

Radial Fourier Analysis (RFA) Descriptor with Fourier-based Keypoint Orientation

S. C. F. Lin

*School of Mechanical and Manufacturing Engineering
The University of New South Wales
UNSW Sydney, NSW, 2052, Australia*

stephen.lin@unsw.edu.au

C. Y. Wong

*School of Mechanical and Manufacturing Engineering
The University of New South Wales
UNSW Sydney, NSW, 2052, Australia*

chin.wong@unsw.edu.au

G. Jiang

*School of Mechanical and Manufacturing Engineering
The University of New South Wales
UNSW Sydney, NSW, 2052, Australia*

guannan.jiang@unsw.edu.au

M. A. Rahman

*School of Mechanical and Manufacturing Engineering
The University of New South Wales
UNSW Sydney, NSW, 2052, Australia*

md.arifur.rahman056@unsw.edu.au

N. M. Kwok

*School of Mechanical and Manufacturing Engineering
The University of New South Wales
UNSW Sydney, NSW, 2052, Australia*

nmkwok@unsw.edu.au

Abstract

Local keypoint detection and description have been widely employed in a large number of computer vision applications, such as image registration, object recognition and robot localisation. Since currently available local keypoint descriptors are based on the uses of statistical analysis in spatial domain, a local keypoint descriptor, namely Radial Fourier Analysis (RFA) keypoint descriptor, is developed with the use of spectral analysis in frequency domain. This descriptor converts image gradients around SIFT keypoints to frequency domain in order to extract the principle components of the gradients and derive distinctive descriptions for representing the keypoints. Additionally, a keypoint orientation estimate is also introduced to improve the rotational invariance of the descriptor rather than simply adopting SIFT keypoint orientations. The introduced orientation estimate employs the starting point normalisation of Fourier coefficients, which are frequency responses, to deduce rotating angles that ensure keypoint correspondences are aligned at the same orientation. Through experiments and comparisons, RFA descriptor demonstrates its outstanding and robust performances against various image distortions. Particularly, the descriptor has extremely reliable performances in dealing with the images, which are degraded by blurring, JPG compression and illumination changes. All these indicate that spectral analysis has strong potential for local keypoint description.

Keywords: Local Keypoint Descriptor, Keypoint Orientation, Fourier Transform, Keypoint Matching, SIFT Descriptor.

1. INTRODUCTION

Local keypoint detection and description have been frequently utilised as an integral system to accomplish diverse computer vision tasks due to its invariance capabilities against image distortions caused by many types of geometric and photometric transformations. For example, they have been widely applied to basic computer vision tasks including image retrieval [1], image registration [2] and object recognition [3]. As an extension from these basic tasks, they can be specifically designed to carry out object categorisation [4], texture classification [5] and face recognition [6]. In engineering, their applications could be related to medical diagnosis [7], robot localisation [8], land cover surveillance [9] and machine operation monitoring [10].

From the perspective of computer vision, local keypoint detection and description should be treated as independent processes as they produce different outputs and serve different purposes. Many keypoint detectors [11] have been proposed to search stable and repeatable keypoints in images, and keypoints generally contain the information of position and other supplementary measures to localise and define the properties of keypoints. One of the supplementary measures could be keypoint orientation that is commonly assigned to keypoints for rotational invariance. For example, SIFT detector [3] estimates keypoint orientation by creating a histogram of the image gradient orientations around a keypoint. Furthermore, ORB detector [12] defines keypoint orientation as the vector between a keypoint position and the intensity centroid of local image patch of the keypoint. Taylor and Drummond [13] compute keypoint orientation by simply choosing the longest vector among the vectors that are formed by intensity differences around a keypoint. All these keypoint orientation estimates are established on gradient and statistical approaches in spatial domain. It is advantageous if a different keypoint orientation estimate, such as using spectral analysis in frequency domain, can be developed to provide accurate keypoint orientations.

For keypoint description, it extracts the features concealed in the local image patches around keypoints and analyses the features to represent that keypoints. A variety of keypoint descriptors [14] have been devised to be invariant to geometric and photometric transformations such as image distortions caused by blur, rotation, scale, illumination and JPG compression changes. Many keypoint descriptors, such as SIFT [3], PCA-SIFT [15], and GLOH [14] descriptors, utilise histograms to determine the distributions of local image patches around keypoints in order to extract and analyse features, while achieving results that are invariant to image distortions. Summation of image gradients is also a robust approach in keypoint description, which is demonstrated by SURF descriptor [16]. Furthermore, BRISK descriptor [17] uses the comparisons of local image intensities around keypoints as features to represent the keypoints. Similar to the issue raised in keypoint orientation estimates, most keypoint descriptors are based on voting or statistical approaches in spatial domain. A local keypoint descriptor that extracts and analyses features in a different domain is essential as it could expand the diversity of local keypoint descriptions to increase its distinctiveness.

To address the overdevelopment of statistical analysis in spatial domain for keypoint orientation estimations and keypoint descriptions, this work relies instead on spectral analysis in frequency domain. For the keypoint orientation estimate, it utilises Fourier transform to convert the image intensities of local image patches to frequency domain from spatial domain. Afterward, starting point normalisation is performed on the intensity frequency to derive keypoint orientations in order to enhance the rotational invariance of keypoints. The proposed keypoint descriptor, which is named as Radial Fourier Analysis (RFA) descriptor, transforms the image gradients of local image patch to frequency domain, then the gradient frequency is decomposed to provide a unique keypoint description. Both of the transformed frequencies are not inverted back to spatial domain as the keypoint orientation and description are directly carried out in frequency domain.

The rest of the paper is organised as follows. In Section 2, the principle of SIFT keypoint detector is briefly revised, and a keypoint orientation estimate is developed to replace SIFT keypoint orientation. Section 3 presents the procedures to generate RFA descriptor with the reasons declared for each procedure. This is followed by Section 4, where an experiment setup to

evaluate the performances of the new descriptor is detailed. The performance results and comparisons are illustrated in Section 5. Significant observations from the results are compared and highlighted for discussions. Finally, in Section 6, all these works are concluded by suggesting possible research directions in accordance with the performances of the proposed keypoint descriptor and orientation estimate.

