CREAK: Color-based Retina Keypoint Descriptor

Yi-An Chen, Chia-Hsin Chan, and Wen-Jiin Tsai, *Member*, *IEEE*
Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan, R.O.C.

Abstract – *Effectively feature matching between images is key to many computer vision applications. Recently, binary descriptors are attracting increasing attention for their low computational complexity and small memory requirement. However, most binary descriptors are based on intensity comparisons of grayscale images and did not consider color information. In this paper, a novel binary descriptor inspired by human retina is proposed, which considers not only gray values of pixels but also color information. Experimental results show that the proposed feature descriptor spends fewer storage spaces while having better precision level than other popular binary descriptors. Besides, the proposed feature descriptor has the fastest matching speed among all the descriptors under comparison, which makes it suitable for real-time applications.*

Keywords: image matching, feature descriptor, keypoint descriptor, human retina, color information

1 Introduction

A great number of computer vision applications, like image search, image recognition, object tracking and image classification depend on describing particular feature points over an image. In order to represent feature points efficiently, applying robust and stable feature descriptor is necessary. However, how to make descriptor more invariant to geometric and lightning transformations while requiring low computation complexity and small amounts of memory is a big challenge. Therefore, many approaches are developed over the last decades.

The most well-known descriptor is Lowe's SIFT [1] feature descriptor which is floating-point based and provides invariance to a variety of common image transformations, but the disadvantages of SIFT are expensive cost of computation and storage. SURF [2] proposed by Bay *et. al.* is designed to improve performance of SIFT and can use less computational time to achieve similar matching rates compared to SIFT. Despite SURF is much faster than SIFT, however it is still impracticable in many real-time applications, such as embedded devices and mobile phones.

In recent years, several binary descriptors have been proposed. Unlike float-point based descriptors which need to represent image information with local gradient histogram, binary descriptors, in contrast, provide the gray value comparison around detected feature points in the image patch, and then image patch information is encoded with a fixed size binary string. Since binary descriptor use Hamming distance and XOR operation for measuring similarity between two descriptors, they can significantly decrease computational time. The performance of binary descriptors can reach as well as float-points ones, while reducing computational costs and memory requirements.

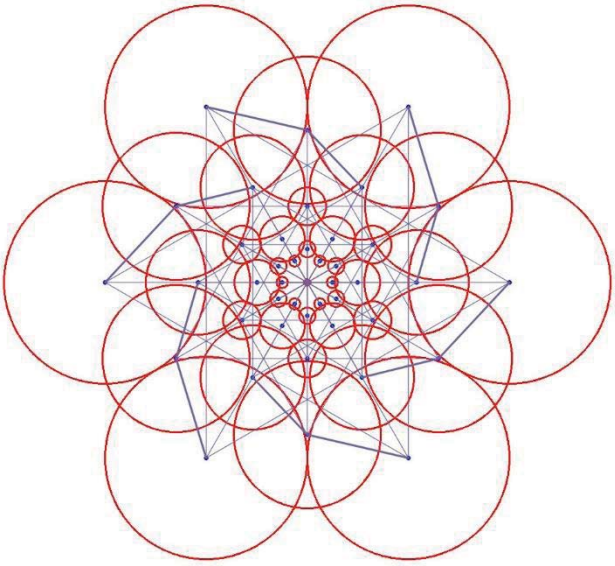


Fig. 1. The proposed CREAK orientation pairs. The lines denote the orientation pairs of FREAK and the bold lines denote the additional pairs for the proposed CREAK descriptor.

For the state-of-the-art binary descriptors, such as BRIEF [4], ORB [6] and FREAK [3], when encoding information of image patch, they only perform gray value comparisons around feature points in the image patch. However, the important information of color is ignored. Besides, their sampling pairs take only the gray value at single pixel into account and therefore sensitive to noise. In order to solve this problem, these binary descriptors offer some alternative smoothing operations before the pixel value are sampled. Although smoothing image can reduce some of noise-sensitive problems, the side effect is that smoothing image will also decrease the details of image and lead to information loss.

Inspired by the above observations, in order to make descriptor more robust and discriminative, the color information is also necessary. In this paper, we propose an alternative binary descriptor named *CREAK (Color-based REtinA Keypoint descriptor)* which is based on the FREAK descriptor and inspired by the photoreceptive cells over the retina, by comparing pixel lightness intensity and color information of pixel rather than single pixel intensity to mimic the retina of human eye. Experiments results show that the proposed feature descriptor not only preserves the ability to fast matching but also spends only less than half size of FREAK

descriptor while having similar or even better matching rate than not only FREAK but also other state-of-the-art binary descriptors.

