Robust Edge Aware Descriptor for Image Matching

Rouzbeh Maani, Sanjay Kalra, Yee-Hong Yang

Department of Computing Science, University of Alberta, Edmonton, Canada

Abstract. This paper presents a method called Robust Edge Aware Descriptor (READ) to compute local gradient information. The proposed method measures the similarity of the underlying structure to an edge using the 1D Fourier transform on a set of points located on a circle around a pixel. It is shown that the magnitude and the phase of READ can well represent the magnitude and orientation of the local gradients and present robustness to imaging effects and artifacts. In addition, the proposed method can be efficiently implemented by kernels. Next, we define a robust region descriptor for image matching using the READ gradient operator. The presented descriptor uses a novel approach to define support regions by rotation and anisotropical scaling of the original regions. The experimental results on the Oxford dataset and on additional datasets with more challenging imaging effects such as motion blur and non-uniform illumination changes show the superiority and robustness of the proposed descriptor to the state-of-the-art descriptors.

1 Introduction

Local feature detection and description are among the most important tasks in computer vision. The extracted descriptors are used in a variety of applications such as object recognition and image matching, motion tracking, facial expression recognition, and human action recognition. The whole process can be divided into two major tasks: region detection, and region description. The goal of the first task is to detect regions that are invariant to a class of transformations such as rotation, change of scale, and change of viewpoint. The detected regions are then described by a feature vector. An ideal region descriptor should not only be invariant to geometric transformations but also be robust to imaging effects such as blurriness, noise, distortions, and illumination changes [1].

This paper focuses on the second task, region description. Local gradients are commonly used by the state-of-the-art methods. Although local gradient information is very powerful, it is susceptible to imaging effects such as noise, illumination changes, and blurriness. The first contribution of this paper is a novel method to compute local edge information. The proposed READ method is similar to local gradients but is computed in a novel way. It provides both magnitude and orientational information at each pixel, and is robust to imaging effects. Similar to using gradients, the proposed READ method can be used in basic image processing operations such as edge detection, segmentation, and

texture analysis. In particular, due to its robustness, READ is useful in conditions where undesirable imaging effects are unavoidable (e.g., noise in magnetic resonance images or blurriness in underwater imaging). The second contribution of this paper is the kernel implementation of the proposed READ method. The presented implementation makes the method fast for practical use. The third contribution of this paper is a novel method for determining support regions around the regions detected by affine detectors. First, it is shown that in theory the support regions can be scaled (with different scaling factors along the eigenvectors of the elliptical affine region) and rotated. Then, through experiments it is demonstrated that this method of support region definition can improve the results compared to the simple method of isotropic scaling of the original regions detected by affine detectors which is suggested in [2]. The advantages of the new descriptor are first explored in common geometric transformations, and then in more challenging imaging effects such as motion blur, non-uniform illumination changes with moving shadows, and images with different levels and types of noise. Although they are very important, these challenging effects are less noted in the evaluation of the previous descriptors.

2 Related Works

There are two main steps in finding corresponding points in two images. In the first step, interest points (regions) are found in the images. Ideal points should be highly discriminative and robustly detectable under different imaging conditions and geometric transformations. Some examples of detection methods include Difference of Gaussian (DoG) [3], Harris-Affine [4], and Hessian-Affine [5]. A review and comparison of region detection methods can be found in [5, 6].

Many detectors provide circular or elliptical regions with different sizes around the detected points for point description. The size of the detected region is determined by the detected scale of the region. By transforming the detected regions (elliptical and circular) to a circular region of a fixed radius, the regions are normalized into a canonical form. As a result, an affine transformation is reduced to a rotation, and an affine invariance on the original image can be obtained by rotation invariance on the canonical region [5]. Hence, region descriptors usually define rotation invariant features to provide descriptors that are invariant to local affine geometric transformations.

One of the most popular descriptors is the Scale Invariant Feature Transform (SIFT) [3]. The main information used in SIFT is the magnitude and orientation of local gradients accumulated in subregions. SIFT is later extended in the Gradient Location and Orientation Histogram (GLOH) method [1]. Mikolajczyk and Schmid [1] demonstrate that SIFT and GLOH outperform other descriptors obtained using shape context, steerable filters, spin images, differential invariants, complex filters, and moment invariants. Some other descriptors include the Center-Symmetric Local Binary Pattern (CS-LBP) [7], the shape of MSER [8], the Local Intensity Order Pattern (LIOP) [9], and KAZE [10].

DAISY is a successful method recently proposed by Tola et al. [11]. Similar to SIFT, DAISY uses the magnitude and orientation of local gradients; however, the weighted sum of gradient orientation is replaced by the convolution of the gradient in specific directions with several Gaussian filters. Recently, it has been shown that the intensity ordinal information is more useful than the fixed location bins used by many descriptors such as SIFT and DAISY. The idea has been used by several descriptors such as LIOP [9], MROGH, and MRRID [12, 2].

A new promising approach is in developing "binary" descriptors such as BRIEF [13], Brisk [14], ORB [15], Freak [16], and BinBoost [17] for real-time applications. A comparative evaluation of these descriptors is presented in [18]. The recent paper by Miksik and Mikolajczyk [19] also compares some of these methods in the accuracy and speed trade-offs suggesting that binary descriptors provide comparable precision/recall results with SIFT and outperform in speed. On the other hand this paper reports that LIOP, MRRID, MROGH are slower but outperform SIFT and other binary descriptors.

