9. Once we reach the last sublevel in each octave, we downsample the image by a factor of 2 using certain smoothing mask and use it as the starting image for the next FED cycle in the next octave. We also need to modify the contrast parameter k , it needs to be multiplied by 0.75 where the smoothing mask reduces the contrast of ideal step edge by 25%.

3. Adaptive Integrated Determinant of Hessian Response

After build the nonlinear scale spaces, in finding the keypoints we need to compute the determinant Hessian response for each of the filtered images L^i . The set of differential multiscale operators are normalized with respect to scale, using a normalized scale factor that takes into account the octave of each particular image in the nonlinear scale space, $\sigma_{i,norm} = \sigma_i / (2^o)$, and

$$L_{Hessian}^i = \sigma_{i,norm}^2 (L_{xx}^i L_{yy}^i - L_{xy}^i L_{xy}^i) \quad (13)$$

Because our system use three channels which are built individually in nonlinear scale spaces, the determinant Hessian response need to be the integration of all channels. It can be achieved by calculating the summation of the weighted determinant Hessian response of each channel using the following formula:

$$L_{Total}^i = w_H(x, y)L_{HessianH}^i + w_L(x, y)L_{HessianL}^i + w_S(x, y)L_{HessianS}^i \quad (14)$$

The weighted value of each channel response needs to be adaptive to each pixel. Briefly, each pixel of the filtered images L^i has its own weighted value. The weighted value can be calculated using two algorithms; Standard Deviation Weight Response and Local Gradient Weight Response. Total of three channels weight of each pixel is equal to 1.

In computing second order derivatives in the determinant Hessian response, we use concatenated Schaar filters with step size $\sigma_{i,norm}$. Schaar filter is used because it approximates rotation invariance better than other filter or central differences differentiation. Afterwards, we search for maxima of the detector response in a window of 3x3 pixels at each evolution level i . The response is also check that the response is higher than pre-defined threshold. And for each potential maxima, we check the response respect to other keypoints at level $i+1$ and $i-1$ in a window of size $\sigma_i \times \sigma_i$ pixels. Finally, the 2D position of the keypoint is estimated with sub-pixel accuracy by fitting a 2D quadratic function to the adaptive integrated determinant of Hessian response in 3x3x3 pixels neighborhood and finding its maximum.

a. Standard Deviation Weight Response

Standard deviation weight response is weight value of each pixel which is calculated in each evolution level to measure the importance of the calculated determinant Hessian response of each channel. The importance is shown by calculating standard deviation of the measured square kernel q^2 , where the kernel

size q is dependent with its corresponding sublevel. We compare each the standard deviation of each channel where higher standard deviation defines the higher of the corresponding channel's importance. Corresponding channel that has highest variation becomes the most influence response in the adaptive integrated determinant of Hessian response.

$$\bar{P}_{x,y} = \frac{\sum_{i=-d}^d \sum_{j=-d}^d P_{x+i,y+j}}{q^2} \quad (15)$$

$$\sigma(x,y) = \sqrt{\frac{\sum_{i=-d}^d \sum_{j=-d}^d (P_{x+i,y+j} + \bar{P}_{x,y})^2}{q^2 - 1}} \quad (16)$$

$$w_{SDTotal}(x,y) = \sigma_H(x,y) + \sigma_L(x,y) + \sigma_S(x,y) \quad (17)$$

$$w_{SD(H)}(x,y) = \frac{\sigma_H(x,y)}{w_{SDTotal}(x,y)} \quad (18)$$

$$w_{SD(L)}(x,y) = \frac{\sigma_L(x,y)}{w_{SDTotal}(x,y)} \quad (19)$$

$$w_{SD(S)}(x,y) = \frac{\sigma_S(x,y)}{w_{SDTotal}(x,y)} \quad (20)$$

Standard deviation of pixel (x,y) is calculated using Eq. 16 where the mean of measured kernel $\bar{P}(x,y)$ is calculated using Eq. 15, the kernel size in pixels q is calculated using formula $q = \sigma_i(o,s) / 2^o$, and $d = (q-1)/2$. The kernel size q should be represented in odd number. The standard deviation weight response of each channel is calculated accordingly in Eq. 18, 19, and 20. The calculated weight response is linear equation of its standard deviation value.

b. Local Gradient Weight Response

Local gradient weight response also determines the contributions of luminance and chromatic components in the calculation of determinant Hessian response as the

measured value in finding extrema on nonlinear scale spaces. This proposed algorithm is using pre-calculated function $|\nabla L_\sigma(x, y)|$ in the diffusivity equation.

$$|\nabla L_\sigma(x, y)| = \sqrt{L_x^2(x, y) + L_y^2(x, y)} \quad (21)$$

$$w_{GTotal}(x, y) = |\nabla L_{H\sigma}(x, y)| + |\nabla L_{L\sigma}(x, y)| + |\nabla L_{S\sigma}(x, y)| \quad (22)$$

$$w_{G(H)}(x, y) = \frac{|\nabla L_{H\sigma}(x, y)|}{w_{GTotal}(x, y)} \quad (23)$$

$$w_{G(L)}(x, y) = \frac{|\nabla L_{L\sigma}(x, y)|}{w_{GTotal}(x, y)} \quad (24)$$

$$w_{G(S)}(x, y) = \frac{|\nabla L_{S\sigma}(x, y)|}{w_{GTotal}(x, y)} \quad (25)$$

$$G_x = \begin{bmatrix} -3 & 0 & +3 \\ -10 & 0 & +10 \\ -3 & 0 & +3 \end{bmatrix} \quad (26)$$

$$G_y = \begin{bmatrix} -3 & +10 & -3 \\ 0 & 0 & 0 \\ +3 & +10 & +3 \end{bmatrix} \quad (27)$$

As shown in Eq. 21, $|\nabla L_\sigma(x, y)|$ is the gradient of a Gaussian smoothed version of the image L^i which is determined with the first derivatives of the image in x and y direction. The derivatives are calculated by convolving 3x3 Schaar operator G_x and G_y as shown in Eq. 26 and 27. Using the same concept with previous algorithm, the weight response of each channel (calculated in Eq. 23, 24, and 25) is the linear equation of its corresponding local gradient value. We compared each local gradient of each channel where the higher local gradient defines the higher of the corresponding channel's importance.

C. Feature Description Algorithm

1. CM-LDB Descriptor

We propose a Color Modified-Local Difference Binary (CM-LDB) that exploits gradient and intensity information from the nonlinear scale space. The LDB descriptor was introduced in [9] and follows the same principle as BRIEF, but using binary tests between the averages of areas instead of single pixels for additional robustness. Our CM-LDB descriptor is using the same concept with AKAZE M-LDB descriptor but in our application, we extend the luminance intensity to HLS color intensity. In addition to the each channel intensity values, the horizontal and vertical derivatives in the areas of each color channel are being compared. Therefore, we use 3 variables in 3 color channels which resulting in $3 \times 3 = 9$ bits per comparison.

Our descriptor proposes using various grids of finer steps, dividing the patch in 2×2 , 3×3 , and 4×4 grids. For each grid level, we compare each grid with others in same level one by one. The averages of those subdivisions are very fast to compute using integral images if the descriptor is upright (not rotation invariant) as in [8]. However, when considering the rotation of the keypoints integral images cannot be used, and visiting all points in a rotated subdivision can be relatively expensive in computation time. Rotation invariance is obtained by estimating the main orientation of the keypoint using the orientation concept of SURF, and the grid of CM-LDB rotated accordingly. Instead of using the average of all pixels inside each subdivision of the grid, we subsample the grids in steps that are a function of the scale s of the feature. This approximation of the average performs well in our experiments. The scale-dependent sampling in turn makes the descriptor robust to changes in scale. CM-LDB uses the derivatives computed in the feature detection step, reducing the number of operations required to construct the descriptor.

Given that CM-LDB computes an approximation of the average of the same areas in the intensity and gradient images, the Boolean values that result from the comparisons are not independent of each other. Reducing the size of the descriptor by choosing a random subset of the bits or with a more elaborated method is expected to improve the results, or at the very least reduce the computational load without decreasing performance.

Total comparison in one keypoint descriptor is $\binom{4}{2} + \binom{9}{2} + \binom{16}{2} = 162$ comparisons. Therefore, the total bits of every descriptor are $162 \times 9 = 1458$ bits. Our descriptor is 3 times bigger in memory consumption and 3 times slower in matching time than A-KAZE M-LDB descriptor. This trade-off need to be done to overcome the reduction of performance if eliminates certain channel or combines the channels' information into one joining form.

IV. PERFORMANCE EVALUATION

A. Simulation Environment

The simulation of our proposed idea is implemented in Visual C++ 2010. MATLAB R2012b is used in creating the figures of graph using the text file result from Visual C++. All simulations are produced using Intel Quad Core @2.83GHz CPU with 3GB RAM in 32-bit Windows operating system. The simulation parameters are summarized in Table 1. In the feature detection stage, we should select our minimum detector response threshold in finding nonlinear scale spaces maxima. The potential maxima based on the adaptive integrated determinant of Hessian response in 3x3x3 pixels neighborhood also need to be higher the defined response threshold. Therefore, this value determines the number of keypoints that is possibly selected. In the diffusion filtering, conductivity function g_2 is selected as our default one.

To measure our proposed algorithm performance, we perform image matching for two corresponding images. In matching the descriptors value of two corresponding images, we implement k -nearest neighbor (kNN) algorithm [14] in finding the closest k descriptors from the measured one. In calculating the distance, our system is using Brute Force [15] (for non-bit-based descriptors) and Brute Force Hamming (for bit-based descriptors). One of the measured performance variables is matching score / number of inliers ratio from the matched keypoints. Inliers ratio can be computed based on ground truth homography which is not always given in our datasets. Hence, homography matrix needs to be computed first using the RANSAC algorithm [16]. In measuring inliers ratio using RANSAC, we need to determine the maximum homography error value and Nearest Neighbor Distance Ratio (NNDR). In kNN descriptor matcher, we select $k=2$. Hence, we need to

check the difference in distance of two candidate matched descriptor. NNDR percentage is the threshold influencing the number of keypoint matches where if the difference of them is too close, it eliminates this descriptor as the matching candidates. In the other hand, maximum homography error determines maximum error in pixels to accept an inlier.

Detector response threshold	0.001 – 0.005 (Default : 0.001)
Octave level	4
Number of sublevels	4
Diffusivity type	g_2
Descriptor matcher	kNN match
Matching Type	Brute Force/ Brute Force Hamming
Compute inliers using RANSAC	Yes, if no ground truth homography
Maximum homography error(in pixels)	2.5
NNDR matching value ratio	80%

Table 1. Simulation parameters

B. Experimental Results and Performance Comparison

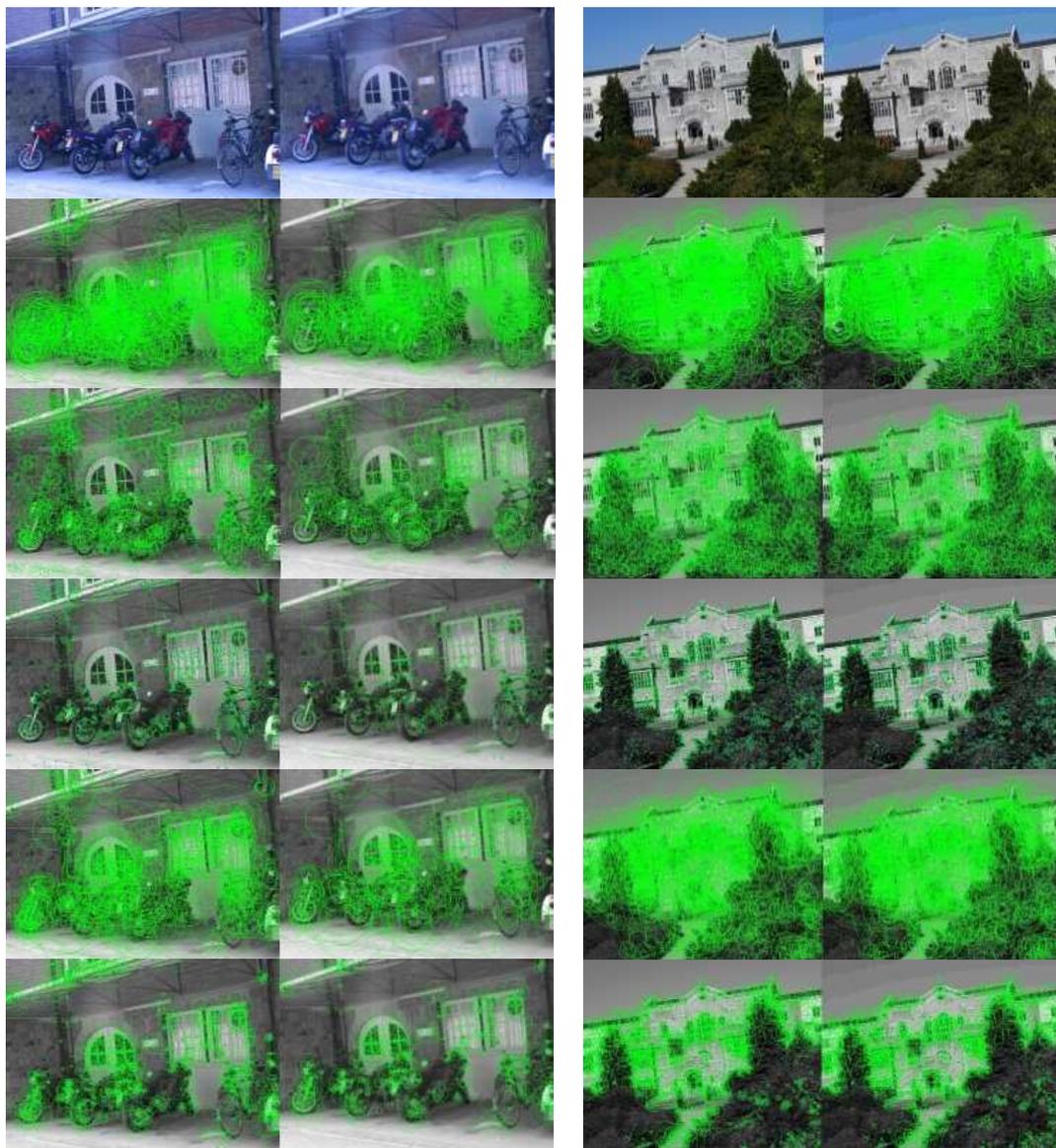
In this section, we present experimental results and performance comparison with other features extraction methods obtained on the standard evaluation set of Mikolajczyk et al. [17, 18] and on color images which has low information in their grayscale images. The standard dataset includes several image sets (each sequence contains 6 images) with different geometric and photometric transformation such as image blur, illumination-changes (lighting), viewpoint, zoom, rotation, and JPEG compression. In addition, the ground truth homographies are also available for every images transformation with respect to the first images of every sequence.