The rest of this paper is organized as follows. Section 2 describes related works. Section 3 describes the proposed method and its implementation. Section 4 evaluates the performance of proposed method with other descriptors and the result of real matching situation. Finally, the conclusion and future works are given in Section 5.

2 Related Work

The feature descriptors can be generally categorized into two groups: one is float-point based descriptor and another is binary descriptor. SIFT [1] is the most popular float-point based descriptor in the last decade and it presents a highly descriptive power and powerful robustness against to a variety of image transformations. First, SIFT uses sequences of DoG (Difference-of-Gaussians) functions to identify potential features that are invariant to rotation and scale, then it computes a grid of oriented gradient histograms to store the descriptor into a 128-dimensional vector. Several float-point based approaches were proposed to improve performance of SIFT, SURF [2] by Bay *et. al*. is a successful one. The computation time of SURF is faster than SIFT, while its matching performance is close to SIFT's by representing features with the responses of Haar wavelets for approximating gradient orientations in the SIFT. However, SURF belongs to float-based descriptor group, it still relies on floating-point calculations to measure Euclidean distance between two descriptors, which increases time to match features across different images and make descriptors impracticable in real-time applications or low-power devices.

Another group of descriptors is called binary descriptor which were proposed to overcome the shortcomings of float-point based descriptors. Recently, binary descriptors are attracting increasingly attention due to their advantages. They calculate Hamming distance and employ fast XOR operation to measure of distance between two binary descriptors. This makes binary descriptors become more faster matching speeds than float-point ones. Furthermore, by using no more than 512 bits, a single binary descriptor requires far less space than SIFT or SURF. BRIEF [4], the first binary descriptor to describe image features achieves great speed acceleration by simply computing the gray value comparisons of random test pairs in the region of interest. Unfortunately, since simple pixel-based test pair is highly sensitive to noise or other change in local appearances, BRIEF is not robust enough to geometric and lightning transformations especially in rotation variation.

According to BRIEF's method, Rublee *et. al*. proposed the ORB [6] descriptor. By estimating first order moments within the patch, ORB can invariant to rotation. It also selects highly uncorrelated pixel pairs for binary test instead of random selected test pairs of BRIEF. Another approach different from ORB and BRIEF is the BRISK [7] proposed by Leutenegger *et. al*. which emphasizes locality by computing intensity differences between two short-

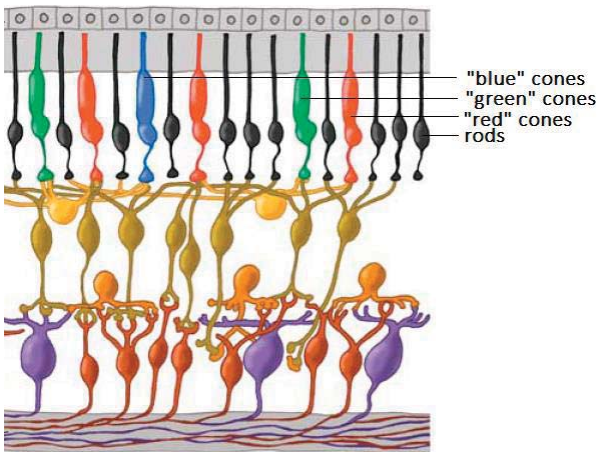
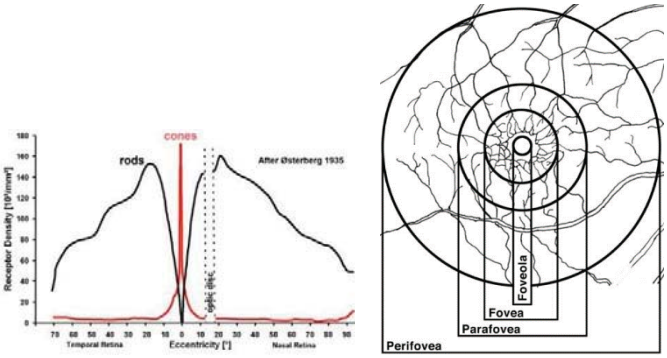


Fig. 2. Cells in the human retina are arrayed in discrete layers [16]



(a) (b)