3 Robust Edge Aware Descriptor

Considering neighbors on a circle (or multiple circles) around a pixel is a popular approach in rotation invariant methods. The values of the circular neighbors are usually encoded in two ways: 1) by applying a threshold (e.g., the center pixel's intensity) similar to the LBP [20] and its variants, 2) by transforming the values into frequency components as suggested by some texture classification methods [21–25]. It is argued that applying a threshold by the LBP-based methods compromises some important information and demonstrated that the latter approach outperforms the first one [24]. The 1D Discrete Fourier Transform (DFT) of the neighbor's intensity values is defined as

$$F(n) = \sum_{k=1}^{P} f(k) e^{\frac{-2\pi i(k-1)(n-1)}{P}}, (n = 1, ..., P), \tag{1}$$

where $F(n)$ consists of P complex numbers, known as the frequency components. Since $f(k)$ consists of real numbers, the frequency components of $F(n)$ are conjugate symmetric about the DC component. That is, the 2^{nd} and the P^{th} components (similarly the 3^{rd} and the $(P-1)^{th}$, and so on) have the same magnitude but opposite phase. A recent work [23] shows that the low frequency components ($F(1)$, $F(2)$, and $F(P)$) comprise more than 90% of the $f(k)$ signal in some well known texture datasets, and therefore can well represent the texture around a pixel. Inspired by this work, we demonstrate that the *second frequency component* (or equivalently the P^{th} component) can be used to robustly compute the edge information. To better explain the edge detection ability of $F(2)$, we need to show the characteristics of function $f(k)$ when it is around an edge. We define an edge as the line separating a dark region from a bright region as shown in Figure 1(a). In this example, we traverse the circular samples in the clockwise direction. We start from a dark region, cross the edge and go into the

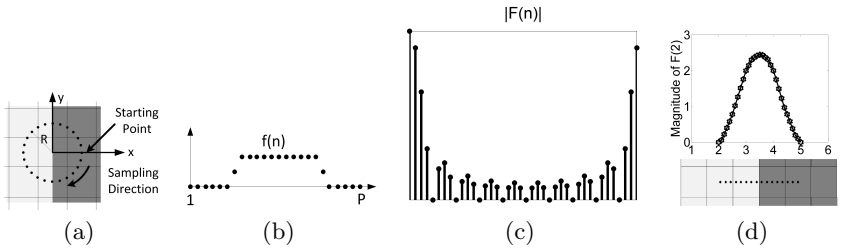


Fig. 1. Using READ operator to compute local edge. a) P sample are located on radius of R around a pixel. b) The function of samples have a rectangular shape. c) The DFT of a rectangular shape function is a sinc. The highest values of the sinc function are at $n = 1$ and $n = \{2, P\}$. d) $|F(2)|$ as a function of the distance of the center of the sampling circle from the edge.

bright region, cross the edge again, and come back to the starting point in the dark region. Using this circular sampling method, the function of an edge can be characterized as a rectangular shape function as shown in Figure 1(b). For simplicity, assume that the rectangular function has a value of one in the bright region. The DFT of the rectangular shape function with width M (using Eq. 1) is a sinc shape function of the following form:

$$F(n)_{rect} = \frac{sin(\frac{\pi(n-1)M}{P})}{sin(\frac{\pi(n-1)}{P})} \times e^{-\frac{i\pi(n-1)(M-1)}{P}}. \tag{2}$$

It can be easily observed that the magnitude of this sinc shape function, $|\frac{sin(\pi(n-1)M/P)}{sin(\pi(n-1)/P)}|$, has the maximum value at $n = 1$ and then at $n = \{2, P\}$ (Figure 1(c)). As a result, an edge manifests itself with maximum values in $F(1)$, $F(2)$, and $F(P)$. It is noteworthy that among these three components, $F(1)$ represents the average intensity (known as DC component), which gives information if the pixel is located in a dark or bright region, while $F(2)$ (or $F(P)$) is more sensitive to the actual edge information around the pixel. Another possible interpretation of $F(2)$ is that it approximates a rectangular shape function better than the other components. To better demonstrate the edge detection ability we plot $|F(2)|$ as a function of the distance of the center of the circle from the edge for the given example shown in Figure 1(d). In this example, we choose points between pixels 2 and 5 with an increment of 0.1 pixels. The magnitude of $F(2)$ is shown for each point in Figure 1(d). As one can see, $|F(2)|$ reaches its maximum value at location 3.5 which is the exact point separating the dark region from the bright region. Now, we formally define READ by setting $n = 2$ in Eq. 1:

$$READ = \sum_{k=1}^{P} f(k) e^{\frac{-2\pi i(k-1)}{P}}. \tag{3}$$

Eq. 3 gives a complex number; hence, it can be further decomposed into real and imaginary parts:

$$Re(READ) = \sum_{k=1}^{P} f(k)cos(\frac{2\pi(k-1)}{P}), \tag{4}$$

$$Im(READ) = -\sum_{k=1}^{P} f(k)sin(\frac{2\pi(k-1)}{P}). \tag{5}$$

The magnitude and the phase of READ can be simply computed from the real and imaginary parts. The magnitude of READ represents the amount of rectangular shape function (i.e., the strength of an edge), while the phase indicates the starting location of the rectangular shape function (i.e., the edge orientation). One may note that the exact value of the phase depends on the neighbor ordering strategy. We start from the x axis and traverse the neighbors in the clockwise direction as shown in Figure 1(a). Using this protocol will result in the same orientation value computed by the conventional gradient orientation formula.