2. IMAGE KEYPOINT DETECTION

In this section, SIFT keypoint detector is briefly explained as SIFT keypoints are utilised to generate RFA descriptors. In order to improve the rotational invariance of SIFT keypoints, a novel estimate of keypoint orientation using Fourier normalisation is proposed to replace SIFT keypoint orientation.

2.1 SIFT Keypoint Detector

SIFT descriptor has been concluded by [18] [19] as the state-of-the-art in the research and development of local keypoint descriptor. As such, SIFT keypoints are used to localise salient points because of its trustworthy characteristics as well as providing a fair comparison base of describing power between our descriptor and SIFT descriptor. Fundamentally, SIFT keypoints are detected by utilising Difference of Gaussian (DoG) operator that was initially introduced by Crowley and Parker [20]. However, Lowe's extension of the operator [3] to detect scale invariant keypoint in a given image is more widely used and is adopted in this work.

SIFT keypoint detector involves the following three steps: (1) scale-space extrema detection; (2) keypoint localisation; and (3) orientation assignment. The first step is DoG operation, where input image is convoluted with Gaussian kernels at different scales to produce a series of Gaussian images. Adjacent images in the series are subtracted from each other, forming DoG images. A $3 \times 3 \times 3$ window is then employed to search the local maximums and minimums in the DoG images, and these extremes are labelled as potential keypoints. The second step filters potential keypoints by measuring their contrast and edge response, within which the keypoints with low contrast and high edge response are discarded. The remaining keypoints are subsequently re-localised to sub-pixel accuracy and identified as SIFT keypoints. In the final step, the dominant gradient orientation of each SIFT keypoint is estimated through binning the gradient orientations around the keypoint into a histogram and the peak of the formed histogram is taken as the gradient orientation of keypoint. The dominant gradient orientations are estimated and assigned to SIFT keypoints to enhance its rotational invariance ability. The last two steps are expanded from DoG operator for stabilising the keypoints in matching and recognition tasks.

2.2 Keypoint Orientation Estimate by Fourier Analysis

Other than histogram-based estimate of keypoint orientation, such as SIFT keypoint orientation, RFA descriptor adopts a simple and effective approach to compute keypoint orientation with Fourier analysis. This Fourier-based keypoint orientation estimate consists of three processing steps: (1) intensity sample collection; (2) Fourier transform; and (3) orientation estimate.

In the first processing step, the neighbouring image intensities around a keypoint are circularly sampled and arranged into a vector. The procedure of this intensity sample collection is presented in Algorithm 1. The intensity sample collection is preformed at the image scale where the keypoint is detected. For each sampling angle, R_θ , image intensities are sampled along the radial direction to represent the intensity signatures at that angle. This sampling process is performed at counter-clockwise direction till a full circular sampling is completed. The intensity vector, $\mathbf{z}(m)$, has the image intensity samples arranged in an ascending order with respect to angular direction, and each sampling angle is represented by R_θ image intensities. The intensity vector, $\mathbf{z}(m)$, can be described as,

$$\mathbf{z}(m) = \{I(u, v)\} , I(u, v) \in \mathbb{R} \quad (1)$$

where, $m = 1 \dots M$, $M = R_o \times 360^\circ / \Delta\theta_o$ and $I(u, v)$ is the image intensity at a given sampling position (u, v) . An illustration of this intensity sample collection is shown in Figure 1 with a radial sampling range of 5 pixel radius (i.e., $R_o = 5$) and an angular sampling interval of 10° (i.e., $\Delta\theta_o = 10^\circ$) for demonstration purpose. Each black line connecting the red circles is a set of samples along the radial direction for that sampling angle. These sets of samples are then arranged into the intensity vector, $\mathbf{z}(m)$, with the order from 0° to $360^\circ - \Delta\theta_o$ as mentioned previously.

ALGORITHM 1: Intensity Sample Collection for Orientation Estimate by Fourier
Input gray image, I
Input keypoint position, (x_0, y_0)
Set radial sampling range, $R_o = 16$
Set angular sampling interval, $\Delta\theta_o = 7.5^\circ$
Set sampling angle index, $\theta_i = 360^\circ / \Delta\theta_o - 1$
Initialise \mathbf{z} to a zero vector of the length $R_o \times 360^\circ / \Delta\theta_o$
For $i = 0$ to θ_i do
Determine sampling angle
$\theta_o = \Delta\theta_o \times i$
For $r = 1$ to R_o do
Determine the sampling position
$u = r \times \cos \theta_o + x_0$
$v = r \times \sin \theta_o + y_0$
Sample image intensity at $I(u, v)$
Distribute the image intensity into \mathbf{z}
End for
End for
Output \mathbf{z}

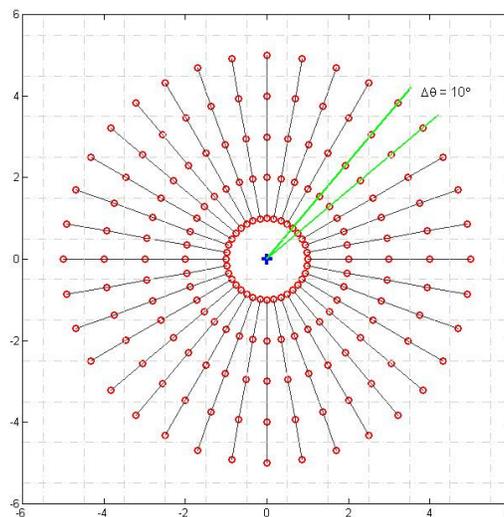


FIGURE 1: Intensity sample collection for keypoint orientation estimate. The red circles indicate the positions of samples and the blue cross sign represents the location of keypoint. Each black line connecting the red circles is a set of samples for that sampling angle.