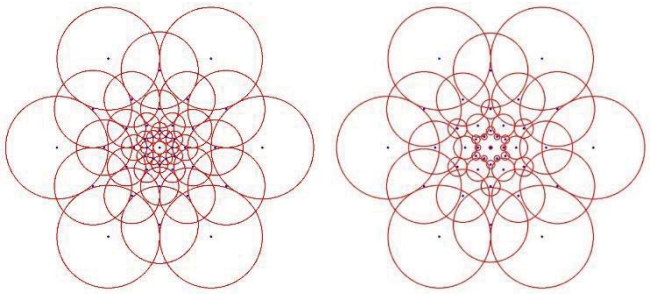
Fig. 3. (a) Topography of the layer of rods and cones in the human retina [17]. (b) Human retina areas [16].

distance or long-distance pixels in a predefined concentric sampling pattern. For computing the patch orientation BRISK uses pixel pairs with large distances while building binary descriptor with short ones.

FREAK (Fast REtinA Keypoint descriptor) [3] is similar to BRISK. It also describes feature point with predefined concentric sampling pattern which was inspired by the retina patterns of human eye. Differing from BRISK, FREAK samples more points exponentially in the inner area to mimic fovea of retina. Besides, using the same learning method of ORB, FREAK also chooses an optimal set of sampling pairs.

Usually, most of binary descriptors perform serval smoothing operations before the pixel pairs are sampled, in order to handle noise-sensitive problem. However, this method also decreases spatial information of image patch. Recently, Gil Levi and Tal proposed the LATCH [8] descriptor which extracts more spatial information from image patch in each descriptor's bit by comparing pixel patches instead of individual pixel value.

Unlike the state-of-the-art binary descriptor based on gray-scale image, and with same concept as LATCH, our CREAK descriptor also extracts more spatial information in the image rather than single pixel value or pixel patch, and compares luminance as well as color information of pixel instead of single pixel intensity to make the descriptor more robust to noise.



(a) (b)

Fig. 4. Comparison of (a) FREAK and (b) the proposed CREAK sampling pattern.

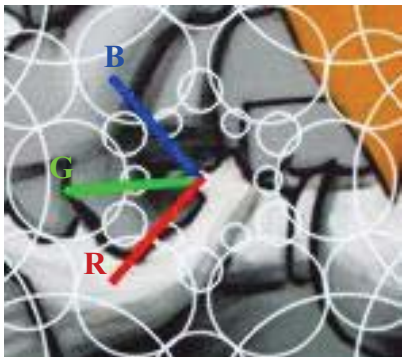


Fig. 5. The same feature point will correspond to three different sampling points (for RGB channels) after orientation estimation.

3 The CREAK Descriptor

3.1 Motivation

According to the research of neuroscience, the retina of human eyes plays a key role in the human visual system (HVS). The purpose of the retina is to receive light, convert them into neural signals, and send these signals to the brain for visual recognition [15]. The retina contains two types of photoreceptive cells: rods and cones. Rod cells have very low spatial resolution but are highly sensitive to light, so they are responsible for the information of illuminance and they are not present in the fovea region. In contrast, cone cells have very high spatial resolution but are relatively insensitive to light, they are primarily located in the fovea region and give us the ability to distinguish colors.

Current understanding that cones also can be subdivided into blue cones, green cones, and red cones based on three different response curves. The topography of the layer of rods and cones in the human retina is shown in Figs. 2 and 3. Consequently, the proposed descriptor is designed by simulating the topology and photoreceptive cells distribution of the retina to describe the features of an image, as described in the next subsection.

3.2 Sampling pattern

In order to design a sampling pattern of binary descriptor which is similar to the human retina, we referenced and modified the FREAK's sampling pattern which is also inspired by the human retina. The characteristic of FREAK's sampling pattern is that the size of its receptive fields (the size of circles in Fig. 4(a)) mimics the density of ganglion cells, which grows exponentially with the distance toward the center of the retina. Furthermore, it also makes receptive fields of the pattern overlapping each other, this can increase spatial redundancy and bring descriptor more discriminative power.

FREAK performs Gaussian smoothing on the sampling points with variable blur kernels according to its receptive field size, however, it is designed for gray level. For color information, the density of the cone cells which have a higher spatial resolution, is higher close to the center of retina than elsewhere, as shown in Fig. 3(a). But large receptive fields adopted in FREAK will decrease the spatial resolution due to the large smoothing area. In order to support color information in the proposed descriptor, we resize each blur kernel to be smaller, especially for the respective fields corresponding to fovea and foveola areas [16] as shown in Fig. 3(b), to simulate the photoreceptive cells distribution of the retina [17]. The comparison of sampling patterns adopted in FREAK and the proposed CREAK is shown in Fig. 4(b).