The rectangular shape function of $f(n)$ (which characterizes an edge) is comparable to the *uniform* patterns in the LBP method which represent edges of varying positive and negative curvatures [20]. However, the uniform patterns are acquired by applying a threshold which makes the patterns sensitive to noise, while in READ, an edge appears as a low frequency component ($F(2)$) which is less affected by noise. Similar to the LBP, the READ can be acquired using different R and P. Figure 2 compares the gradient calculation of the READ operator with that of the central difference ($\Delta_h f(x) = f(x+h/2) - f(x-h/2)$) on a synthetic and a real image. READ is computed with setting ($P = 8, R = 1$) and $h = 1$ pixel in the central difference method. To make a fair visual comparison, the magnitude of the gradients are normalized in the range [0 1].

As one can observe the magnitude of the READ is zero in flat regions, while it is maximized on pixels located on an edge. However, this change is gradual in contrast to the central difference (this was also demonstrated in Figure 1(d)). The phase of READ faithfully represents the orientation of the local edge. One may consider the orientational values around the circle in the synthetic image. The color map in the bottom right of the phase image shows the color representing a given angle. Similarly the change in phase is smoother than that of the central difference. The second row of Figure 2 compares edge characterization on a real image. It can be observed that the outputs of the two methods look very similar both in magnitude of the detected edges and in their orientations. However, one may note that the READ is more sensitive to finding faint edges. For example, the edges in the background on the upper left side and above the hat are more clear in the magnitude of READ compared to that of the local gradients.

3.1 Properties of READ

READ has several advantages that make it favorable for computer vision and image processing applications. The first advantage is its robustness to noise. The

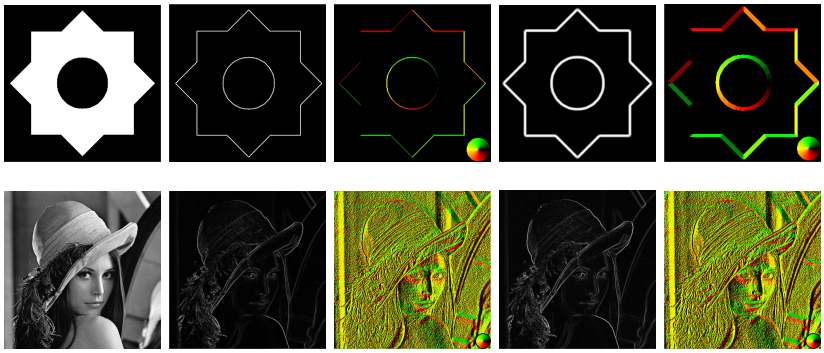


Fig. 2. Computing gradient on a synthetic (first row) and a real image (second row). The first column is the original image. The 2^{nd} and 3^{rd} columns are the magnitude and orientation of gradient computed by the central difference. The 4^{th} and 5^{th} columns are the magnitude and orientation of gradient computed by READ.

reason is that noise appears in high frequency components. However, READ is defined based on a low frequency component, $F(2)$, which is not sensitive to noise. This property will be further demonstrated in Section 4. The next property of READ is its invariance to linear changes of illumination. One can simply observe that any linear change of illumination linearly changes the magnitude of function $f(k)$. This effect, however, does not change the phase of READ. To keep the magnitude of READ invariant to linear illumination changes, the image intensity is normalized to have zero mean and unit standard deviation. Finally, READ is robust to blur effect. The reason is that the blur effect mainly dampens the high frequency components and the low frequencies are less affected. These properties are further explored in Section 4.

3.2 Efficient Implementation

Speed is one important factor for descriptors. As a result a good descriptor should also have a reasonable runtime. In this regard, we present an efficient implementation for the READ operator. The idea is to define kernels to represent Eqs. 4 and 5 and finding the gradients by convolution of the defined kernels with the image. For a radius of R, the kernel has a size of $N \times N$ where $N = 2R + 1$. There are two factors to compute in the kernel as defined in Eqs. 4, 5: the value of the samples, $f(k)$, and the cos/sin coefficients. Since, the location of the samples are known for a specific R and P, both factors could be found easily.

The value of each sample is found using bilinear interpolation from its four nearest neighbors if it is not exactly located on a pixel. To compute the value of $f(k)$ we consider a $P \times N^2$ matrix. The k^{th} row of the matrix represents the k^{th} sample and each column represents the weight of each element in the kernel for the bilinear inertpolation (we order the elements of the kernel in a $1 \times N^2$ row vector). We call this matrix B representing the bilinear weights of the kernel.

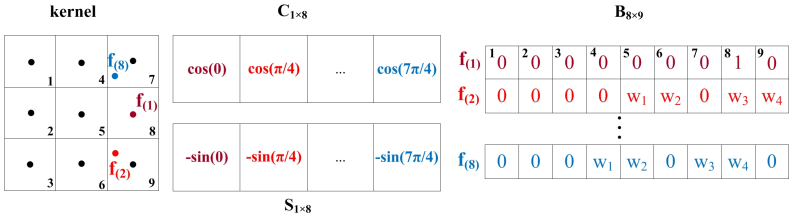


Fig. 3. Construction of matrix B, C, and S for $R = 1$ and $P = 8$ setting. The kernel size is 3×3. Samples are either located on a pixel ($f(1)$) or between pixels ($f(2)$,$f(8)$). In the latter case, we need bilinear interpolation to compute the value of the sample.