We compare our HLS-AKAZE features detector and descriptor against ORB, BRISK, SIFT, SURF, KAZE, and A-KAZE features. For ORB, BRISK, SIFT, and SURF, we use the OpenCV based implementation. And for KAZE and A-KAZE, we use the original open-source library from the author's website. For ORB, there are some parameters which we initially set. We set the maximum number of keypoints detect to 1500, the scale factor to 1.5, the number detection octaves to 3, the edge threshold to 31.0, the number of points producing each element of the oriented BRIEF descriptor to 2 points, and the patch size is 31. For BRISK, we initially set the FAST/AGAST detection threshold score to 10.0 and the number of detection octaves to 3. For SIFT, we set the maximum number of keypoints detect to 4000, the number of detection sublevel to 3, the contrast threshold to 0.04, the edge threshold to 10, and the sigma value to 1.6. For SURF, we set the determinant Hessian threshold to 1000, the number of detection octaves to 4, the number of detection sublevel to 3, the selected descriptor size is 128 bytes (extended SURF) and the rotation invariance of detector is rotation invariant (not upright).

For KAZE, A-KAZE, and HLS-AKAZE, we set most of the parameter in the same value which basically the three methods have the same system. In KAZE, the descriptor type selected is M-SURF 64bytes descriptor, the nonlinear scheme is using AOS, and the sigma smoothing derivatives is set to 1.0. In A-KAZE, the descriptor type used is M-LDB descriptor and the nonlinear scheme is FED. For three of them, the feature detection thresholds of different methods are set to proper values to detect approximately the same number of features per image. Our approach is using local gradient weight response as our default weighted method which is determined as the best based on experiment.

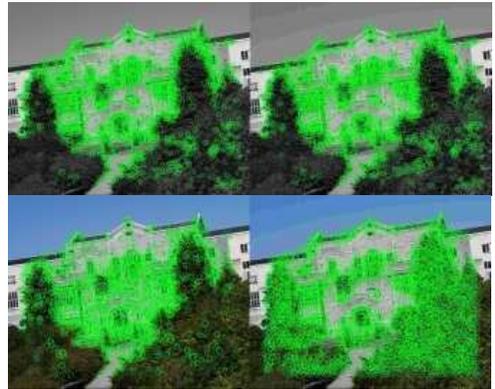
Figure 5 shows the original images followed by the features detection results for each features extraction methods in the standard datasets. From each dataset, we

choose the first image and the fourth image as our experiment data. For the next experiment, the same images will be used. From left to right in column sequence, it shows blur, JPEG compression, lighting, viewpoint, and zoom + rotation images. From top to down in row sequence, it shows ORB, BRISK, SIFT, SURF, KAZE, A-KAZE, and HLS-AKAZE image detection results. The image matching results are shown in Figure 6 which also have the same sequence with figure 5.

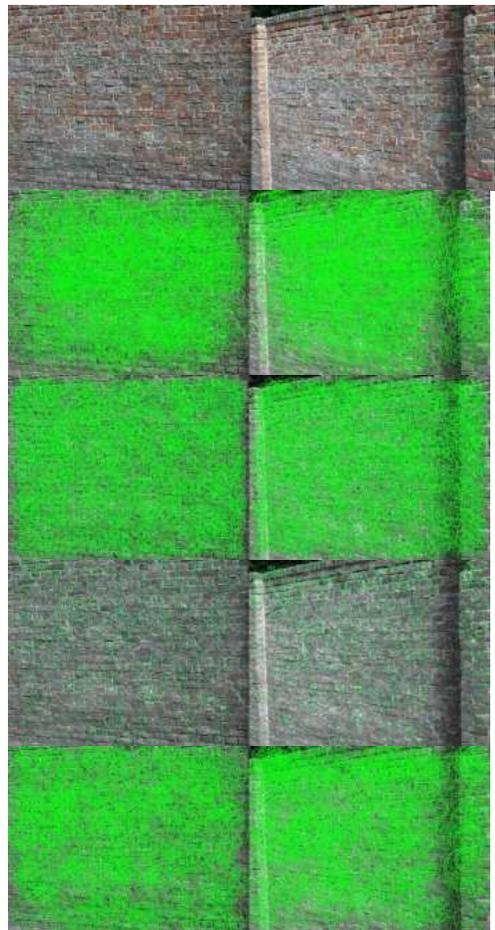




(a)



(b)

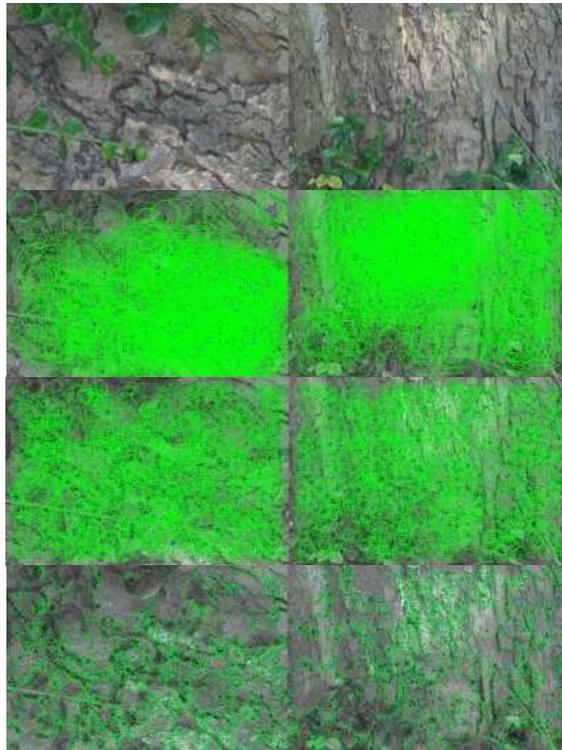


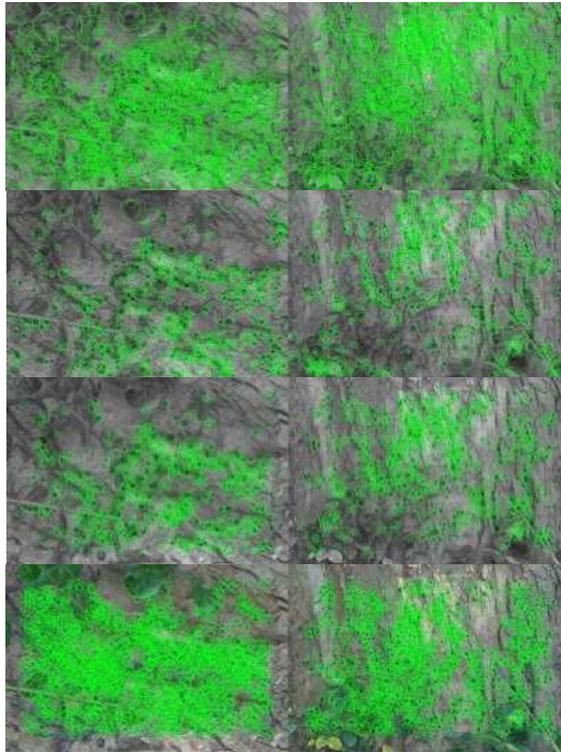


(c)



(d)

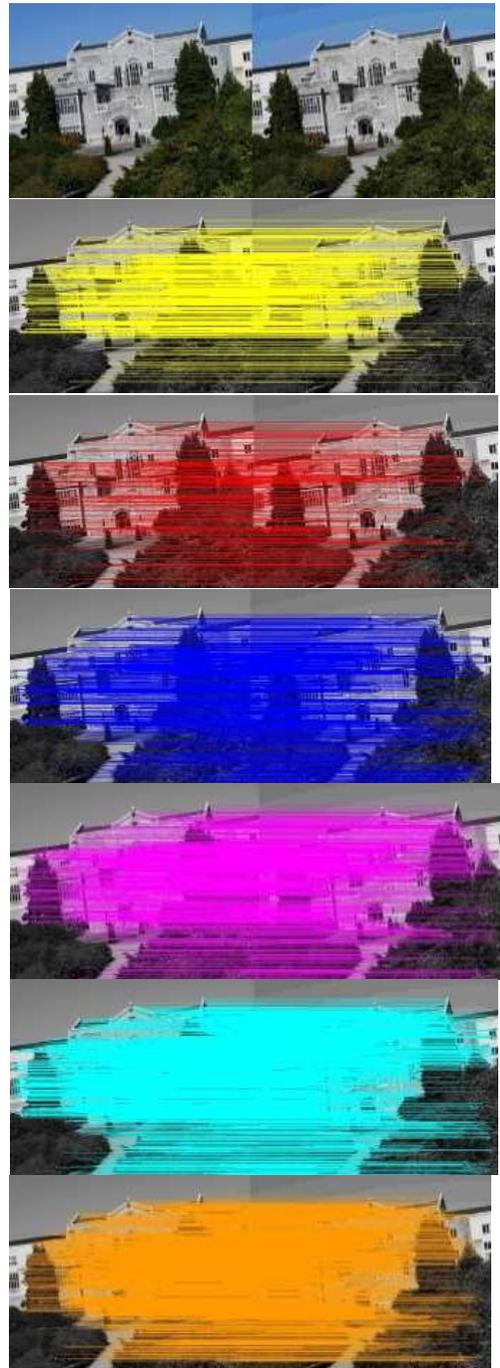


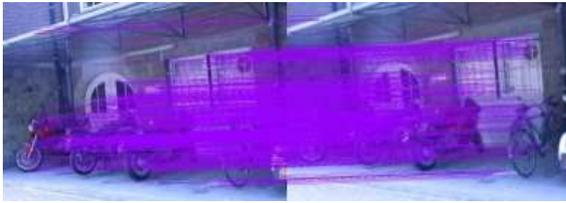


(e)

Figure 5. Image detection results of the image datasets (first row) (a) blur; (b) JPEG compression; (c) Illumination; (d) Viewpoint; (e) Zoom + rotation. In row sequence from second row to down: ORB, BRISK, SIFT, SURF, KAZE, A-KAZE, HLS-AKAZE.

To evaluate the detector performance, we measure the detector repeatability score between two correspondences image as defined in [17]. This system measures the ratio between corresponding keypoints and the minimum number of keypoints visible in both images. The overlap error is defined as the ratio of the intersection and union of the regions $\epsilon_s = 1 - (A \cap H^t B H) / (A \cup H^t B H)$, where A and B are the two regions and H is the corresponding homography between the images. When the overlap error between two regions is smaller than 60% (determined threshold), a correspondence is considered.