3.3 Orientation pairs

Since we reduce the receptive field sizes considering the newly added color information, it may decrease the descriptor tolerance for some homographic transforms, such as rotation and scaling. After investigating the orientation pairs in FREAK, we observe that: (i) the angles between feature point center and the orientation pairs are limited to several specific angles, (ii) there are less pairs in perifovea area than fovea and foveola areas, and (iii) the pairs are linked in the same layers. Therefore, we proposed to add 12 *cross-layer orientation pairs* in the perifoveal area, making the proposed CREAK descriptor having 57 pairs, while FREAK have 45 pairs, as shown in Fig. 1. These pairs can not only generate more angles to improve the tolerance of rotation, but also utilize the inter layer information of receptive fields to improve the tolerance of scaling.

For the orientation estimation, we use the same method as that of FREAK which estimates local gradients over selected pairs

$$O = \frac{1}{M}\sum_{Po\in G}\left(I(P_o^{r1}) - I(P_o^{r2})\right)\frac{P_o^{r1}-P_o^{r2}}{\left\|P_o^{r1}-P_o^{r2}\right\|} \qquad (1)$$

where M is the number of pairs in the set of all the pairs used to compute the local gradients G and P_o^{ri} is the 2D vector of the spatial coordinates of the center of receptive field.

We perform the orientation computation separately for each color channel because even with the same orientation pairs, different color channels could have different gradients and that means the same feature point will correspond to three different sampling points after orientation estimation. By doing this, more spatial information can be retrieved for same feature point to increase the performance. An example is shown in Fig. 5.

3.4 Building the Descriptor

We construct our descriptor by performing tests for the intensities of the predefined test pairs, which is a common binary descriptor construction process. Let $I(P_a^1)$, $I(P_a^2)$ denote the smoothed intensity of the pair $P_a = (P_a^1,\ P_a^2)$, the binary test function T(P_a) is formulated as

$$T(P_a) = \begin{cases} 1, & if\ I(P_a^1) \geq I(P_a^2) \\ 0, & otherwise \end{cases} \quad (2)$$

For the proposed CREAK descriptor which consists of three color channels, the binary tests are performed for each channel. Then, the complete binary descriptor D of size N is formed by concatenating three N/3 binary test results and defined as

$$\mathrm{D} = \sum_{i=0}^{\frac{N}{3}-1} \left[2^i T(B_i) + 2^{i+\frac{N}{3}} T(G_i) + 2^{i+\frac{2N}{3}} T(R_i) \right] \quad (3)$$

where B, G, R represent color channels blue, green, and red, while B_i, G_i, R_i are color test pairs of receptive fields with their corresponding channels, respectively. For example, if the total descriptor size N = 192, then the number of bits used by each color channel will be N/3 = 64.

In order to choose a set of test pairs that is best for describing the feature point, we employ the same training method in FREAK, but separately applied for each color channel. The training process consists of the following steps:

1) For each feature point, compute a descriptor composed of all possible test pairs. Create a matrix M whose rows are associated to the feature point and columns are associated to all the possible test pairs.
2) Compute the means of each column in M.
3) According to ORB [6], a higher variance is desired in order to produce a discriminant feature, and the mean value of 0.5 will have the highest variance for a binary distribution. Therefore, the matrix columns are sorted by the absolute value minus 0.5.
4) Keep the best column and iteratively add columns having low correlation with the selected columns.

In the proposed descriptor, test pairs are selected by training from approximately 500k feature points which are drawn from images in the PASCAL 2006 dataset [14]

4 Experimental results

The proposed CREAK descriptor have been implemented in C++ and integrated into OpenCV 3.1 for performance evaluations. The experiments are conducted following the evaluation framework presented in [12]. The framework consists of applying Gaussian blur, brightness change, rotation, and scale change to each image from the Oxford datasets proposed by Mikolajczyk and Schmid [11]. All the experiments are executed using a Desktop PC with an Intel i7 3.4 GHz processor and 12 GB of RAM.

4.1 Color Space Selection

Firstly, we measure the performance of using different commonly used color spaces, including YCrCb, Lab, and RGB. The image graffiti [11] with 12 different levels of blur, 85 levels of brightness, 73 levels of rotation degree, and 71 levels of scales are adopted in the experiments and the results in terms of matching correctness ratio (i.e., # of correct matches divided by # of matches) are shown in Table. 1. It is observed that when using the same bit-length of descriptors, RGB color space obtains the best performance on average. We also compare different descriptor lengths for RGB color space, and observe that a descriptor with length more than 192 bits does not make apparent improvement. Therefore, 192 bits (24 bytes) of descriptor length and the RGB color space are recommended and used for the proposed CREAK descriptor in the later experiments of this paper.