Now, we consider two $1 \times P$ row vectors called C and S to represent the cos and $-sin$ weights in Eqs. 4 and 5. Figure 3 illustrates the construction of B, C, and S for $R = 1$, $P = 8$ setting. The two kernels representing the equations can be simply computed by matrix multiplication $G_x = CB$ and $G_y = SB$. To use the kernel in the convolution operation we reshape G_x and G_y from $1 \times N^2$ to $N \times N$ matrices and reflect the values around the center of the matrix.

3.3 Region Descriptor

In this section we present a region descriptor for an arbitrary affine region detector. The orientation and magnitude of the underlying structure can be found by READ. However, although the magnitude of READ is rotation invariant, the phase of READ changes by rotation. Assume that the phase of READ at an arbitrary point x is α. When a rotation by θ° occurs, the phase will change to $\alpha' = \alpha + \theta$. Rotation invariance can be obtained by decomposing the phase of READ (i.e., α, α') into two components: a constant part related to the underlying structure (β) and a variable part related to the the location of the point (γ, γ') as shown in Figure 4. This approach is similar to the local rotation invariant coordinate system used by some descriptors such as MROGH [12], RIFT [26], and RIFF [27]. Instead of considering a new coordinate system, this can be easily done in READ by just subtracting the angle of the location of the point from the phase of READ computed in the regular coordinate system (i.e., $\beta = \gamma - \alpha = \gamma' - \alpha'$) which makes the computation fast.

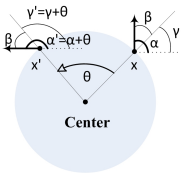


Fig. 4. Rotation invariant phase information.

By considering d orientational bins centering at ori_i, $(1 \leq i \leq d)$, the phase of READ is linearly assigned to the two closest orientational bins:

$$ori_i = (2\pi/d) \times (i - 1). \tag{6}$$

$$Bin_i(x,y) = \begin{cases} \frac{(2\pi/d)-|ori_i - \angle READ(x,y)|}{(2\pi/d)} & \text{, if } |ori_i - \angle READ(x,y)| < 2\pi/d \\ 0 & \text{, otherwise.} \end{cases} \tag{7}$$

The intensity ordinal information is used to form subregions. First, the pixels are sorted in a non-descending order of their intensity values, $X_1, ..., X_n$. Then, the ordered pixels, X_i, are divided into k partitions,

$$Pr(p) = \{X_i | X_{\lceil n(p-1)/k+1 \rceil} \leq X_i \leq X_{\lceil np/k \rceil}\}, \tag{8}$$

where $\lceil\ \rceil$ denotes the ceiling operator. The orientational histograms in each partition, $Pr(p)$, is accumulated and weighted with the average magnitude of the READ in that partition,

$$Hist(p,i) = \sum_{\forall (x,y) \in Pr(p)} Bin_i(x,y).\mu_{READ}(p), \tag{9}$$

where $Bin_i(x,y)$ is computed by Eq. 7 and $\mu_{READ}(p)$ by

$$\mu_{READ}(p) = \frac{1}{|Pr(p)|} \sum_{\forall (x,y) \in Pr(p)} |READ(x,y)|, \tag{10}$$

where $|Pr(p)|$ denotes the number of pixels in partition p. The final descriptor is a $d \times k$ feature vector constructed by concatenating the orientational histograms in all subregions.

Some descriptors (e.g., MROGH, MRRID) use support regions defined as the scaled version of the original detected region to improve their performance. Here, a novel and more flexible support region definition is presented. We suggest that the support region can be obtained by rotating and scaling with different scaling factors along the eigenvectors of the elliptical affine region (anisotropic scaling).

Using vector notation, a point X_L in an ellipse satisfies $X_L^T M_L X_L = 0$ in the homogeneous representation, where M_L is a symmetric matrix. As shown by Mikolajczyk and Schmid [4], when two elliptical regions $X_L^T M_L X_L = 0$ and $X_R^T M_R X_R = 0$ are corresponding, their canonical regions, $X_{Lc} = M_L^{1/2} X_L$ and $X_{Rc} = M_R^{1/2} X_R$, are related by a rotation:

$$\begin{aligned} X_{Rc} =& R(\alpha) X_{Lc} \\ \Rightarrow X_R =& M_R^{-1/2} R(\alpha) M_L^{1/2} X_L. \end{aligned} \tag{11}$$

Since M_L (and similarly M_R) is a symmetric matrix, it can be decomposed as $M_L = \Sigma_L \Lambda_L \Sigma_L^T$, where Σ_L is the orthogonal eigenvector matrix, and Λ_L the diagonal eigenvalue matrix. We define the transformation H for the scale matrix $S = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix}$:

$$H = \Sigma_L S^{-1} \Sigma_L^T, \tag{12}$$

Lemma 1. *Transformation H maps $X_L^T M_L X_L = 0$ into a new ellipse, the eigenvectors of which are the same as the old ellipse but the eigenvalues are scaled by the $(s_1)^2$ and $(s_2)^2$ factors.*

Proof. If the ellipse $X_L^T M_L X_L = 0$ undergoes the H transformation, the new ellipse is defined as $X_L'^T M_L' X_L' = 0$, where $X_L' = H X_L$ and $M_L' = H^{-T} M_L H^{-1}$. Substituting $M_L = \Sigma_L \Lambda_L \Sigma_L^T$ in the M_L' equation will results in:

$$M_L' = (\Sigma_L S^{-1} \Sigma_L^T)^{-T} \Sigma_L \Lambda_L \Sigma_L^T (\Sigma_L S^{-1} \Sigma_L^T)^{-1} \tag{13}$$

After a few steps of reduction this equation results in $M_L' = \Sigma_L S \Lambda_L S^T \Sigma_L^T$. Considering the new eigenvalue matrix, $\Lambda_L' = S \Lambda_L S^T$, results in $M_L' = \Sigma_L \Lambda_L' \Sigma_L^T$ which is claimed in Lemma 1.