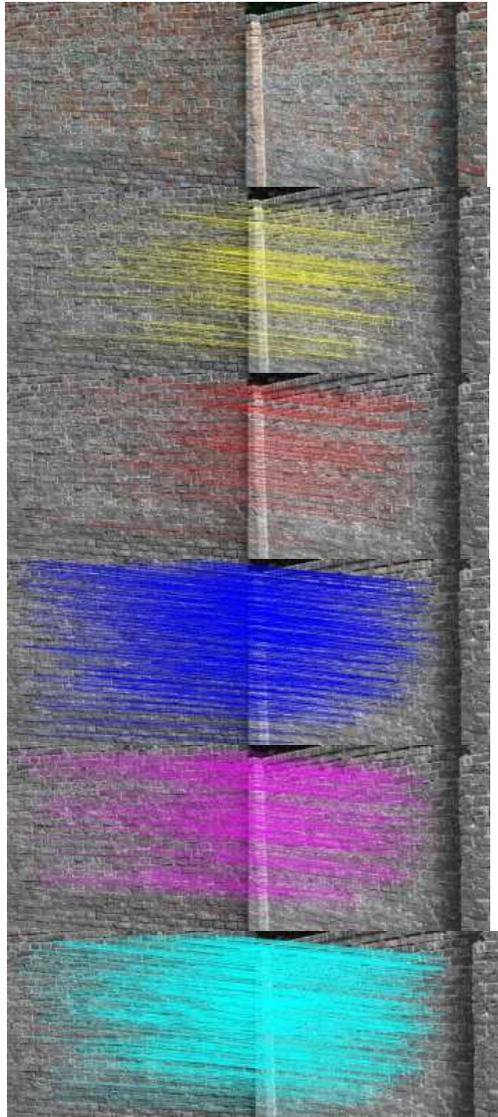




(a)

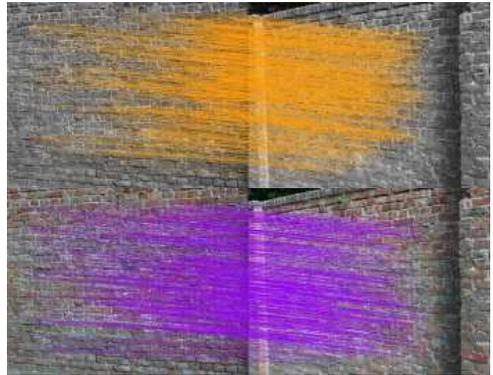


(b)

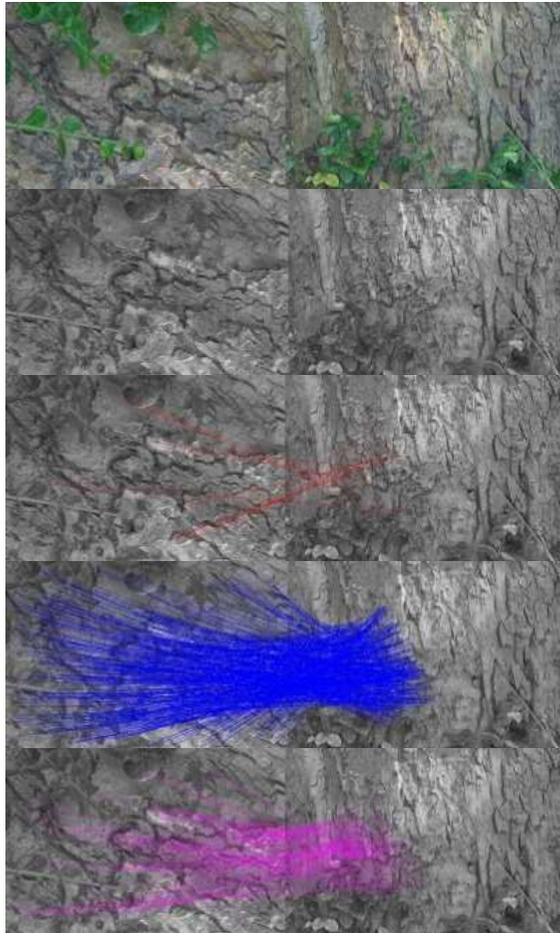


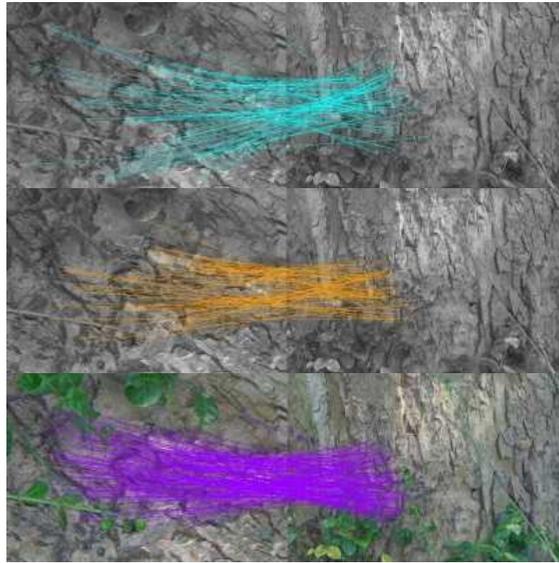


(c)



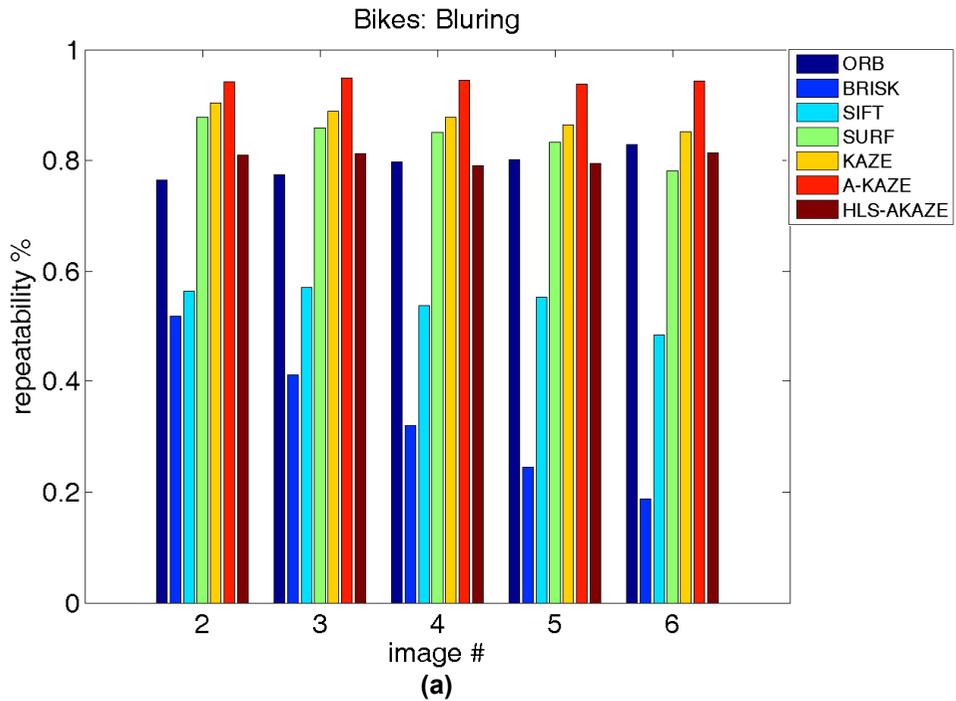
(d)

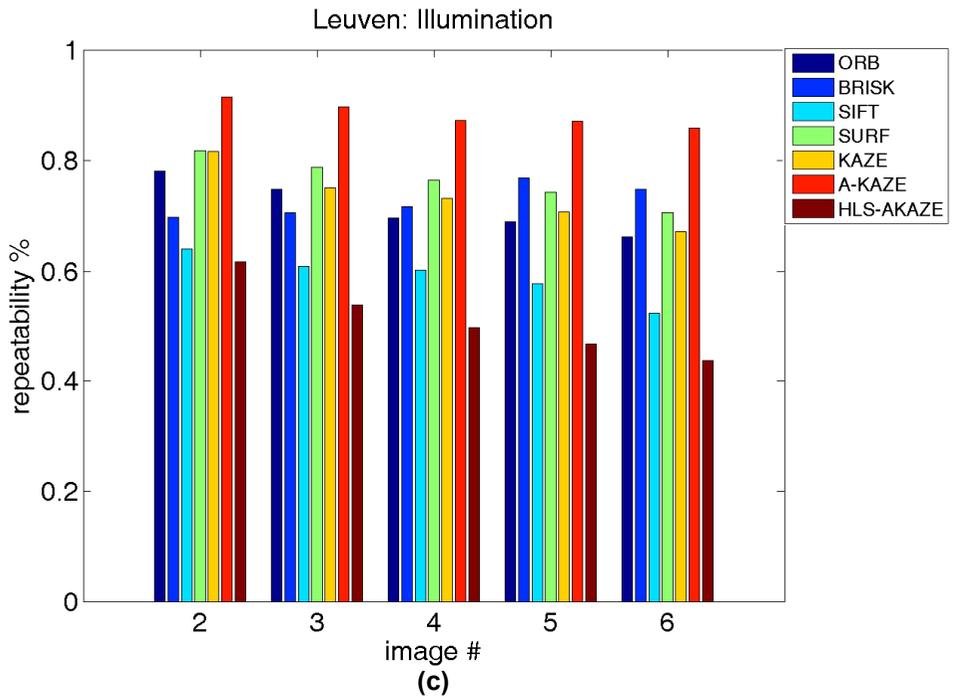
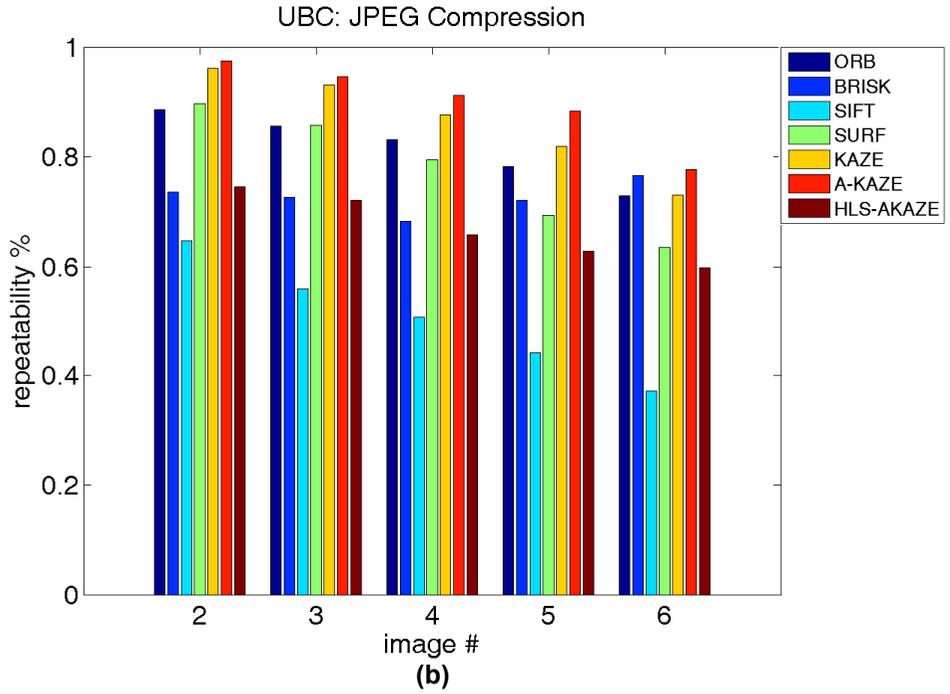




(e)

Figure 6. Image matching results of the first row image datasets (a) blur; (b) JPEG compression; (c) Illumination; (d) Viewpoint; (e) Zoom + rotation. In row sequence from second row to down: ORB, BRISK, SIFT, SURF, KAZE, A-KAZE, HLS-AKAZE.