For the second processing step, a discrete Fourier transform is performed on the obtained intensity vector, $\mathbf{z}(m)$, that is,

$$Z(h) = \sum_{m=0}^{M-1} z(m) e^{-\frac{j2\pi mh}{M}}, \quad Z(h) \in \mathbb{C} \quad (2)$$

where, $h = -M/2 \dots M/2 - 1$. Here, $Z(h)$ denotes the h^{th} Fourier coefficient (i.e., frequency response) based on the information given by $\mathbf{z}(m)$ that is obtained by Algorithm 1. Note that, the Fourier coefficients, $Z(h)$, are shifted to the centre.

The final processing step estimates the keypoint orientation by using the phase of first positive Fourier coefficient (i.e., $Z(1)$). The phase is calculated by,

$$\rho = \tan^{-1} \left(\frac{\text{Im}(Z(1))}{\text{Re}(Z(1))} \right) \quad (3)$$

where, $\text{Re}(\dots)$ and $\text{Im}(\dots)$ denote the real and imaginary parts of the Fourier coefficient respectively. Since this calculated phase is in frequency domain, it must be converted back to spatial domain to properly express the phase as keypoint orientation with regard to the arrangement of image intensity samples in the intensity vector, $\mathbf{z}(m)$. The conversion from phase, ρ , to keypoint orientation, θ , is defined as,

$$\theta = \rho \times \frac{M}{2\pi} \times \frac{\Delta\theta_o}{R_o} \quad (4)$$

The multiplication of $\rho \times M/2\pi$ implies that the phase is converted to spatial domain from frequency domain. After the conversion, the term $\Delta\theta_o/R_o$ aligns the converted phase to coincide with the angular sampling interval and radial sampling range in intensity sample collection.

The principle of this orientation estimate is based on the property of starting point of Fourier coefficients. For example, a complex signal is defined as $\mathbf{a}(t)$ and its Fourier coefficients are defined as $\mathbf{b}(w)$. Considering the signal is shifted by an amount of t_0 in its sequence that merely changes the starting point of the signal to $t = t_0$ from $t = 0$. The shifted signal is expressed as $\mathbf{a}_p(t) = \mathbf{a}(t - t_0)$, whose Fourier coefficients are

$$\begin{aligned} b_p(w) &= \sum_{\substack{t=0 \\ t=T-1}}^{T-1} a_p(t) e^{-\frac{j2\pi tw}{T}} \\ &= \sum_{t=0}^{T-1} a(t - t_0) e^{-\frac{j2\pi(t-t_0)w}{T}} \\ &= b(w) e^{\frac{j2\pi t_0 w}{T}} \end{aligned} \quad (5)$$

for $t = 0, 1, 2, \dots, T - 1$. Thus, shifting the starting point of the signal causes a linear phase shift in all Fourier coefficients that depends on the coefficient index (i.e., w) and the amount of shift (i.e., t_0).

It can be seen from Algorithm 1 that the keypoint orientation estimate is transformed to starting point normalisation as the neighbouring image pixels around a keypoint are sampled and arranged in the order corresponding to angular direction. Then, Eq. 3 determines the amount of linear phase shift that is imposed on the Fourier coefficients, $Z(h)$, obtained in Eq. 2. Lastly, Eq. 4 converts the linear phase shift to the change of starting point in the intensity vector, $z(m)$. It also reduces the computation cost by avoiding the use of inverse Fourier transform to determine the change of starting point in spatial domain.

Another issue to be considered while applying Fourier transform is the relationship between input and output of the transform. As the input of Fourier transform is a complex function, the output coefficients are difficult to predict and do not obey any symmetries since they are the sum of all the real and imaginary parts of the input. However, if the input of Fourier transform is a real function, the real and imaginary parts of the output coefficients are even and odd functions respectively. These two symmetries make the magnitude and phase of the output coefficients to become even and odd functions respectively.

For our keypoint orientation estimate, the phase calculated in Eq. 3 could either account for the first positive or negative Fourier coefficients (i.e., $Z(1)$ or $Z(-1)$) as they are the same in magnitude but opposite in direction. Nevertheless, the intensity sample collection in Algorithm 1 samples image intensities in counter-clockwise direction and analyses a non-zero area. The first positive Fourier coefficient, $Z(1)$, is therefore chosen to estimate keypoint orientation as well as to define its direction. The use of this coefficient is also suggested by Folkers and Samet [21] to normalise the Fourier coefficients that undergo the change in starting point.

3. IMAGE KEYPOINT DESCRIPTION: RFA DESCRIPTOR

In principle, keypoint descriptor aims to represent a keypoint by extracting main information within a minimum sampling window that is centred at the keypoint. The extracted information is further analysed to outline its features while maintaining the discrimination among other keypoint descriptors. The information can be extracted either in spatial domain or frequency domain with the uses of statistical or spectral analysis. For example, SIFT descriptor computes the image gradient orientations within the local image patch of a keypoint in spatial domain. Then, a statistical analysis is applied as the gradient orientations are distributed into orientation histograms that yield a 128 dimensional vector for representing the keypoint. In this work, RFA descriptor obtains the Fourier coefficients of image gradients to determine the main components of local image patches through spectral analysis in frequency domain.

The RFA descriptor is generated with the following steps: (1) gradient sample collection; (2) Fourier transform; and (3) principle frequency extraction. As previously mentioned, RFA descriptors are generated by using SIFT keypoints that contain the details of position, scale and dominant gradient orientation. In particular, the keypoint orientation estimated by Fourier analysis replaces the dominant gradient orientation in SIFT keypoints as it is more reliable and accurate.