□

Before presenting the theorem we present the following equations (used in the proofs). Assume that S and D are diagonal and Q is orthogonal, then it is easy to show:

$$(QDQ^T)^{1/2} = QD^{1/2}Q^T = QD^{1/2}Q^{-1}, \text{and} \tag{14}$$

$$(QSDS^TQ^T)^{1/2} = QD^{1/2}S^TQ^T = QD^{1/2}SQ^T. \tag{15}$$

Theorem 1. *Assume that the original ellipses defined by M_L and M_R undergo the H_L and H_R transformations, $X_L' = R\Sigma_L S^{-1} \Sigma_L^T X_L$, and $X_R' = R\Sigma_R S^{-1} \Sigma_R^T X_R$, where R is an arbitrary rotation matrix. The canonical regions of X_L' and X_R' are related by a rotation.*

Proof. We need to show that Eq. 11 holds for X_L' and X_R' with the new elliptical regions defined by M_L' and M_R'. We start by multiplying Eq. 11 with H_R

$$\begin{aligned}
H_R X_R &= H_R M_R^{-1/2} R(\alpha) M_L^{1/2} X_L \\
&= H_R M_R^{-1/2} R(\alpha) M_L^{1/2} H_L^{-1} H_L X_L \\
&= (R\Sigma_R S^{-1} \Sigma_R^T)(\Sigma_R \Lambda_R^{-1/2} \Sigma_R^T) R(\alpha) (\Sigma_L \Lambda_L^{1/2} \Sigma_L^T)(\Sigma_L^{-T} S \Sigma_L^{-1} R^{-1}) H_L X_L \\
&= (R\Sigma_R S^{-1} \Lambda_R^{-1/2} \Sigma_R^T R^{-1}) R R(\alpha) R^{-1} (R\Sigma_L \Lambda_L^{1/2} S \Sigma_L^{-1} R^{-1}) H_L X_L \\
&= (H_R^{-T} M_R H_R^{-1})^{-1/2} R R(\alpha) R^{-1} (H_L^{-T} M_L H_L^{-1})^{1/2} H_L X_L. \\
\Rightarrow X_R' &= M_R'^{-1/2} R(\gamma) M_L'^{1/2} X_L'.
\end{aligned} \tag{16}$$

□

Figure 5 illustrates the concept. The original regions detected by an affine detector (yellow) are related by rotation. The red regions are anisotropically scaled and rotated version of the original regions detected by affine detectors (yellow). It can be observed that the red regions are related by rotation as well. This idea can be considered as a generalized form of support regions suggested in MROGH in which R in the H transformation is an identity matrix and $s_1 = s_2$ in the S matrix. Nonetheless, this generalized form gives more flexibility to choose the support regions.

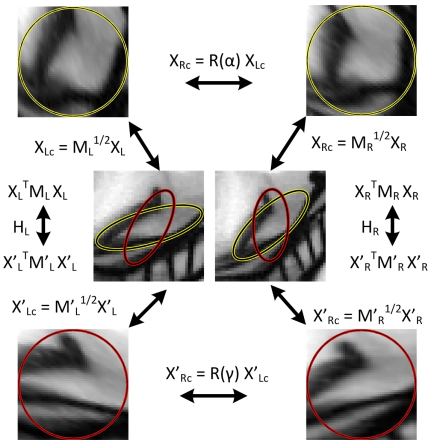


Fig. 5. Using new support regions (red) as anisotropically scaled and rotated version of the original regions detected by affine detectors (yellow). The new red regions are related by rotation.

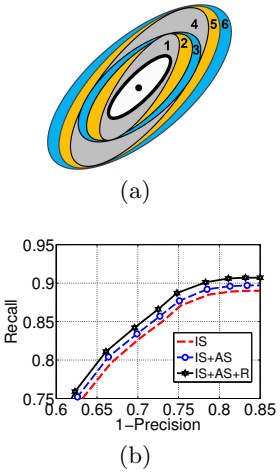


Fig. 6. a) Support regions. b) Comparing different support region strategies: IS, and its combination with AS, and R.