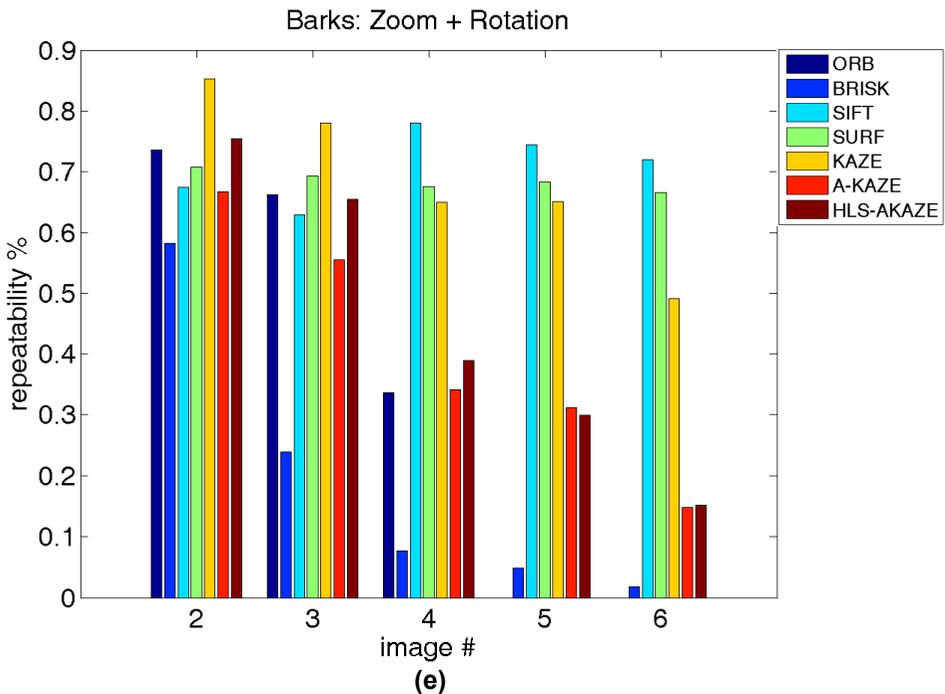
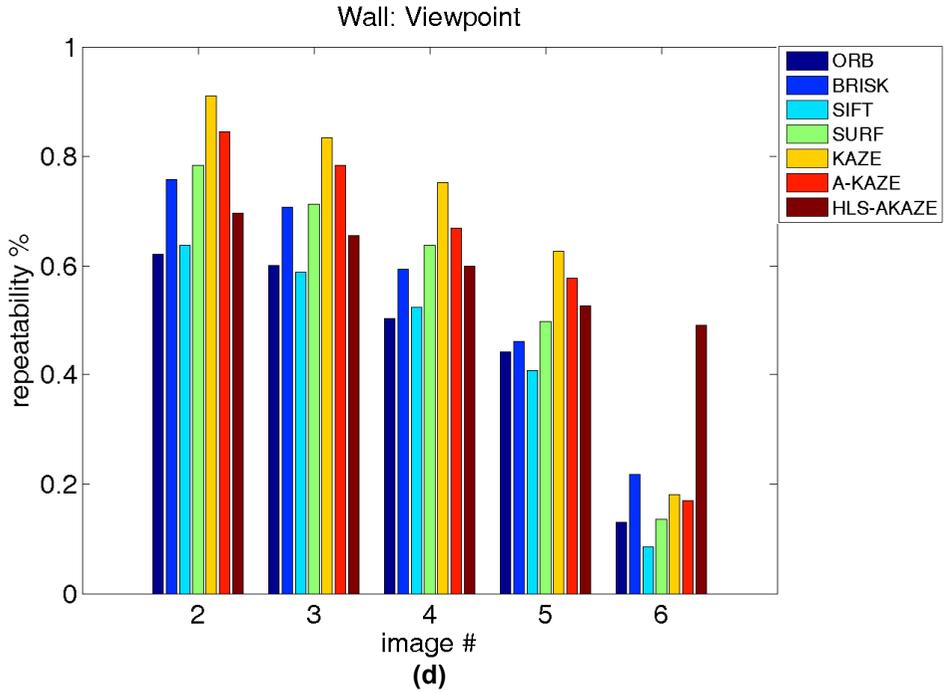
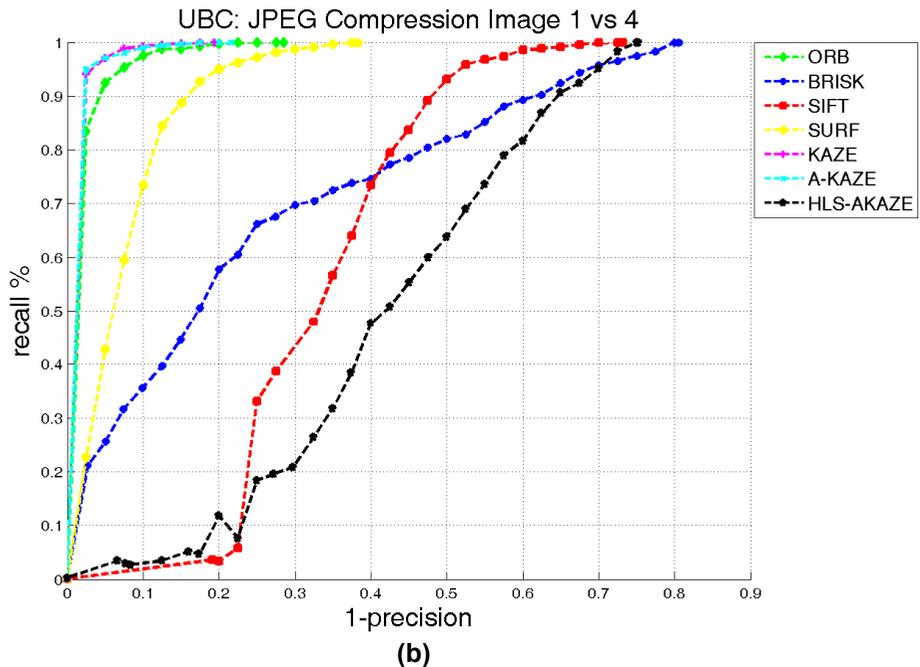
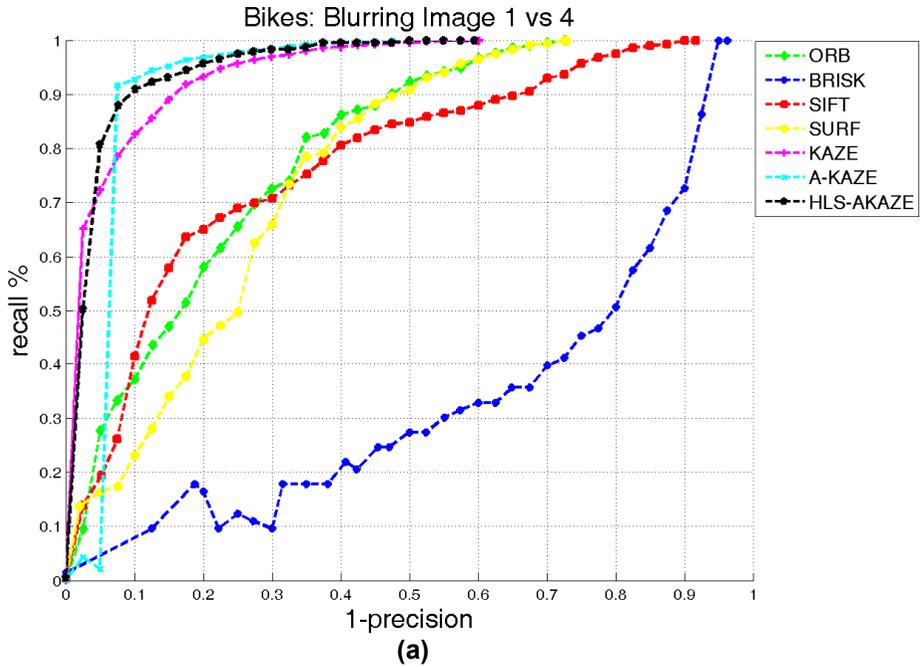
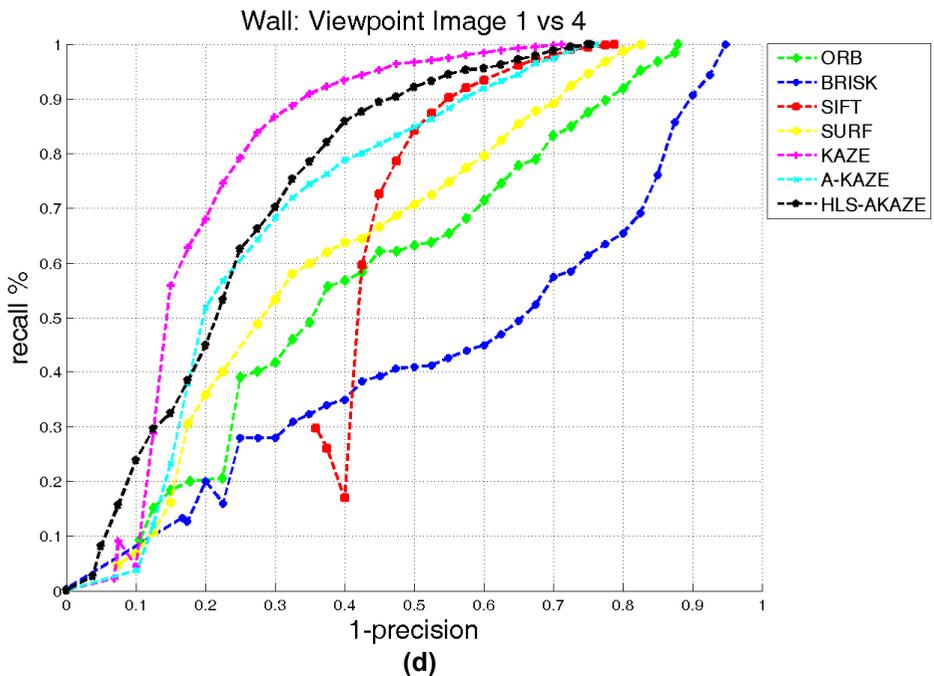
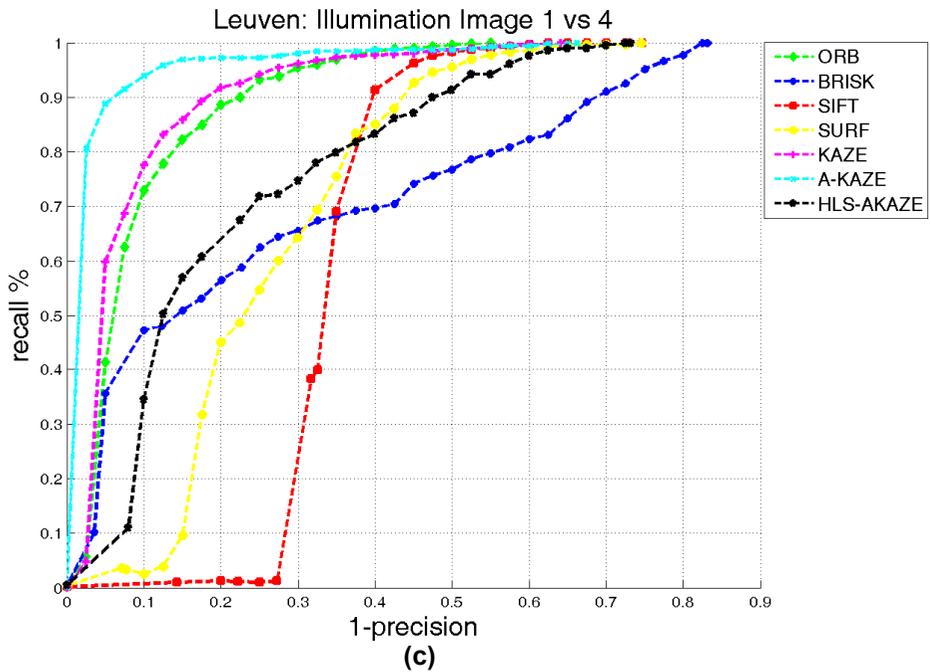


Figure 7. Detector repeatability score for an overlap area error 60%. Best view in color. (a) blur; (b) JPEG compression; (c) Illumination; (d) Viewpoint; (e) Zoom + rotation.

Figure 7 depicts the repeatability scores for all sequences from standard datasets. Each category of dataset is comparing the first image with the rest of images which is using the available ground truth homography which is needed to calculate the repeatability. As it can be observed, the repeatability scores of our proposed algorithm is comparable result in the blur images, but the performance in JPEG compression and illumination-changed images is not comparable with others. In the case of zoom and rotation, the performance is only good for the first two images. The reason of the case of poor repeatability performance is in the adaptive integrated determinant Hessian responses calculation. Integrating the H, L, and S determinant Hessian responses using adaptive local gradient weight response only works on certain image conditions. In JPEG compression and illumination-changed images, the hue and saturation of the corresponding images will be very different which cause different weight value calculated, and it means the maxima of the corresponding images will cause some differences. Otherwise, in viewpoint images, our system shows a significant stability in the repeatability score. The other methods show a significant degradation performance in the most changed viewpoint image. It means our features detection system is robust in the viewpoint images. The main reason is the 3 channels intensity values are not changed for the same corresponding point which means the calculation of determinant Hessian response using our proposed idea is working well.