Firstly, the local image gradients around a keypoint are circularly sampled in a manner that is similar to the intensity sample collection for Fourier keypoint orientation estimate. Here, image gradients specifically mean both horizontal and vertical gradients. The gradient sample collection process is detailed in Algorithm 2. Through the algorithm, it can be seen that the keypoint orientation is taken into account to determine the position of gradient samples to ensure that image gradients are correctly sampled with the same starting point as its corresponding keypoint. Figure 2 displays the sampling pattern of this gradient sample collection with a radial sampling range of 5 pixel radius (i.e., $R_g = 5$) and an angular sampling interval of 10° (i.e., $\Delta\theta_g = 10^\circ$) for demonstration purpose. A slightly different sampling pattern is applied in gradient sample collection that the image gradients along angular direction are treated as the signatures at that sampling radius. It is shown by each black line that connects the red circles in Figure 2. As for the output of this step, 16 gradient vectors are obtained and they are express as,

$$s_r(n) = dx_r(n) + j dy_r(n) \tag{6}$$

where, sampled horizontal and vertical gradients are substituted into the real and imaginary parts of these complex vectors respectively. For Eq. 6, it has the variables $n = 1, 2, \dots, N$ and $r = 1, 2, \dots, R_d$, where $N = 360^\circ/\Delta\theta_d$ and $R_d = 16$. The radial sampling range, R_d , indicates the number of gradient vectors and is set to this value due to the suggestion given by the SIFT descriptor performance evaluation for local keypoint matching [22].

ALGORITHM 2: Gradient Sample Collection for RFA Descriptor
Input gray image, I
Input keypoint details, (x_0, y_0, θ)
Set radial sampling range, $R_d = 16$
Set angular sampling interval, $\Delta\theta_d = 5^\circ$
Set sampling angle index, $\theta_i = 360^\circ/\Delta\theta_d - 1$
For $r = 1$ to R_d do
Initialise s_r , to a zero complex vector of the length $R_d \times 360^\circ/\Delta\theta_d$
For $i = 0$ to θ_i do
Determine the sampling angle, $\theta_d = \Delta\theta_d \times i$
Determine the sampling position
$u = r \times \cos(\theta_d - \theta) + x_0$
$v = r \times \sin(\theta_d - \theta) + y_0$
Determine and sample image gradients at $I(u, v)$
Distribute the image gradients into s_r
End for
End for
Output s_r

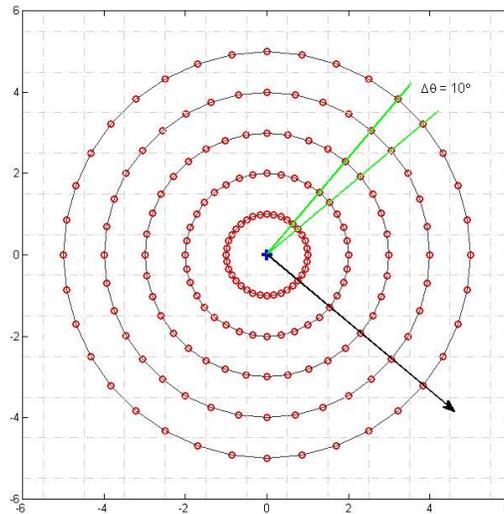


FIGURE 2: Gradient sample collection for RFA descriptor. The red circles indicate the positions of samples, the blue cross sign represents the location of keypoint and the black arrow directs keypoint orientation.

Secondly, prior to Fourier transform, unit vector normalisations are separately employed on the real and imaginary parts of each of the obtained gradient vectors, $s_r(n)$. The normalisations are used to increase gradient contrast while preserving the characteristics of the gradients. Then, a discrete Fourier transform is independently performed on each of the obtained gradient vectors to

observe the Fourier coefficients (i.e., frequency responses) at different radial distances from the keypoint. The transformations are defined as,

$$S_r(k) = \sum_{n=0}^{N-1} s_r(n) e^{-\frac{j2\pi nk}{N}} \quad (7)$$

for $k = -N/2 \dots N/2 - 1$. The variable $S_r(k)$ denotes the k^{th} Fourier coefficient of the r^{th} gradient vector that is based on the information given by the $s_r(n)$ in Eq. 6. Again, these Fourier coefficients have been shifted to centre and are the output of the second step.

The last step manipulates the Fourier coefficients, $S_r(k)$, in order to extract the principle frequencies in these coefficients. It translates the coefficients into more meaningful analysis and distinctive representation of the neighbourhood of keypoints. RFA descriptor is generated by extracting four coefficients with the indices of $k = -2, -1, 1, 2$ from each set of the obtained Fourier coefficients, $S_r(k)$. The descriptor, V , is mathematically expressed as,

$$V = \{\text{Re}(S_r(k)), \text{Im}(S_r(k))\} \quad (8)$$

for $r = 1, 2, \dots, R_g$ and $k = -2, -1, 1, 2$. Consequently, it results in a 128 dimensional vector since the real and imaginary parts of the extracted coefficients are treated as individual numbers. The vector is then normalised to unit length to suppress the influences induced by illumination changes. This normalised vector becomes the RFA descriptor.

It is essential to avoid or minimise the information loss associated with the generation of keypoint descriptor. For example, SIFT descriptor uses the gradient orientations weighted by their gradient magnitudes to analyse the local image gradients around keypoints. While the magnitude and orientation information of the gradients are preserved, this methodology also changes the characteristics of the gradients through the weighting operation. To prevent this situation, both horizontal and vertical gradients are taken into account for the gradient sample collection in the generation of RFA descriptor. Therefore, the information of gradients can be entirely preserved and used for Fourier analysis without altering its characteristics from the original.

During the manipulation of Fourier coefficients in the last step of RFA generation, all the centre coefficients (i.e., $S_r(0)$) have been ignored as these coefficients straddle in the line between positive and negative frequencies. Phrased differently, the coefficients contain the responses from both positive and negative frequencies that make them unstable and sensitive to geometric translation changes [23]. In addition, the same property is applied to the last Fourier coefficients (i.e., $S_r(-N/2)$).