TABLE 1. Average matching correctness ratio comparison for various color spaces

	Blur	Brightness	Rotation	Scale
Lab-192	69.86%	97.63% (2nd)	93.98%	93.76%
YCrCb-192	69.21%	96.09%	92.66%	95.13%
RGB-192	70.60% (1st)	97.54%	94.07% (2nd)	95.38% (2nd)
RGB-384	70.54% (2nd)	97.66% (1st)	94.57% (1st)	96.84% (1st)

4.2 Comparison with different descriptors

In this section, the performance of the proposed CREAK descriptor is compared with a wide range of binary feature descriptors available in OpenCV, including BRIEF [4], BRISK [7], ORB [6], FREAK [3], and LATCH [8], with their default parameters. The same feature point (keypoint) detector is employed for a fair comparison of descriptor performance. The feature detector proposed in ORB is adopted in our experiments, due to its good performance and high speed. The experiment results are shown in Figs. 6-9 for test case "graffiti 1". It is observed that except for the blur transformations, in all testing conditions the proposed CREAK descriptor presents the most robust performance, compared to other binary competitors. For blur transformations, CREAK has lower performances than BRIEF and LATCH, and has comparable performances to ORB. However, both BRIEF and LATCH have extremely low performances for scale transformations, while ORB and BRIEF have bad performances for rotation transformations. A real matching case example is provided in Fig. 10.

Furthermore, we also compare the proposed CREAK descriptor with state-of-the-art floating-point based descriptors including SIFT and SURF with their default feature detectors, as shown in Figs.11-14. The experiment results show that CREAK achieve the better performance among all.

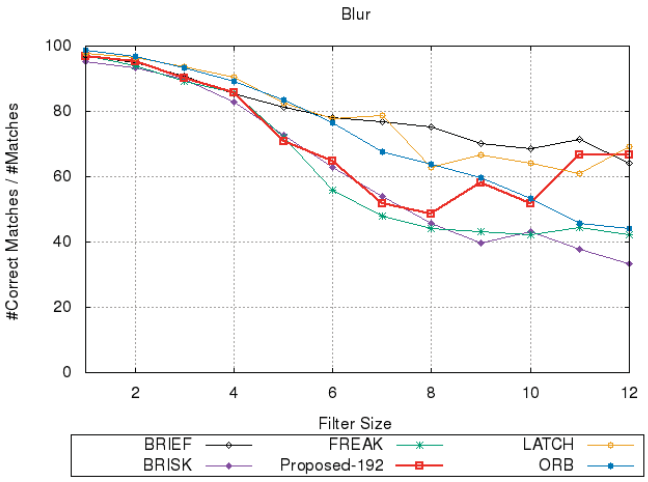


Fig. 6. Performance comparison of binary descriptors under blur transformations.

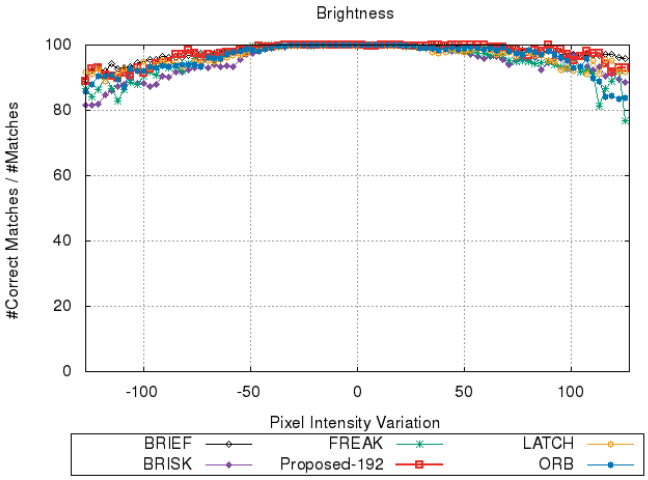


Fig. 7. Performance comparison of binary descriptors under brightness transformations.