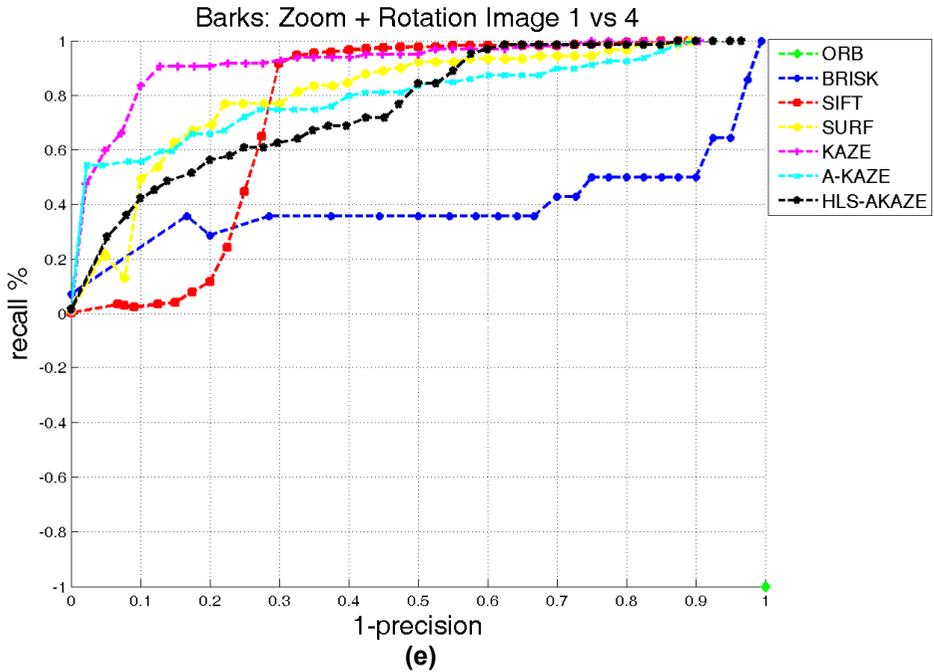
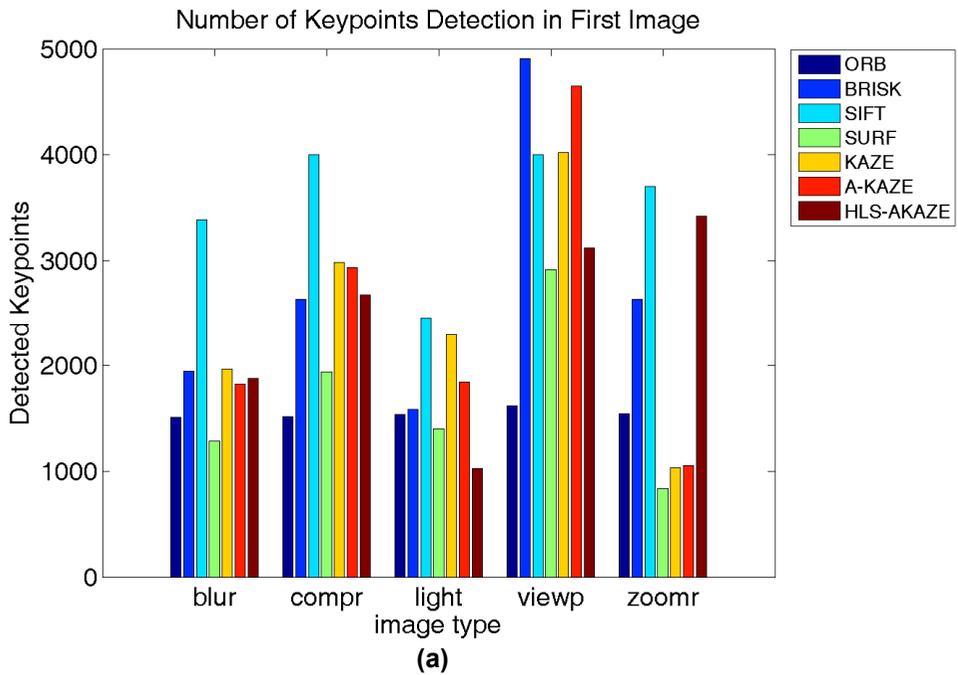
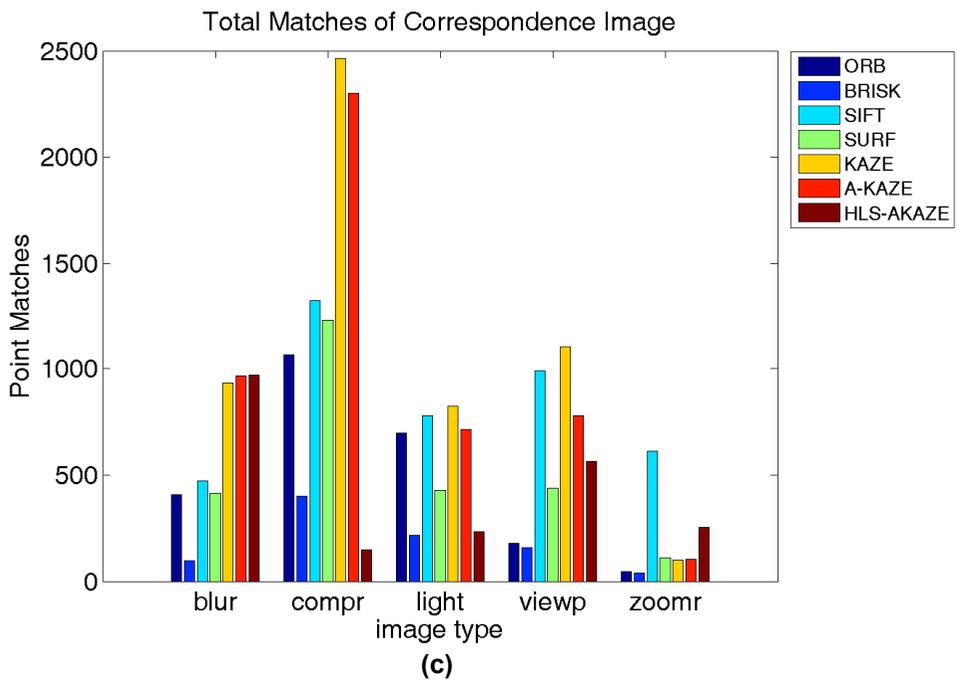
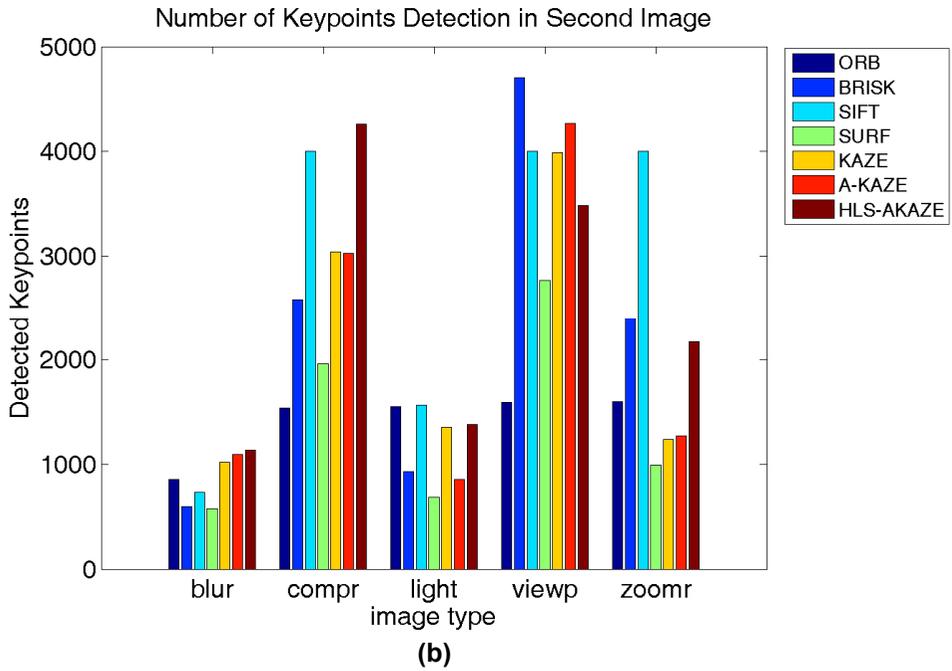
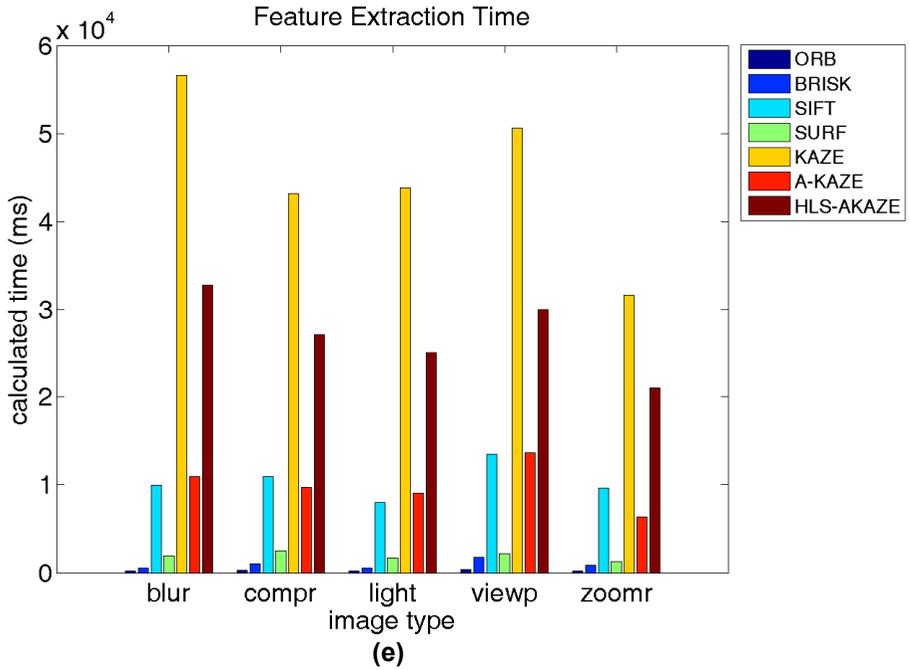
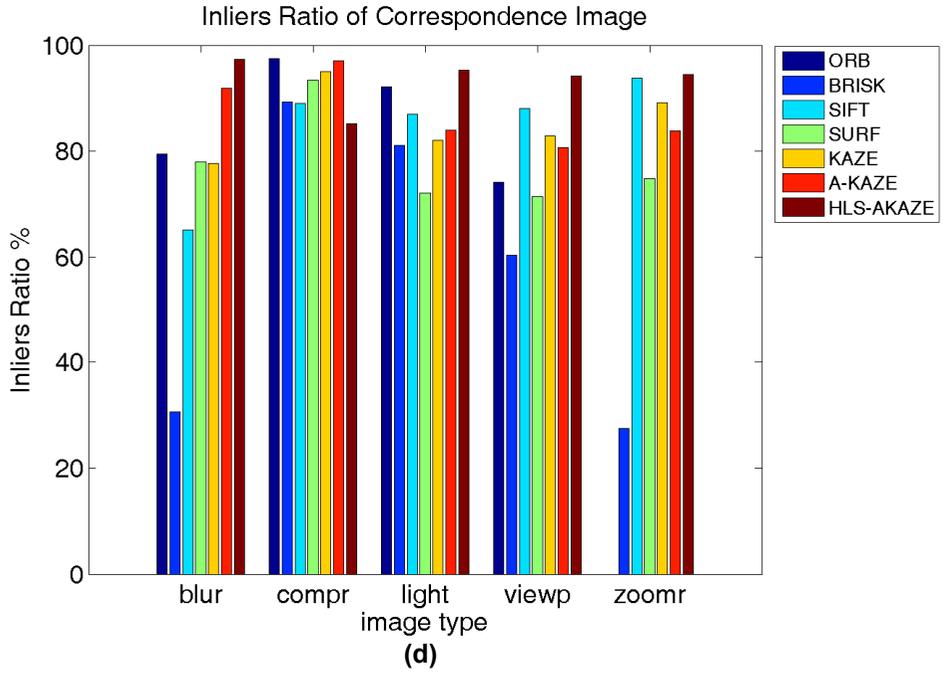


Figure 8. Recall vs 1-precision graphs for nearest neighbor matching strategy of image 1 vs image 4. Best view in color. (a) blur; (b) JPEG compression; (c) illumination; (d) Viewpoint; (e) Zoom + rotation.







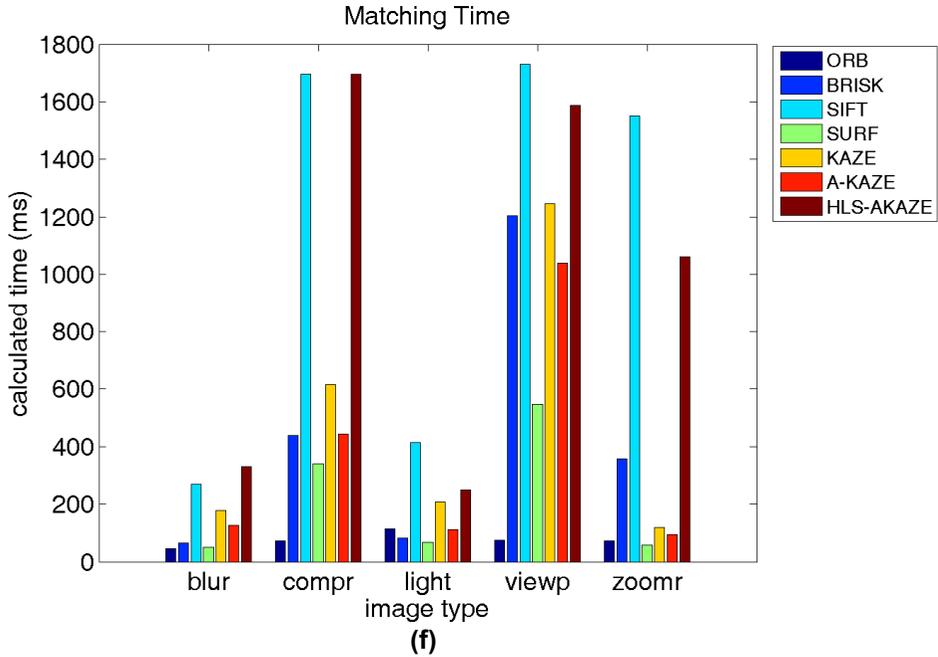


Figure 9. Benchmark image matching performance of image 1 vs image 4. Best view in color. (a) Number of detected keypoint image 1; (b) Number of detected keypoint image 2; (c) Total Matches of correspondence images; (d) Inliers Ratio; (e) Feature detection and description time; (f) Matching descriptor time.