Basically, the low frequency band of a spectrum represents the approximation of the spectrum itself, and the high frequency band stands for the detail of the spectrum. According to this principle, RFA descriptor analyses the low frequency band of local image gradients around keypoints by extracting and utilising the four coefficients as explained in the last step of RFA generation. Specifically, the extraction of the four coefficients acts like a low-pass filter convoluting with image gradients in spatial domain, but the extraction directly doing so in frequency domain. High frequency band is discarded because the details of local image gradients would be extremely dissimilar in the cases that input images are not exactly identical and the fact that images are smoothed by different Gaussian scales during the generation of RFA descriptor. Whereby, the descriptor extracts low frequency coefficients to describe the principle component in local image gradients.

Since the input of the Fourier transform in Eq. 7 is a complex function, the resulting Fourier coefficients do not obey any symmetrical properties as mentioned previously. To properly analyse low frequency band, for RFA descriptor, it is necessary to include the coefficients at both of the positive and negative frequencies. They represent for the same frequency response and contain the information at that frequency, but only in opposite direction or sign.

4. EXPERIMENT SETUP

This section describes the image database, matching techniques and evaluation metrics used to manifest the performances of RFA descriptor.

4.1 Image Database

The image database from a performance study of keypoint descriptors [14] was adopted to evaluate the performances of RFA descriptor with the keypoint orientation estimated by Fourier analysis. The image database includes a few sets of images that are individually affected by blur, illumination, JPG compression, rotation, and scale distortions. Thus, RFA descriptor's performances can be completely examined under these image distortions. Example images of this database are shown in Figure 3.

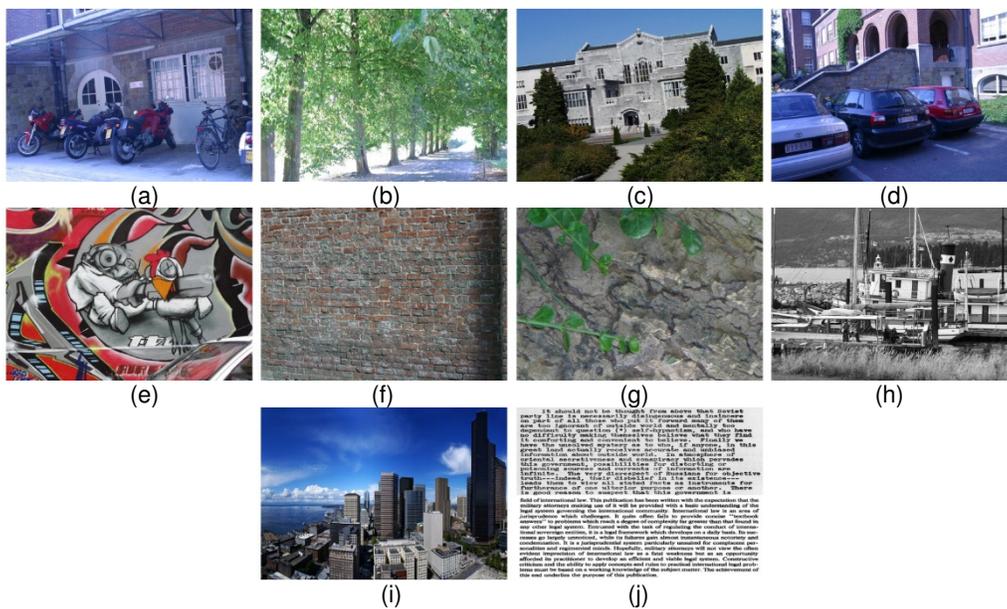


FIGURE 3: Example images of the database used for the evaluations of RFA descriptor. Blur changes: (a) Bikes and (b) Trees. JPG compressions: (c) UBC. Illumination changes: (d) Leuven. Viewpoint changes: (e) Graffiti and (f) Wall. Rotation and scale changes: (g) Bark and (h) Boat. Rotation changes: (i) Building and (j) Text.

4.2 Matching Technique

In order to obtain descriptor matches, it requires a similarity measure and a matching strategy to calculate the likeliness between descriptors and determine most likely descriptor matches among a pool of descriptors respectively. There are numerous similarity measures [24] [25] and matching strategies [26] [27] available for the matching of keypoint descriptors.

In this paper, however, the most familiar similarity measure and matching strategy are utilised to generate descriptor matches. They are Euclidean distance and threshold-based nearest neighbour (TNN) matching strategy respectively. As soon as the Euclidean distances between a descriptor and a pool of descriptors are calculated, TNN matching strategy would assess the descriptor pair that has the shortest distance between them. If the distance of that pair is below a pre-defined threshold, the pair is then recognised as a match, otherwise, no match is generated.

4.3 Evaluation Metrics

The RFA descriptor matches generated by Euclidean distance with TNN matching strategy are used to measure the performances of RFA descriptor. The matches are evaluated by recall-precision, which was initially employed in [28] to verify the performances of keypoint descriptors. It is the most popular performance metrics in the evaluation of keypoint descriptors because the major examinations on descriptor matches are their sensitivity (i.e., matching rate) and accuracy (i.e., matching accuracy). Recall represents the sensitivity of the generated descriptor matches in a given image pair that contains the same scene. It is defined as the ratio between the number of correct matches and the number of correspondences, such that,

$$\text{recall} = \frac{\# \text{ correct matches}}{\# \text{ correspondences}} \quad (9)$$

where, the correct matches and correspondences are verified by projecting keypoints from its image to the corresponding image. If the projected keypoints are equal as or close to their corresponding keypoints, then they are recognised as correct matches or correspondences. Another metrics in recall-precision is 1-precision that indicates the accuracy of generated matches and is the ratio between the number of false matches and the number of total matches. It is expressed as,

$$1 - \text{precision} = \frac{\# \text{ false matches}}{\# \text{ correct matches} + \# \text{ false matches}} \quad (10)$$

5. EXPERIMENTAL RESULTS

This section presents the performance evaluations of RFA descriptor under various image distortions. Apart from the evaluations of RFA descriptor, the performances of other benchmark descriptors, such as SIFT, VLFeat SIFT and PCA-SIFT, are also provided for comparison purposes. Since PCA-SIFT requires a training process, it was trained by 26,000 image patches that are not related to the image database before it is used in the following evaluations. Twenty and 128 dimensional PCA-SIFT descriptors are illustrated in the evaluations for the completeness of the comparisons. The former is recommended by its author and the latter is to fairly compare with the other descriptors as they have dimensions of 128. Significant observations and discussions about the evaluations are also given in this section.