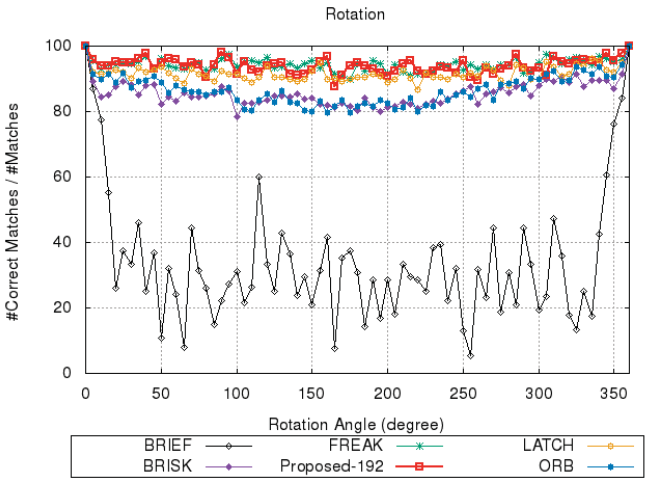


Fig. 8. Performance comparison of binary descriptor under rotation transformations.

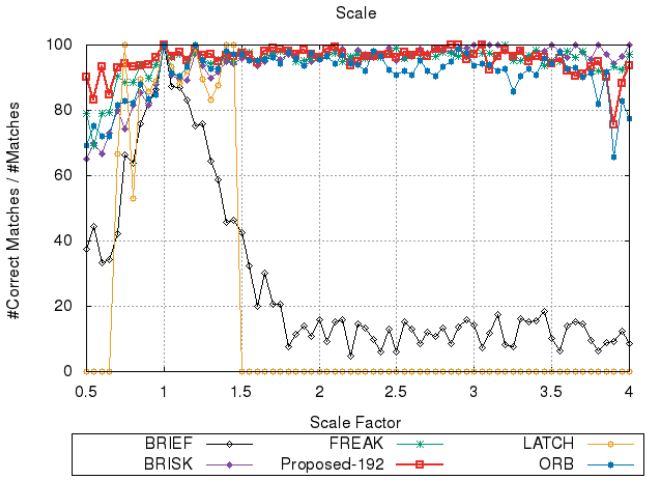
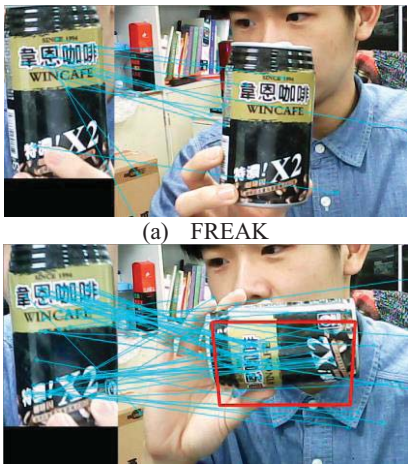


Fig. 9. Performance comparison of binary descriptors under scale transformations.



(a) FREAK

(b) CREAK

Fig. 10. A real matching case example. (a) FREAK fails for the simple upright matching test while (b) the proposed CREAK matches successfully even for the object under 90-degree rotation with view point change.

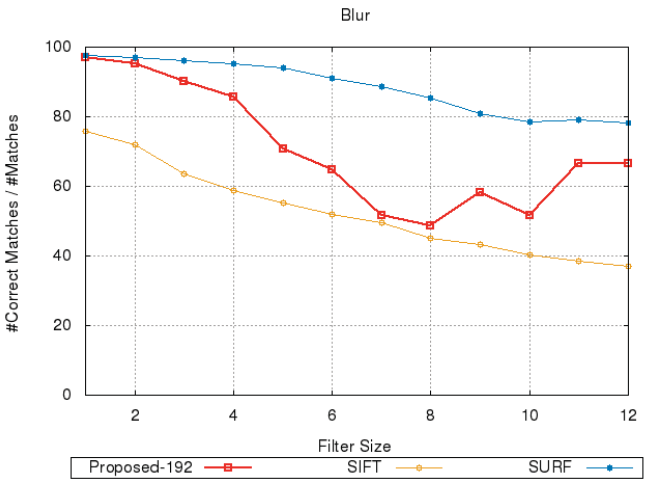


Fig. 11. Comparison of float-point based descriptors under blur transformations.