As we can observed in figure 8, we evaluate the joint performance of the detection, description, and matching for each of the analyzed method. Descriptors are evaluated by means of precision-recall graphs as proposed in [18]. This criterion is based on the number of correct matches and the number of false matches obtained for an image pair:

$$recall = \frac{\#correct\ matches}{\#correspondences} , \quad 1 - precision = \frac{\#false\ matches}{all\ matches} \quad (28)$$

where the number of correct matches and correspondences is determined by overlap error. The overlap error for this precision-recall graph is using 50%. The correct and false matches consider the descriptor matching algorithm which we implement nearest neighbor matching strategy. The precision-recall performance of our proposed system shows low performance on JPEG compression images. And in

illumination-changed and zoom + rotation images, our proposed system shows a comparable performance with others. In the other hand, in blur and viewpoint images, our proposed system shows a good performance which is faster than other methods in achieving the maximum recall which is 1. In the case of low performance results, we can conclude the low repeatability score in detection influence the precision-recall performance. As we can see in this figure, the same category of low performance is shown. And another reason is our 9-bits per comparison descriptors of all HLS channels sometimes too fit (overfit). This condition produces a good result in certain conditions and also otherwise.

Figure 9 depicts 6 categories to measure image matching performance based on the number of detected keypoints, total matches, inliers ratio, and time performance both of extraction and matching. Our proposed system shows moderate number of detected keypoints from both corresponding images. This result just shows the qualitative data, which the number of keypoints detected depends on the pre-defined threshold. The total matching number of JPEG compression images shows a significantly low detection in Figure 9(c). The reason is the same with previous explanation about precision-recall graph. However, it doesn't affect the inlier ratio, where inlier ratio is the ratio of inliers respect to total matches. As we can see in Figure 9(d), our system shows robust performance of each category where almost all categories our system has the best performance. In feature detection and description time calculation (Figure 9(e)), our system is the second highest computational cost. The reason is because we build the nonlinear scale space for each channel individually. But using the FED scheme, our time performance is better than KAZE which is using AOS scheme. In Figure 9(f), our system is also the second highest computational cost in matching time. Descriptor size is the main factor of this performance. Other detector are represented in bits and also using a

smaller bits than our system. As we already know, SIFT is using 128 bytes descriptor which cause this method is slower than ours.



(a)



(b)



(c)



(d)

Figure 10. Two color images with their less-informative grayscale images. (a) wall_indoor images in color; (b) its grayscale images with the markers which show the less-informative region; (c) wall_street images in color; (d) its grayscale images with the markers which show the less-informative region.

The sample images of wall_indoor and wall_street are shown in Figure 10. The sample images are represented in color images (real images) and followed by their grayscale version. As we can see, in the grayscale version of both of images, there are some regions which are lost some important information (highlighted with markers). It means in color images, those regions have important edges which are not seen in the grayscale version because the intensity values are homogenous. In example of the wall_indoor images, there is no difference in grayscale pixel value of the clothes and his arms, which in color image shows very prominent difference. And also in wall_street images, the black line textures with blue strip lines inside is only showing black line textures in grayscale version.

Figure 11 shows the detection results of wall_indoor images using A-KAZE as the grayscale-based feature extraction method and our proposed methods as the color-based feature extraction methods (HLS-AKAZE Grad and HLS-AKAZE SD). As we can observe, Both HLS-AKAZE Grad and HLS-AKAZE SD methods can detect important keypoints in the same grayscale value regions where in the other

hand A-KAZE method only can extract luminance important keypoints. Figure 12 shows the detection results of wall_street images A-KAZE, HLS-AKAZE Grad, and HLS-AKAZE SD. With the same explanation as above, we can see A-KAZE method cannot detect important information in the blue strips region.



(a)

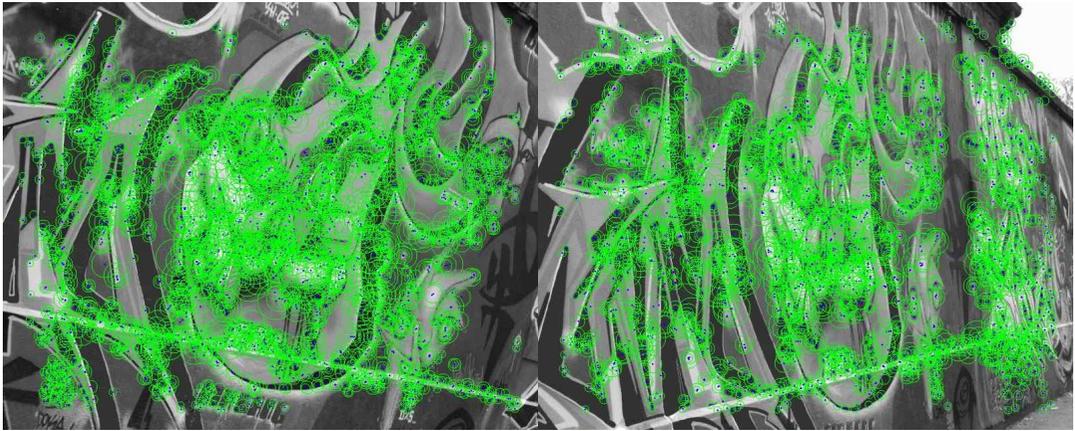


(b)

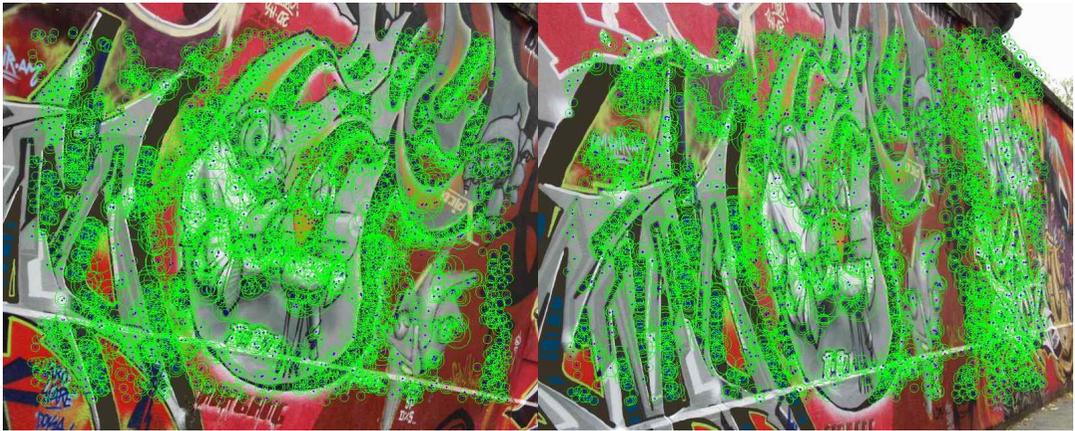


(c)

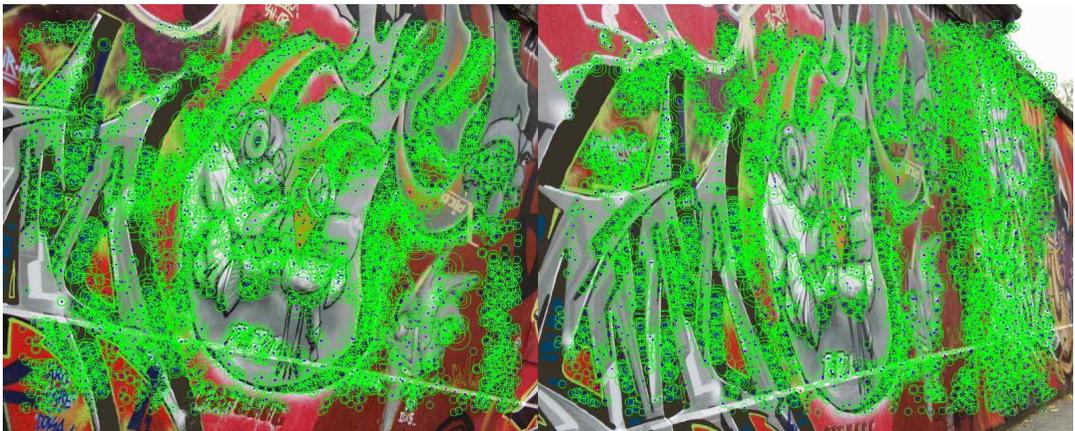
Figure 11. Image detection results of the wall_indoor images. (a) A-KAZE method; (b) HLS-AKAZE Grad method; (c) HLS-AKAZE SD method.



(a)



(b)



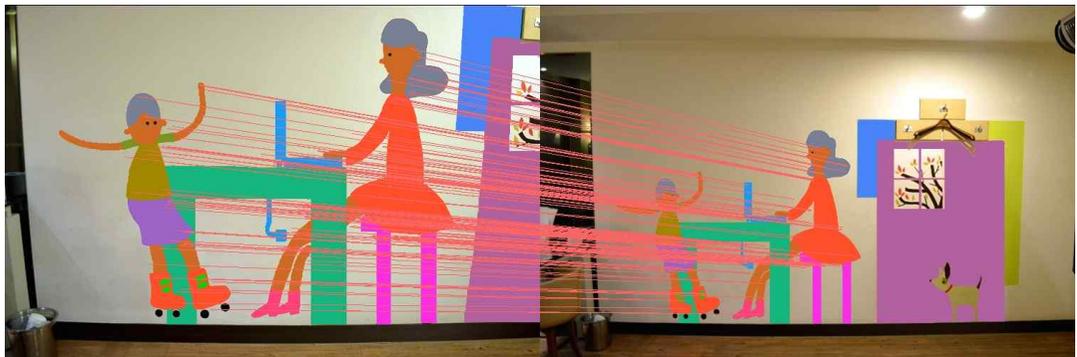
(c)

Figure 12. Image detection results of the wall_street images. (a) A-KAZE method; (b) HLS-AKAZE Grad method; (c) HLS-AKAZE SD method.

Figure 13 and 14 shows the image matching results of both images using A-KAZE, HLS-AKAZE Grad, and HLS-AKAZE SD. In the feature detection, A-KAZE method cannot detect some color important keypoints. It cause in the matching step, the experimental results just reflect the previous detection performance.



(a)



(b)



(c)

Figure 13. Image matching results of the wall indoor images. (a) A-KAZE method; (b) HLS-AKAZE Grad method; (c) HLS-AKAZE SD method.



(a)

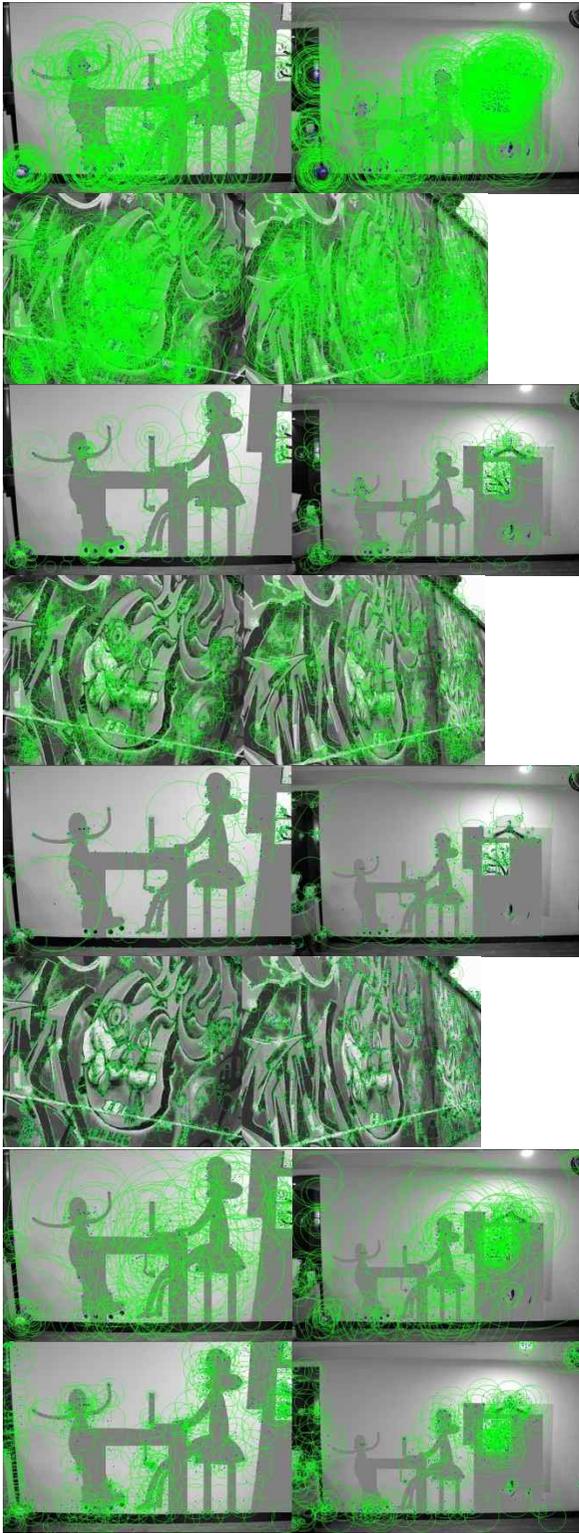


(b)

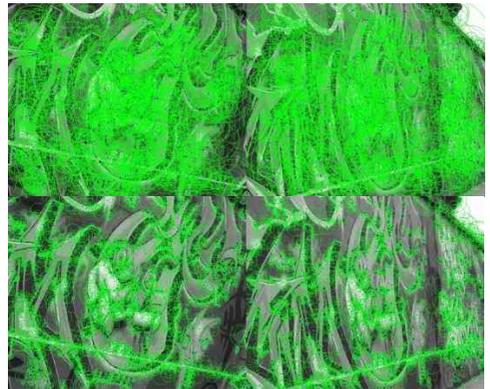


(c)

Figure 14. Image matching results of the wall_street images. (a) A-KAZE method; (b) HLS-AKAZE Grad method; (c) HLS-AKAZE SD method.