The descriptors were assessed while the images of Figure 3(a) and Figure 3(b) were blurred at different degrees. The performance results obtained by these blurred images demonstrate the abilities of the descriptors against blur changes, and they are presented in Figure 4. From the recall-precision graphs, it can be seen that RFA descriptor particularly outperforms the other keypoint descriptors in high precision region, between 0 and 0.5 of 1-precision. Since the smoothing effect caused by blurring occurs at high frequency band, the RFA descriptor is benefited from using Fourier transform and low-pass filter that remove the high frequency band of local image gradients around keypoints. Thus, the performances of RFA descriptor is hardly degraded by the image distortions induced by blur changes.

The quality of Figure 3(c) is degraded by JPG compression, which is based on discrete cosine transform to trade off image quality for a smaller file size. Figure 5 displays the performance results of the descriptors under this image quality degradation. As expected, RFA descriptor demonstrates a better performance than the other descriptors in this case since discrete cosine transform is a variant of Fourier transform. Both JPG compression and RFA descriptor process their input image in frequency domain and manipulate information with frequency decomposition. Whereby, the remaining information after JPG compression can be easily extracted and used with RFA descriptor.

The performance results from altering the illuminations of Figure 3(d) are depicted in Figure 6. Once more, it is observed that RFA descriptor surpasses the other descriptors in high precision region. The high recall at high precision region indicates the representations of image gradients by RFA descriptor are distinct and accurate under this type of image distortion. It is advantaged by the unit vector normalisation prior to Eq. 7 as the normalisation minimises the effect of illumination changes by locally standardising the magnitudes of horizontal and vertical gradients. Another contributor for this high performance is the linearity of Fourier transform that allows the scaling of an input signal to be factorised, then it is normalised by the unit vector normalisation in the last step of RFA descriptor generation.

Figure 7 shows the performances of the descriptors against the image distortions caused by viewpoint changes. As Figure 3(e) and Figure 3(f) were captured at different viewpoints, sophisticated geometrical transforms are imposed onto these images, increasing the possibility of generating incorrect matches. In Figure 7(a), the performance of RFA descriptor is extensively higher than the other descriptors in high precision regions. Although the other descriptors have higher recall in low precision region, it should be noted that their recall approximately starts to rise at the 1-precision of 0.7 and it can be said that the most of their matches are incorrect. With Figure 7(b), however, RFA descriptor only demonstrates an acceptable performance, which is in-between SIFT and VLFeat SIFT descriptors, as serious image deformations have been imposed in this dataset.

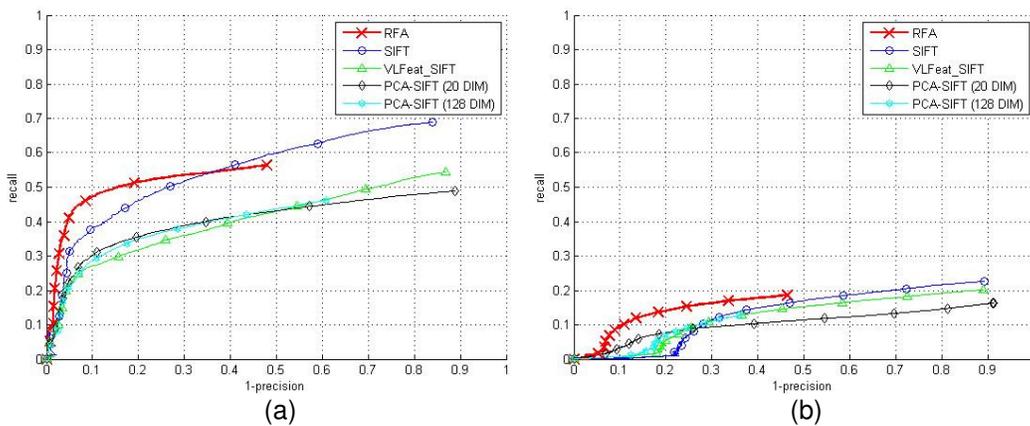


FIGURE 4: Recall-Precision evaluations for blur changes using: (a) Bikes and (b) Trees.

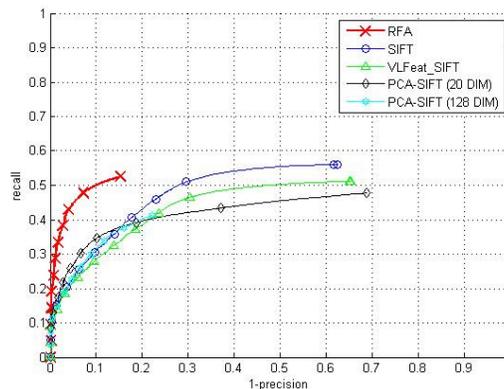


FIGURE 5: Recall-Precision evaluations for JPG compression using UBC.