4 Experimental Results

4.1 Gradient Calculator Robustness

In the first experiment, we examine the robustness of the READ operator for gradient calculation for different imaging effects and artifacts. We compare the *READ* gradient operator with the central difference, the first order derivative of Gaussian (i.e., $\Delta f(x) = f(x) * \Delta_G(\sigma)$, where $*$ is convolution and $\Delta_G(\sigma) = \frac{-2x}{\sqrt{2\pi}\sigma^3} e^{-\frac{x^2}{2\sigma^2}}$), and Sobel. Central difference is an old operator which is still used by some descriptors (e.g., MROGH). First order derivative of Gaussian is used in the Canny edge detector (a popular edge detector) and Sobel is a known gradient operator. For the experiments we use Flower, Foliage, Friut, Winter, and Man Made datasets from the McGill color image collections[1]. This includes 821 color images which are converted to the gray scale format. For the noise experiments, two types of noise are added: Gaussian noise with a specific standard deviation ($\sigma = 1, 1.5, ..., 3$) and the Salt & Pepper noise with different noise densities ($density = 0.05, 0.10, ..., 0.35$). For the blur effect experiment, the images are smoothed with a Gaussian kernel (i.e., $K(x, y) = e^{-\frac{x^2+y^2}{2\sigma^2}}$) with a window size of $W \times W$ ($W = (1.5 \times \sigma + .5) \times 2 + 1$). The experiment is performed for $\sigma = 1, 1.5, ..., 3$. Finally, the motion blur effect is generated by Matlab using distances of 4 to 36 pixels with a step size of 4 pixels. To make the result of the methods comparable, the gradient vector at pixels ($G_i = [g_x, g_y]^T$) is normalized

[1] http://tabby.vision.mcgill.ca/html/welcome.html

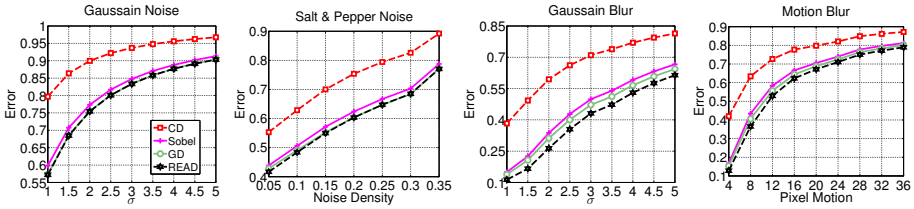


Fig. 7. The robustness of Central Difference (CD), Gaussian Derivative (GD), Sobel, and READ for different imaging effects.

in each image I:

$$G_i = \frac{G_i}{\sum_{\forall G_i \in I} |G_i|}. \tag{17}$$

Then we measure the normalized error:

$$Err = \frac{1}{N} \sum_{i=1}^{N} \frac{|G_i^{cor} - G_i^{orig}|}{|G_i^{orig}|}, \tag{18}$$

where G_i^{cor} and G_i^{orig} are the gradient vectors in the corrupted and the original image, and N the number of pixels in the image. To avoid instability due to small values in the denominator, the vectors with small magnitudes ($|G_i^{orig}| < 10^{-6}$) are excluded. Figure 7 compares the normalized error of the compared methods. In the noise conditions (both Gaussian and Salt & Pepper) the proposed READ operator and the first order derivative of Gaussian are equally the most robust methods. However, in blur conditions (Motion and Gaussian) the READ operator outperforms the other methods. The central difference is the most sensitive method in all experiments.

4.2 Tuning Parameters

The next experiment is performed to find the best tuning parameters for the gradient scale, rotation, and scaling factors of the supporting regions. A total of 50 image pairs with different transformations (mainly rotation and zoom) are used [2]. Six regions are considered (Figure 6(a)). All regions undergo isotropic scaling (IS) by a factor of 1.5 from the previous region, regions 1-3 undergo anisotropic scaling (AS) in the direction of eigenvectors of the elliptic region, regions 1 and 4 and regions 3 and 6 are rotated by θ° and $-\theta^\circ$. The best gradient scale is searched for $R = 1, 2, 3, 4, 5$ with corresponsing $P = 6, 8, 10, 12, 14$. The best s_1 and s_2 are searched in the range $[0.7, 1.3]$ in steps of 0.05, and the best rotation angle from the range $[0, 25]^\circ$ with a step of 5°, respectively. One may note that other configurations are also possible, and similar to DAISY the best parameters can be learned systematically [28]. Nonetheless, this specific

[2] Images downloaded from http://lear.inrialpes.fr/people/mikolajczyk/

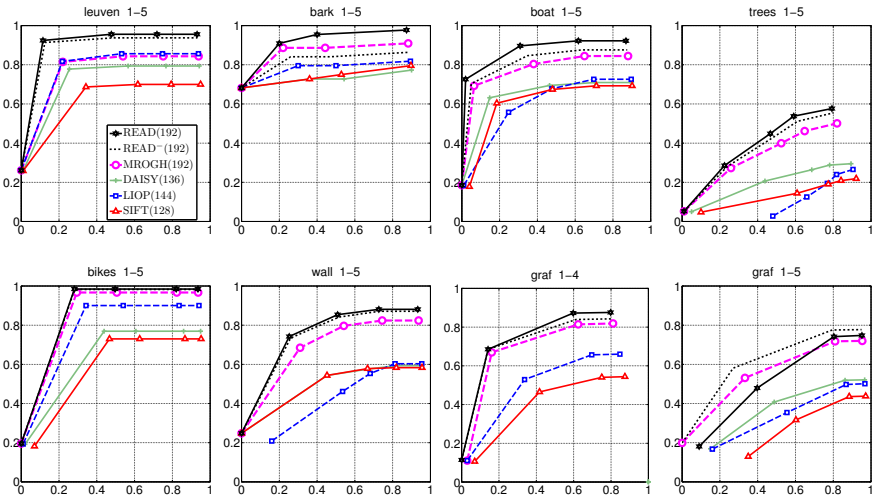


Fig. 8. The performance of the descriptors. The $READ^-$ is our descriptor with no anisotropic scaling or rotation of support regions. The y axis is recall and the x axis is 1-precision.

configuration is sufficient to show the capability of the new descriptor. The best configuration is found to be $R = 4$, $P = 12$, $s_1 = 0.75$, $s_2 = 1.25$, and $\theta = 20°$. Figure 6(b) compares the configuration with only isotropic scaling (IS), and its combinations with anisotropic scaling (AS) and rotation (R). As one can see, the suggested strategy to define support region improves the performance of the descriptor.