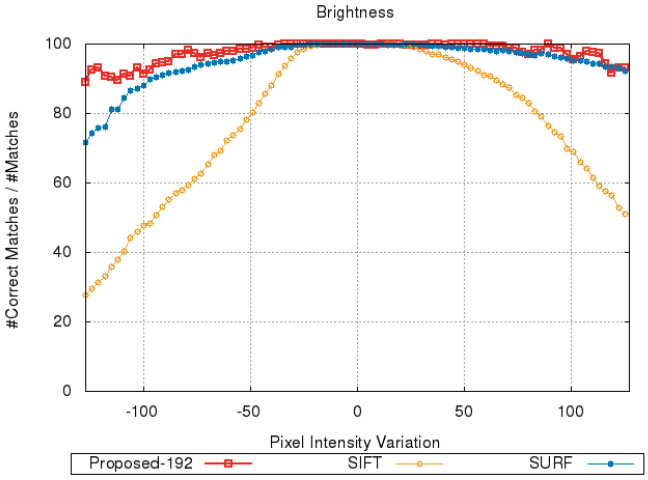


Fig. 12. Comparison of float-point based descriptors under brightness transformations.

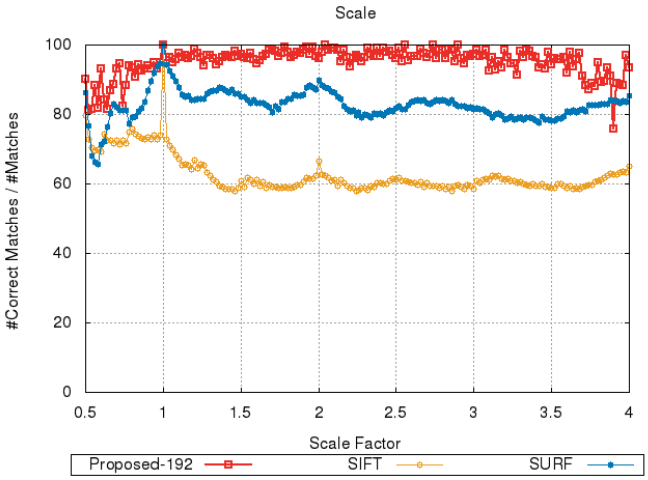


Fig. 13. Comparison of float-point based descriptors under scale transformations.

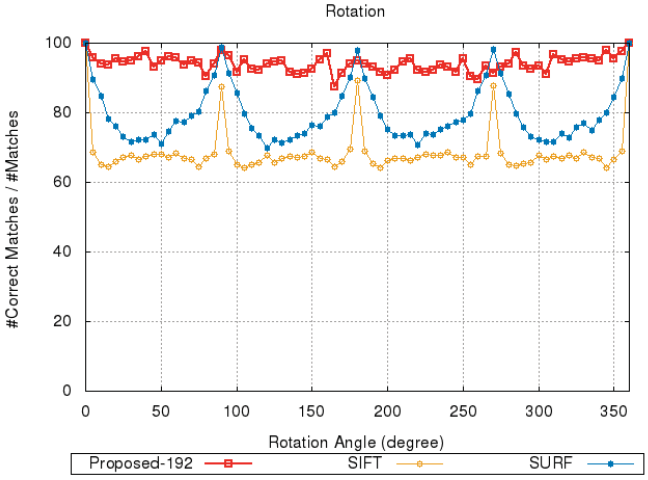


Fig. 14. Comparison of float-point based descriptors under rotation transformations.

TABLE 2. Performance comparison of computation times (in milliseconds) for different feature descriptors

Descriptor	Description	Matching	Total
SIFT	0.553	4.14	4.69
SURF	0.089	2.04	2.13
BRIEF	**0.005**	1.24	1.25
BRISK	0.013	1.53	1.54
ORB	0.024	1.17	1.19
FREAK	0.018	0.62	0.64
LATCH	0.048	1.31	1.36
Proposed	0.026	**0.32**	**0.35**

TABLE 3. Performance comparison of storage requirements (in bytes) for different feature descriptors

Descriptor	Storage
SIFT	512
SURF	256
BRIEF	64
BRISK	64
ORB	32
FREAK	64
LATCH	32
Proposed	**24**

Tables 2 and 3 show the comparison of computation time and storage requirement, respectively, for the descriptors. It is observed that the proposed CREAK descriptor not only requires the least storage space, but also achieves the least total processing time. Considering the computation time and storage requirement, we believe that the proposed CREAK descriptor will be more suitable than other descriptors for the applications requiring real-time feature matching.

In summary, compared to our predecessor, FREAK, CREAK achieves apparent improvements for both blur and brightness transformations, having comparable performances for rotation and scaling, running at about two times faster, while saving more than half of the storage spaces.

4.3 Extremely matching case examples

Moreover, we also display two extremely matching case examples using the graffiti image from [11] with homography matrices to verify inlier matches, as shown in Table 4. Comparing with ORB, CREAK can match two more features whereas FREAK matched none in pair 1|5. It also brings us a great accomplishment that when we apply the most harder image pair 1|6, CREAK matched two features correctly, however, for other descriptors, even the SIFT and SURF, there is no match obtained. The matched feature points by using CREAK are shown in Fig. 15(a) for pair 1|5 and Fig. 15(b) for pair 1|6.