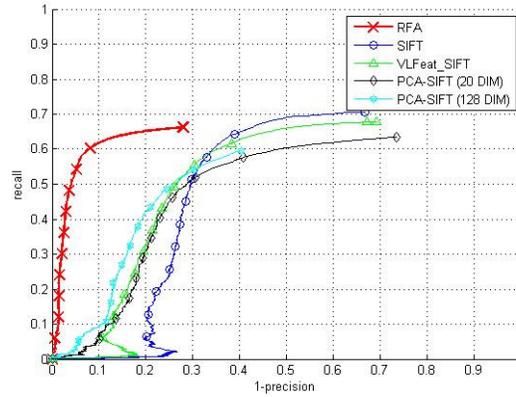


FIGURE 6: Recall-Precision evaluation for illumination changes using Leuven.

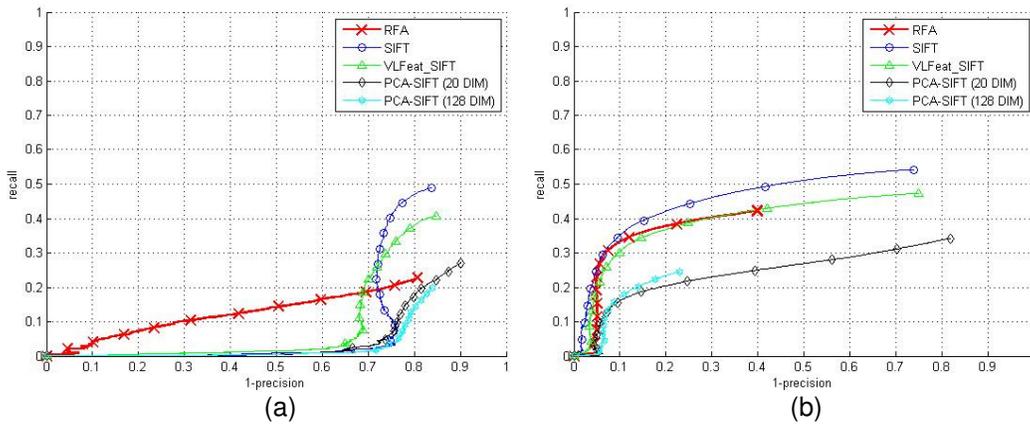


FIGURE 7: Recall-Precision evaluations for viewpoint changes using: (a) Graffiti and (b) Wall.

Figure 8(a) and Figure 8(b) present the performance results of the descriptors when the images of Figure 3(g) and Figure 3(h) simultaneously experienced the image distortions caused by rotation and scale changes. From Figure 8(a), RFA descriptor has achieved a performance that is much lower than SIFT and VLFeat SIFT descriptors. In Figure 8(b), RFA descriptor has a higher recall in high precision region when compared with the other descriptors. It is an acceptable performance, especially if accuracy is more important than sensitivity in keypoint matches.

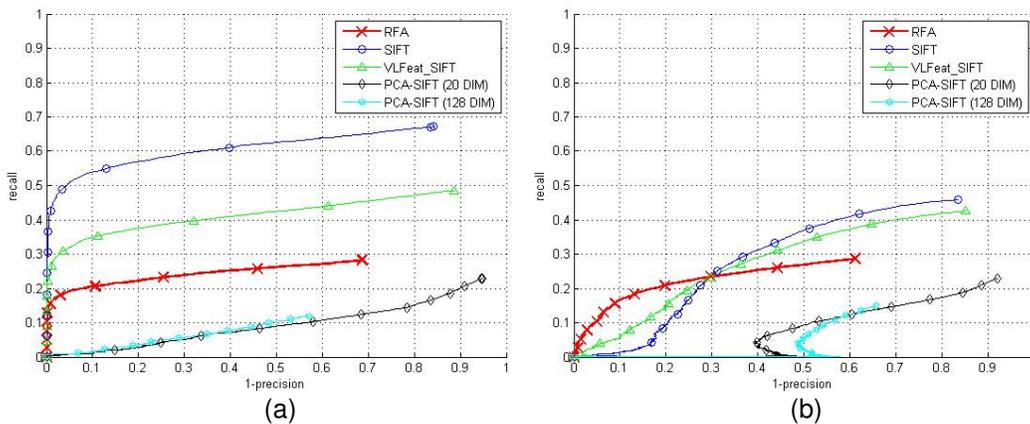


FIGURE 8: Recall-Precision evaluations for rotation and scale changes using: (a) Bark and (b) Boat.

Apart from the performance results given in Figure 7 and Figure 8, Figure 9 shows the performances of the descriptors when they were subjected under the influences of purely rotation changes. From both the recall-precision graphs, RFA descriptor demonstrates equivalent performances as VLFeat SIFT descriptor, but it outperforms SIFT and PCA-SIFT descriptors by significant margins. It is a solid evidence that our keypoint orientation estimate provides a reliable and robust reference orientation and allows the local image patches of descriptor matches to be rotated to a similar orientation. With the assistance of this keypoint orientation estimate, RFA descriptor has been able to easily deal with the image distortion induced by rotation changes.

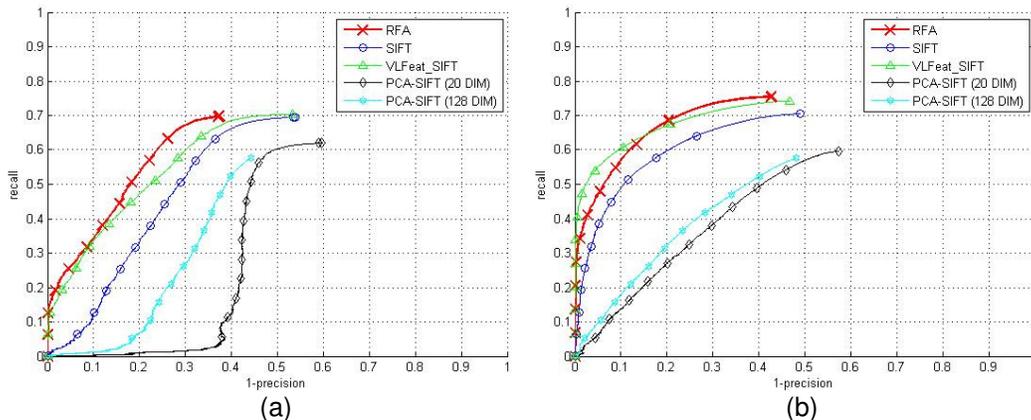


FIGURE 9: Recall-Precision evaluations for rotation changes using: (a) Building and (b) Text.