4.3 Oxford dataset

To evaluate the performance of the proposed descriptor we follow the evaluation protocol described by Mikolajczyk and Schmid [1] using the standard Oxford dataset[3]. The dataset includes image sets to evaluate different geometric transformations and imaging effects. The first image in each set is considered as a reference image and the other images are acquired under the designated change. A match is considered correct if the overlap error in the image area covered by two corresponding regions is less than 50% of the union of the regions and the recall/1-precision is reported.

We compare our method to SIFT (as the baseline), DAISY, and LIOP and MROGH which have the highest performance according to the recent descriptor comparison study by Miksik and Mikolajczyk [19]. Figure 8 shows the performence of the descriptors. Two versions of our method are shown: 1) using only isotropic scaling for support regions ($READ^-$), and 2) adding anisotropic

[3] Available at http://www.robots.ox.ac.uk/ vgg/research/affine/

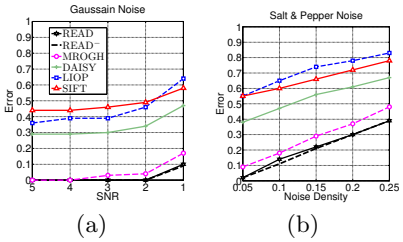
scaling and rotation to the previous version ($READ$). As one can see, the proposed method outperforms the other methods in all cases including illumination change (leuven), rotation and zoom (bark, boat), blur effect (trees, bikes), and view point change (wall, graf). An interesting case is the "1-5" pair in the graf dataset in which adding anisotropic scaling and rotation degrades the performance. This case shows that the performance of rotated and anisotropically scaled support regions relies on the accuracy of the affine region detector. The regions detected by the Hessian-affine detector is not very precise on the graf dataset due to the textureless nature of the scene. Therefore, due to a large viewpoint change, a small inaccuracy produces a large error when we use rotation and anisotropic scaling to define the support regions. Without using rotation or scaling ($READ^-$), however, we can get a much better result for this special case. Nevertheless, if the viewpoint change is smaller (e.g., less than 40° as shown in "1-4" pair in graf) or if the scene has some texture to help better detection (e.g., the wall), anisotropic scaling and rotation improve the result as shown in the other cases. With the exception of leuven, the MROGH is the second best method. In general, all the examined descriptors perform better than SIFT. The average runtime to compute the descriptors on a PC with an Intel quad core 2.60 GHz CPU with 16GB RAM running Windows 7 Professional is 2.4 ms for READ, 3.1 ms for MROGH, 2.1 ms for LIOP, 1.9 ms for DAISY, and 1.0 ms for SIFT. Therefore, our method not only outperforms MROGH but also is quite faster.

4.4 Noise

To evaluate the performance of the descriptors in the presence of noise, the "1-5" image pairs of the Oxford dataset and two types of noise (i.e., Gaussian and the salt & pepper, SP) are considered. The Gassian noise with different Signal to Noise Ratio (SNR), and SP noise with different noise density are added to the "5" image while the reference image "1" is unchanged. We compute the area under the curve (AUC) for the recall/1-precision graphs for the original (AUC^{orig}) and the noisy (AUC^{noise}) conditions. The AUC drop ratio ($(AUC^{orig} - AUC^{noise})/AUC^{orig}$) is shown in Figure 9. As one can see, the READ method is the most robust one in all levels of both types of noise. In some levels of noise $READ^-$ is slightly more robust than READ. After READ, the next robust methods are MROGH, and DAISY. SIFT and LIOP are the most sensitive methods in the Gaussian and SP noise, respectively.

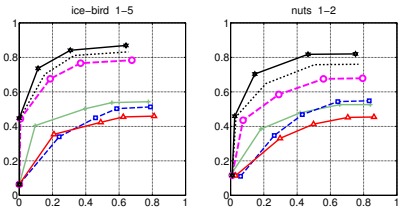
4.5 Motion Blur

Motion blur is one of the common and challenging problems in many computer vision applications. In spite of its importance, and to the best of our knowledge, the effect of motion blur on the performance of descriptors have not been explored. To do so, we apply the motion blur effect function in MATLAB to our images and generate and add motion blur effect with distances of 4 to 16 pixels with a step of 4. Figure 10(a) shows the the performance of the descriptors



(a) (b)

Fig. 9. The AUC drop ratio in different levels of noise for a) Gaussian noise, b) salt & pepper noise.



(a) (b)

Fig. 10. Performence of descriptors on a) motion blur, b) non-uniform illumination change. The y axis is recall and the x axis is 1-precision.

for the "1-5" pair (16 pixels distance). One can see that the READ descriptor notably outperforms the other methods.

4.6 Non-uniform Illumination

Non-uniform illumination and shadows are among the most challenging effects. To evaluate the performance of the descriptors we use the "1-2" image pairs from the nuts dataset[4]. As can be seen in Figure 10(b), the READ descriptor remarkably outperforms the other methods with about 14% higher maximum recall than the second best method.