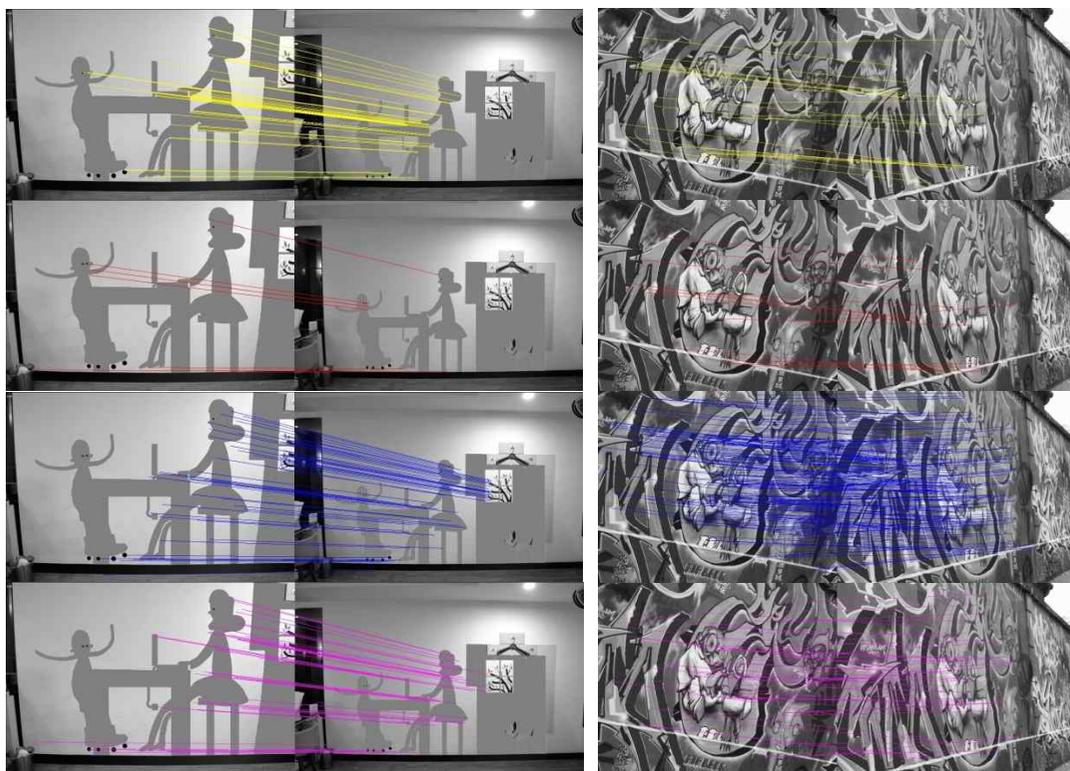
(a)



(b)

Figure 15. Image detection results of the less-informative grayscale image datasets (a) wall_indoor; (b) wall_street using other grayscale-based feature detector methods. In row sequence from top to down: ORB, BRISK, SIFT, SURF, KAZE.

Figure 15 and 16 shows the detection and matching results of the wall_indoor and wall_street images using the other grayscale-based feature extraction methods which are ORB, BRISK, SIFT and SURF. As we can see in those figures, the performance and analysis of those experimental results is the same with A-KAZE method. The performance of matching results is dependent with the detection results. Same with A-KAZE, the other grayscale-based feature extraction methods cannot extract and detect the important keypoints in terms of color because of the homogenous intensity in the grayscale images.



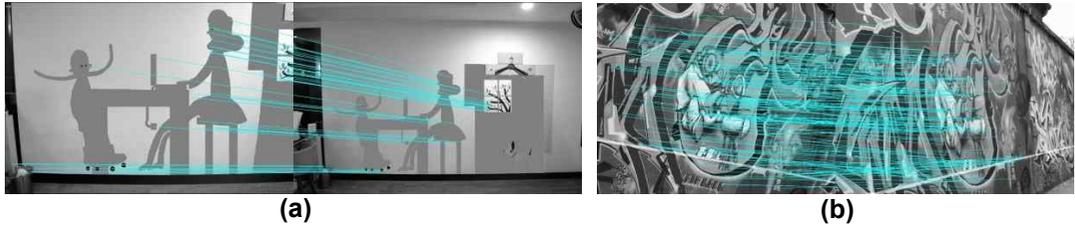
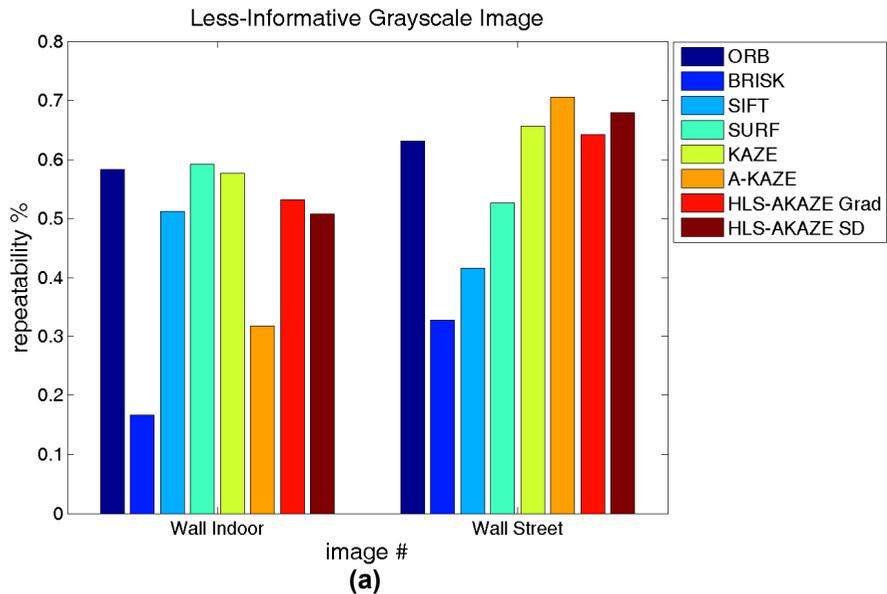


Figure 16. Image detection results of the less-informative grayscale image datasets (a) wall_indoor; (b) wall_street using other grayscale-based feature detector methods. In row sequence from top to down: ORB, BRISK, SIFT, SURF, KAZE.

Figure 17 shows the matching results of both images using all feature extraction methods. The previous detection results of grayscale-based feature detector cause the matching results cannot match some important color keypoints. The detector performance is shown in figure 17(a) which shows the repeatability score. The grayscale-based feature detector methods have almost the same score with our proposed method. The reason is because there are still many important keypoints beside the important color keypoints. However, in figure 17(b) and (c), in the recall-precision graph, our proposed method shows a high performance which means our system is robust in facing this kind of conditions.



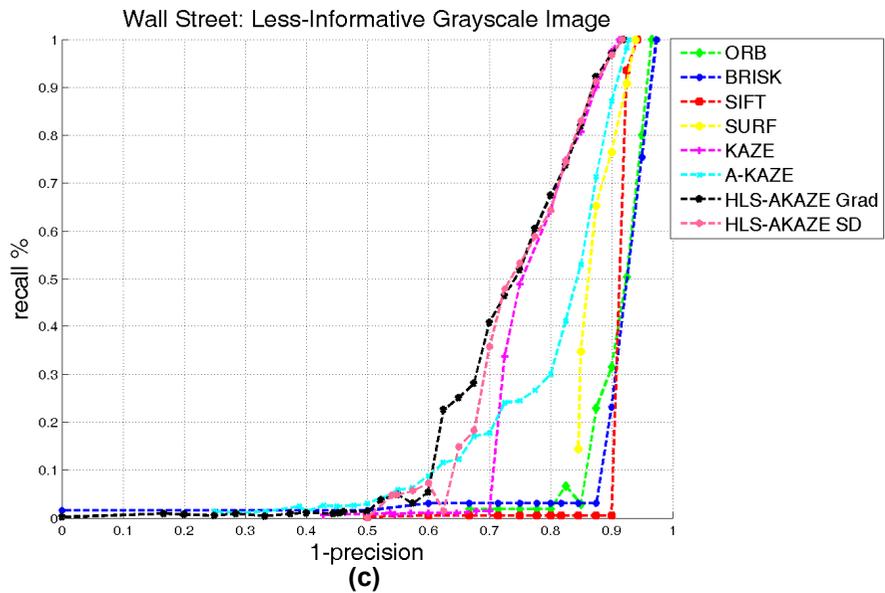
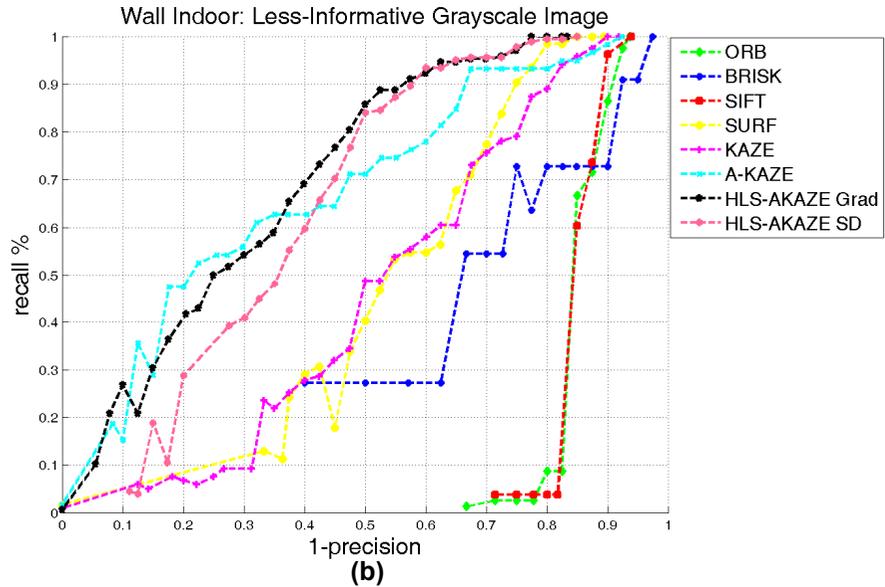
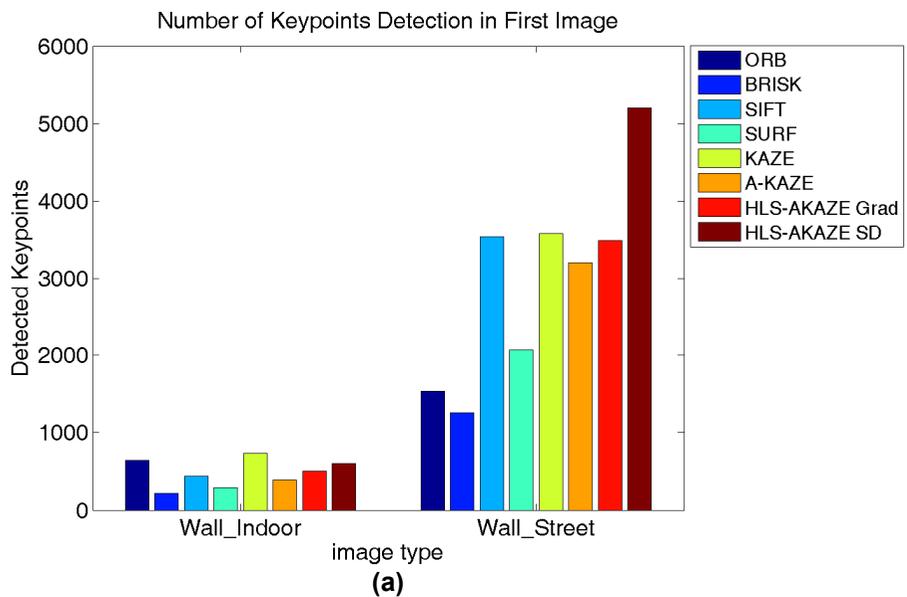
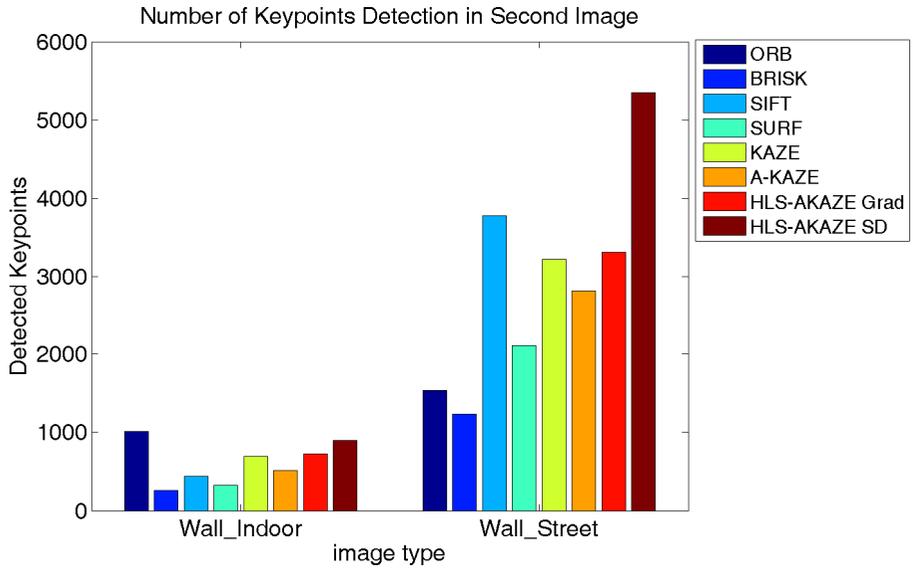


Figure 17. (a) Detector repeatability score for an overlap area of 50%; (b) Recall vs 1-precision graphs for wall_indoor images; (c) Recall vs 1-precision graphs for wall_street images.