TABLE. 4. Number of match features for graffiti 1|5 and 1|6

	#Match(1|5)	#Inlier(1|5)	#Match(1|6)	#Inlier(1|6)
ORB	**18**	2	**16**	0
FREAK	5	0	10	0
CREAK	12	**4**	7	**2**

(a) Viewpoint change pair 1|5 (5th strength)

(b) Viewpoint change pair 1|6 (6th strength)

Fig. 15. Matched points generated by the proposed CREAK descriptor for the test case "graffiti" under different view changes.

5 Conclusion

In this paper, a novel binary descriptor is presented, which is inspired from human retina. According to the distribution of photoreceptive cells over the retina, more precisely, rods and cones, we comparing color values of the pixels around the feature point instead of the pixel gray value with our sampling pattern based on FREAK's retina sampling pattern. Experimental results show that, the proposed descriptor has better recognition rate than other widely-used binary descriptors even compared with SIFT or SURF, while having low requirements especially in storage. Besides, the matching time of the proposed method outperforms other descriptors due to it have only 192 test pairs, that makes it more suitable for visual algorithm requires real-time performance. Our future work will continue make improvement of feature descriptor based on the principle of the human retina. Last but not least, there are still space for improvement for the blur transformation case for the proposed descriptor. We will also make more effort to study how to improvement the performance for blur transformation.

ACKNOWLEDGMENT

This research is supported in part by MOST-103-2221-E-009-129-MY2 of the Ministry of Science and Technology, Taiwan, R.O.C.

6 References

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int'l Journal of Computer Vision*, vol. 60, issue 2, pp. 91-110, 2004.

[2] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," *European Conf. Computer Vision*, pages 404–417. Springer, 2006.

[3] A. Alahi, R. Ortiz, and P. Vandergheynst., "Freak: Fast retina keypoint," *Computer Vision and Pattern Recognition*, pp. 510–517, 2012.

[4] M. Calonder, V. Lepetit, C. Strecha, and P. Fua., "Brief: Binary robust independent elementary features," *In European Conf. Comput. Vision*, pp. 778–792, 2010.

[5] E. Tola, V. Lepetit, and P. Fua., "Daisy: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 32, issue 5, pp. 815–830, 2010.

[6] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski., "Orb: an efficient alternative to sift or surf," *Int'l Conf Comput. Vision*, pp. 2564–2571, 2011.

[7] S. Leutenegger, M. Chli, and R. Y. Siegwart., "Brisk: Binary robust invariant scalable keypoints," *Int'l Conf Com-put. Vision*, pp. 2548–2555, 2011.

[8] Gil Levi and Tal Hassner., "LATCH: Learned Arrangements of Three Pat*ch Codes,*" *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid, NY, USA, March, 2016

[9] Matheus A. Gadelha, Bruno M. Carvalho., "DRINK: Discrete Robust INvariant Keypoints," *Int'l Conf on Pattern Recognition (ICPR)*, 2014

[10] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool., "A comparison of affine region detectors," *Int'l Journal of Comput. Vision*, vol. 65, pp. 43–72, 2005

[11] K. Mikolajczyk and C. Schmid., "A performance evaluation of local descriptors," *Trans. Pattern Anal. Mach. Intell*., 27(10):1615–1630, 2005

[12] I. Barandiaran, C. Cortes, M. Nieto, M. Graña, and O. Ruiz., "A New Evaluation Framework and Image Dataset for Key Point Extraction and Feature Descriptor Matching," *Int'l. Conf. Computer Vision Theory and Applications* , (VISAPP), 2013

[13] K. Mikolajczyk and C. Schmid., "An affine invariant interest point detector," *Computer Vision* (ECCV), 2002

[14] M. Everingham., "The Pascal Visual Object Classes (VOC) Challenge,"[Online].Available:http://pascallin.ecs.soton.ac.uk /challenges/VOC/databases.html

[15] Helga., "Kolb How the Retina Works", *American Scientist*, 2003

[16] A. Hendrickson., "Organization of the Adult Primate Fovea," *Macular Degeneration,* 2005

[17] G. Osterberg, "Topography of the layer of rods and cones in the human retina," *Acta ophthalmologica., Supplementum 6*, Levin & Munksgaard, Copenhagen, 1935.