5 Conclusions

In this paper we propose a new gradient operator called READ which is robust to imaging effects. A kernel implementation of the READ is presented which makes the computation efficient. We demonstrate the ability of READ by defining a region descriptor and comparing it with other state-of-the-art descriptors. This paper presents a novel method to define support regions for the regions detected by an affine detector. The experimental results show that the READ descriptor outperforms the state-of-the-art descriptors such as SIFT, LIOP, DAISY, and MROGH in ordinary geometric transformations and common imaging effects. Additional experiments on noise, motion blur, and non-uniform illumination change further demonstrate the robustness and superior performance of the proposed method. The proposed READ gradient calculator can be used in other image processing and computer vision applications such as segmentation, texture classification, object recognition, and in applications that need robust gradient estimation.

[4] Accessible at http://lear.inrialpes.fr/people/mikolajczyk/Database/

References

1. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Machine Intell. **27** (2005) 1615 –1630
2. Fan, B., Wu, F., Hu, Z.: Rotationally invariant descriptors using intensity order pooling. IEEE Trans. Pattern Anal. Machine Intell. **34** (2012) 2031–2045
3. Lowe, D.: Distinctive image features from scale-invariant keypoints. Intl. J. of Computer Vision **60** (2004) 91–110
4. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. Intl. J. of Computer Vision **60** (2004) 63–86
5. Mikolajczyk, K., Tuytelaars, T., Schmid, e.a.: A comparison of affine region detectors. Intl. J. of Computer Vision **65** (2005) 43–72
6. Tuytelaars, T., Mikolajczyk, K.: Local invariant feature detectors: a survey. Foundations and Trends in Computer Graphics and Vision **3** (2008) 177–280
7. Heikkilä, M., Pietikäinen, M., Schmid, C.: Description of interest regions with local binary patterns. Pattern Recognition **42** (2009) 425–436
8. Forssen, P., Lowe, D.: Shape descriptors for maximally stable extremal regions. In: Intl. Conf. on Computer Vision, ICCV. (2007) 1–8
9. Wang, Z., Fan, B., Wu, F.: Local intensity order pattern for feature description. In: Intl. Conf. on Computer Vision, ICCV. (2011) 603–610
10. Alcantarilla, P.F., Bartoli, A., Davison, A.J.: Kaze features. European Conference on Computer Vision, ECCV (2012) 214–227
11. Tola, E., Lepetit, V., Fua, P.: Daisy: An efficient dense descriptor applied to wide-baseline stereo. IEEE Trans. Pattern Anal. Machine Intell. **32** (2010) 815–830
12. Fan, B., Wu, F., Hu, Z.: Aggregating gradient distributions into intensity orders: A novel local image descriptor. In: IEEE Conf. on Computer Vision and Pattern Recognition , CVPR. (2011) 2377–2384
13. Calonder, M., Lepetit, V., Strecha, C., , e.a.: Brief: Binary robust independent elementary features. In: European Conference on Computer Vision, ECCV. Springer (2010) 778–792
14. Leutenegger, S., Chli, M., Siegwart, R.Y.: Brisk: Binary robust invariant scalable keypoints. In: Intl. Conf. on Computer Vision, ICCV. (2011) 2548–2555
15. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: Orb: an efficient alternative to sift or surf. In: Intl. Conf. on Computer Vision, ICCV. (2011) 2564–2571
16. Alahi, A., Ortiz, R., Vandergheynst, P.: Freak: Fast retina keypoint. In: IEEE Conf. on Computer Vision and Pattern Recognition , CVPR. (2012) 510–517
17. Trzcinski, T., Christoudias, M., Fua, P., Lepetit, V.: Boosting binary keypoint descriptors. In: IEEE Conf. on Computer Vision and Pattern Recognition , CVPR. (2013) 2874–2881
18. Heinly, J., Dunn, E., Frahm, J.M.: Comparative evaluation of binary features. In: European Conference on Computer Vision, ECCV. (2012) 759–773
19. Miksik, O., Mikolajczyk, K.: Evaluation of local detectors and descriptors for fast feature matching. In: Intl. Conf. on Pattern Recognition, ICPR. (2012) 2681–2684
20. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Machine Intell. **24** (2002) 971–987
21. Arof, H., Deravi, F.: Circular neighbourhood and 1-d dft features for texture classification and segmentation. IEE Proc Vis Image Signal Process **145** (1998) 167–172

22. Liao, S., Chung, A.: A new subspace learning method in fourier domain for texture classification. In: Intl. Conf. on Image Processing, ICIP. (2010) 4589 –4592
23. Maani, R., Kalra, S., Yang, Y.: Rotation invariant local frequency descriptors for texture classification. IEEE Transactions on Image Processing **22** (2013) 2409–2419
24. Maani, R., Kalra, S., Yang, Y.H.: Noise robust rotation invariant features for texture classification. Pattern Recognition **46** (2013) 2103–2116
25. Maani, R., Kalra, S., Yang, Y.: Robust volumetric texture classification of magnetic resonance images of the brain using local frequency descriptor. IEEE Transactions on Image Processing **23** (2014) 4625–4636
26. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. IEEE Trans. Pattern Anal. Machine Intell. **27** (2005) 1265–1278
27. Takacs, G., Chandrasekhar, V., Tsai, S.S., Chen, D., Grzeszczuk, R., Girod, B.: Fast computation of rotation-invariant image features by an approximate radial gradient transform. IEEE Transactions on Image Processing **22** (2013) 2970–2982
28. Winder, S., Hua, G., Brown, M.: Picking the best daisy. In: IEEE Conf. on Computer Vision and Pattern Recognition , CVPR. (2009) 178–185