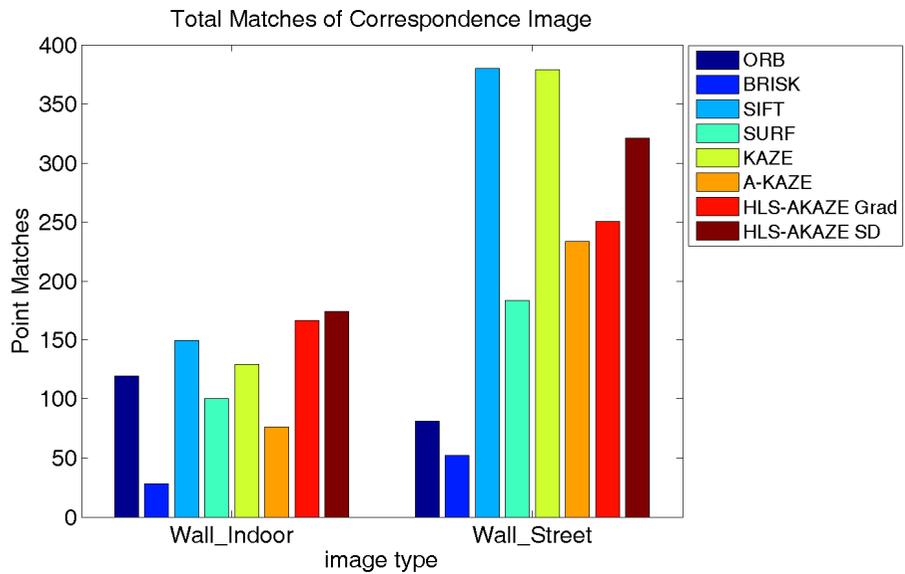
As we can see in Figure 18, the number of keypoints detection in the first and second images shows a significant detection in HLS-AKAZE SD. This condition occurs because the determinant Hessian response of Hue and Saturation of the both

sample images is very high, concurrently with the weighted response using SD. Our system is significantly better in inlier ratio performance than others in the wall_indoor images. As we can see in Figure 18(d), wall_indoor sample image has more important color information than its luminance that cannot be detected with grayscale-based feature detectors. In wall_street images, the loss of color important information is not comparable with its luminance information. In time performance, HLS-AKAZE SD is shown as the highest computational cost.

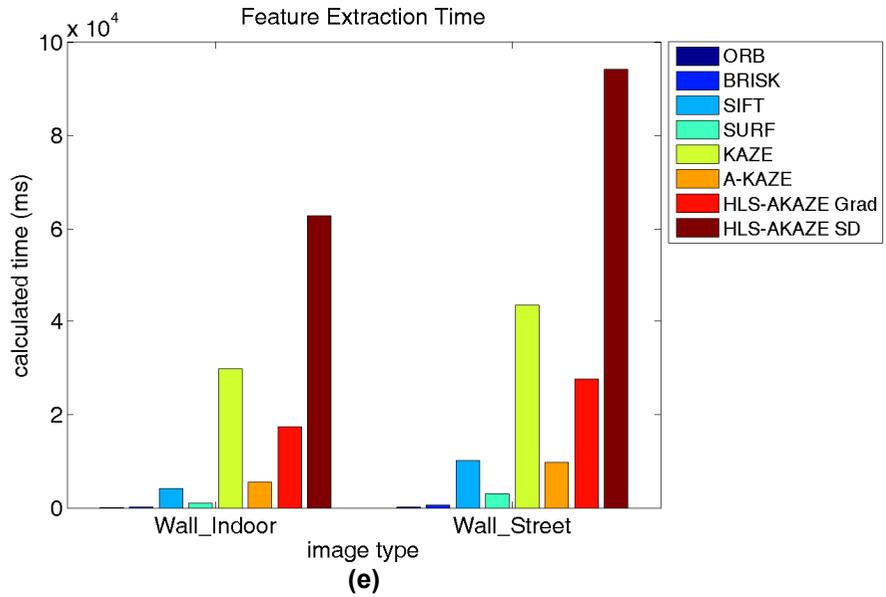
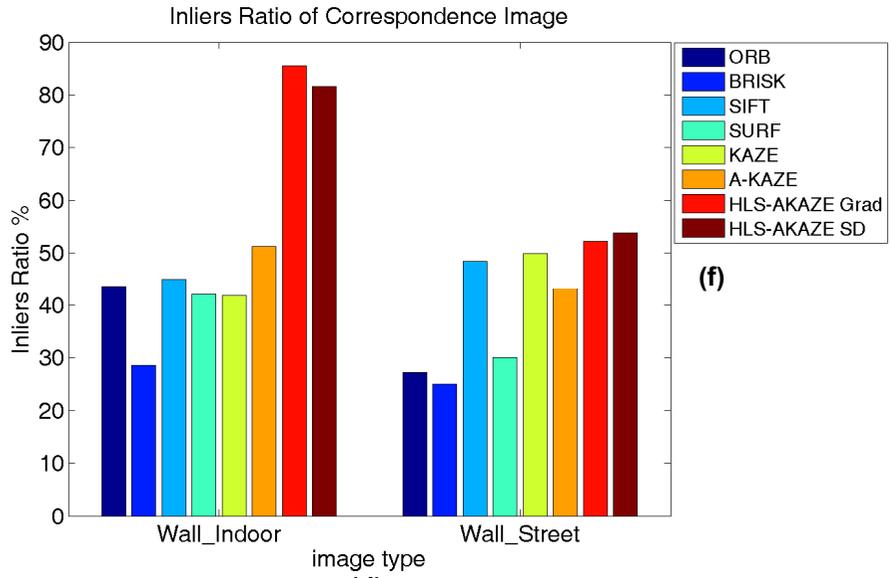




(b)



(c)



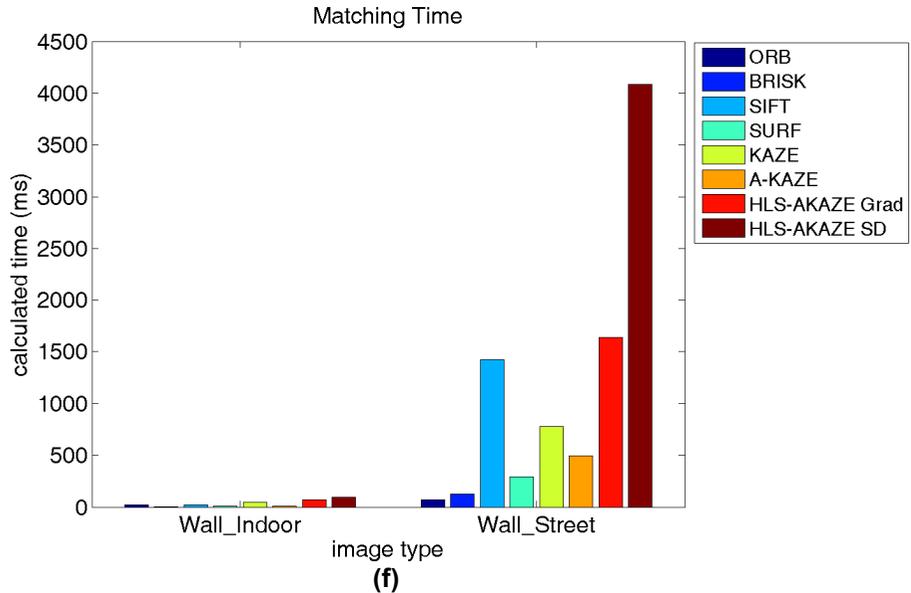


Figure 18. Benchmark image matching performance of both less-informative grayscale images. Best view in color. (a) Number of detected keypoint image 1; (b) Number of detected keypoint image 2; (c) Total Matches of correspondence images; (d) Inliers Ratio; (e) Feature detection and description time; (f) Matching descriptor time.

V. CONCLUSIONS

In this paper, we have presented HLS-AKAZE feature, an invariant color features detection and description in nonlinear scale spaces. The nonlinear scale spaces are built using FED schemes in each of Hue, Lightness, and Saturation channel individually. By means of FED schemes embedded in pyramidal approach, we can reduce computationally time compared with using AOS schemes. By using nonlinear scale spaces, the scale invariant detector can be created using the diffusion algorithm which can reduce the noise in the image but still retain the detail information.

In the feature detection, we find local maxima based on adaptive integrated determinant of Hessian response. The determinant Hessian response is the integration of each HLS channel which is weighted adaptive to its importance. There are two proposed weighted response methods; standard deviation and local gradient weight response. In the descriptor level, Color Modified-Local Difference Binary (CM-LDB) descriptor is introduced as our scale-, rotation-, and illumination-invariant descriptor which exploits HLS color information concurrently with its gradient information from nonlinear scale spaces.

The performance of detection, description, and matching based on experimental results using standard datasets show that our system is robust in the viewpoint-changed and blur images. In the certain image conditions where there are many important color information which have the same grayscale value, our system is showing significant performance result compared with other grayscale-based feature detectors.

BIBLIOGRAPHY

- [1] Davison. Real-time simultaneous localization and mapping with a single camera,” in Proc. IEEE Intl. Conf. Comput. Vis., 2003, pp. 1403–1410.
- [2] Kanan, C. & Cottrell, G. W. (2012). Color-to-Grayscale: Does the Method Matter in Image Recognition? PLoS ONE, Edited by Eshel Ben-Jacob, vol. 7, issue 1, 2012.
- [3] G. Lowe. Distinctive image features from scale-invariant keypoints. Intl. J. of Computer Vision, 60(2):91–110, 2004.
- [4] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. Computer Vision and Image Understanding, 110(3):346–359, 2008.
- [5] S. Leutenegger, M. Chli, and R. Y. Siegwart. BRISK: Binary robust invariant scalable keypoints. In IEEE Intl. Conf. on Computer Vision (ICCV), pages 2548–2555, 2011.
- [6] Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: an efficient alternative to SIFT or SURF. In IEEE Intl. Conf. on Computer Vision (ICCV), pages 2564–2571, 2011.
- [7] P. F. Alcantarilla, A. Bartoli, and A. J. Davison. KAZE features. In Eur. Conf. on Computer Vision (ECCV), pages 214–227, 2012.
- [8] P. F. Alcantarilla, J. Nuevo, and A. Bartoli, and. A-KAZE features. In Eur. Conf. on Computer Vision (ECCV), 2013.
- [9] X. Yang and K. T. Cheng. LDB: An ultra-fast feature for scalable augmented reality. In IEEE and ACM Intl. Sym. on Mixed and Augmented Reality (ISMAR), 2012.

- [10] D. A. Rojas, F.S. Khan, J. Weijer, and T. Gevers. The impact of Color on Bag-of-Words based Object Recognition. 2010
- [11] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(7):1651–1686, 1990.
- [12] Weickert, J. Efficient image segmentation using partial differential equations and morphology. *Pattern Recognition* 34, 1813–1824, 2001.
- [13] S. Grewenig, J. Weickert, and A. Bruhn. From box filtering to fast explicit diffusion. In *Proceedings of the DAGM Symposium on Pattern Recognition*, pages 533–542, 2010.
- [14] D. Coomans; D.L. Massart (1982). "Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-Nearest neighbour classification by using alternative voting rules". *Analytica Chimica Acta* 136: 15–27.
- [15] Knuth, D.E., Morris Jr, J.H., Pratt, V.R. 1977. Fast pattern matching in strings. *SIAM J. Comput.* 6(1):323–350.
- [16] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [17] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A comparison of affine region detectors. *Intl. J. of Computer Vision* 65, 43–72 , 2005.
- [18] Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Machine Intell.* 27, 1615–1630. 2005.