6. CONCLUSION

In this paper, a keypoint orientation estimate and a keypoint descriptor have been proposed to analyse the local image patch around keypoints in frequency domain. The keypoint orientation is estimated from the information of image intensities with the starting point normalisation of Fourier coefficients. For the keypoint descriptor, it extracts low order Fourier coefficients of image gradients to approximate overall appearance of the gradients. Meanwhile, image detail that resides in high order Fourier coefficients are ignored as excessive detail would make the descriptor superfluously distinctive.

When evaluating the performances of RFA descriptor, it was found that the descriptor outperforms benchmark descriptors, such as SIFT, PCA-SIFT and VLFeat SIFT descriptors, in images suffering from a variety of image distortions. RFA descriptor is particularly stable and reliable in dealing with the image distortions caused by blur, rotation, illumination and JPG compression changes. This is evidenced from the experimental results presented in this paper. In addition, RFA descriptor encodes image gradients into a distinctive and accurate representation as its matches always yield a recall that rapidly rises up in high precision region, indicating high sensitivity and accuracy during matching.

Through this paper, it has been proven that spectral analysis, such as Fourier transform, Fourier normalisation and low-pass filtering, can be employed to estimate keypoint orientations and to describe local keypoints. This research shows that spectral analysis serves well as an alternative or supplementary tool for statistical analysis of keypoint description and orientation estimation.

7. REFERENCES

- [1] J. Li, N. Allinson, D. Tao, and X. Li, "Multitraining support vector machine for image retrieval," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3597–3601, 2006.
- [2] J. Bauer, H. Bischof, A. Klaus, and K. Karner, "Robust and fully automated image registration using invariant features," in *Proceedings of International Society for the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2004, pp. 1682–1777.

- [3] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [4] G. Dorko and C. Schmid, "Selection of scale-invariant parts for object class recognition," in *Proceedings of International Conference on Computer Vision*, 2003, pp. 634–639.
- [5] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using affine-invariant regions," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2003, pp. 319–324.
- [6] M. Yang, L. Zhang, S. K. Shiu, and D. Zhang, "Monogenic binary coding: An efficient local feature extraction approach to face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1738–1751, 2012.
- [7] Y. Fan and M. H. Meng, "3d reconstruction of the wce images by affine sift method," in *Proceedings of World Congress on Intelligent Control and Automation*, 2011, pp. 943–947.
- [8] S. Muramatsu, D. Chugo, S. Jia, and K. Takase, "Localization for indoor service robot by using local-features of image," in *Proceedings of ICCAS-SICE*, 2009, pp. 3251–3254.
- [9] A. Sedaghat, M. Mokhtarzade, and H. Ebadi, "Uniform robust scaleinvariant feature matching for optical remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4516–4527, 2011.
- [10] X. M. Mu, J. P. Liu, W. H. Gui, Z. H. Tang, C. H. Yang, and J. Q. Li, "Machine vision based flotation froth mobility analysis," in *Proceedings of Chinese Control Conference*, 2010, pp. 3012–3017.
- [11] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.
- [12] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *Proceedings of IEEE International Conference on Computer Vision*, 2011, pp. 2564–2571.
- [13] S. Taylor and T. Drummond, "Binary histogrammed intensity patches for efficient and robust matching," *International Journal of Computer Vision*, vol. 94, no. 2, pp. 241–265, 2011.
- [14] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [15] Y. Ke and R. Sukthankar, "Pca-sift: a more distinctive representation for local image descriptors," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2004, pp. 506–513.
- [16] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [17] S. Leutenegger, M. Chli, and R. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *Proceedings of IEEE International Conference on Computer Vision*, 2011, pp. 2548–2555.
- [18] J. Li and N. M. Allinson, "A comprehensive review of current local features for computer vision," *Neurocomputing*, vol. 71, no. 10, pp. 1771–1787, 2008.

- [19] L. Juan and O. Gwun, "A comparison of sift, pca-sift and surf," *International Journal of Image Processing*, vol. 3, no. 4, pp. 143–152, 2009.
- [20] J. L. Crowley and A. C. Parker, "A representation for shape based on peaks and ridges in the difference of low-pass transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 2, pp. 156–170, 1984.
- [21] A. Folkers and H. Samet, "Content-based image retrieval using fourier descriptors on a logo database," in *Proceedings of International Conference on Pattern Recognition*, vol. 3, 2002, pp. 521–524.
- [22] S. Lin, C. Wong, T. Ren, and N. Kwok, "The impact of information volume on sift descriptor," in *Proceedings of International Conference on Wavelet Analysis and Pattern Recognition*, 2013, pp. 287–293.
- [23] A. K. Jain, *Fundamentals of digital image processing*. Prentice-Hall, Inc., 1989.
- [24] A. Strehl, J. Ghosh, and R. Mooney, "Impact of similarity measures on web-page clustering," in *Proceedings of Workshop on Artificial Intelligence for Web Search*, 2000, pp. 58–64.
- [25] A. Huang, "Similarity measures for text document clustering," in *Proceedings of New Zealand Computer Science Research Student Conference*, 2008, pp. 49–56.
- [26] O. Chum, J. Matas, and S. Obdrzalek, "Enhancing ransac by generalized model optimization," in *Proceedings of Asian Conference on Computer Vision*, vol. 2, 2004, pp. 812–817.
- [27] M. Cho and H. Park, "A robust keypoints matching strategy for sift: An application to face recognition," in *Neural Information Processing*, ser. Lecture Notes in Computer Science, 2009, vol. 5863, pp. 716–723.