Photometric invariant feature descriptor based on SIFT

Ming-liang Gao (高明亮), Xiaomin Yang (杨晓敏), Yanmei Yu (余艳梅), and Daisheng Luo (罗代升)*

College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China

**Corresponding author: dshluo@hotmail.com*

Received August 30, 2011; accepted October 28, 2011; posted online April 27, 2012

For many years, various local feature descriptors have been proposed. Among them, Lowe's scale invariant feature transform (SIFT) descriptor is the most successful one and has been proven to be performed better on the distinctiveness and robustness than other descriptors. However, SIFT descriptor is based on gray level images and pays little attention to the color information which can be a powerful cue in the distinction and recognition of objects. To increase the discriminative power, color features have been plugged into the feature descriptors only recently. In this letter, we study the photometric invariant properties of the Lowe's SIFT, HueSIFT, *rg*SIFT and CSIFT based on color diagonal offset model. Theoretical and experimental results show that the four descriptors are not fully invariant to photometric transformation. To solve this problem, a new color invariant framework based on color diagonal offset model is proposed in this letter. Experimental results validate our proposed framework.

OCIS codes: 100.2000, 330.0330.

doi: 10.3788/COL201210.S11003.

Recently, studies of local feature descriptors have drawn more and more attentions. This can be attributed to the fact that the local features and descriptors can be used to perform reliable matching under different conditions. Local feature descriptors have been applied successfully in a wide range of systems and applications, such as the context of the object recognition[1], scene classification[2],image registration[3], and so on.

Generally, systems based on local feature usually involve three main stages, i.e., feature detection, feature description and feature matching. The detection of interest points determines the stable features that are to be matched and the description of interest points involves creating a unique descriptor for each point by describing it and its surroundings. The matching of interest points which are used for further process matches the descriptors to decide whether these points belong to the object of interest or not.

For good features, image's local properties (such as intensity, gradient, geometric, and shape information) with unique local neighborhood are usually involved. The most successful local image descriptor so far is Lowe's SIFT descriptor[4]. Features extracted using the SIFT algorithm are invariant to image scale, rotation, and partially robust to changing viewpoints and illumination. Since the proposal of the SIFT descriptor, many modified descriptors have appeared in the literature. For example, PCA-SIFT[5], the gradient location orientation histogram (GLOH)[6], speeded up robust feature descriptor (SURF)[7], rotation invariant feature transform (RIFT)[8]. However, most of the existing approaches are based on gray level images and pay little attention to the color information. So they are not fully invariant to photometric transformations.

It is well known that color is a powerful cue in the distinction and recognition of objects. Human visual system can only discern tens of gray intensities but thousands of color values. In many cases, objects can be well recognized by their color information[9,10]. Photometric changes can greatly affect the performance of object recognition if descriptors used are not robust to these changes.

A lot of works have been done to solve the color constancy problem[11−13], among which Finlayson's color diagonal offset model[13] is one of the most successful one. Five types of common photometric changes, i.e. light intensity changes, light intensity shifts, light intensity changes and shifts, light color changes and light color changes and shifts, can be successfully modeled by the color diagonal offset model.

To increase the discriminative power, various color invariants have been plugged into the feature descriptors to make the descriptors invariant to photometric changes. For example, HueSIFT, CSIFT[14] and *rg*SIFT[15]. In this letter, we study the photometric invariant properties of Lowe's SIFT, HueSIFT, CSIFT and *rg*SIFT based on the color diagonal offset model. Theoretical and experimental results show that the four descriptors are partially invariant to the photometric changes. When suffered light color changes and light color changes and shifts, the discriminative power of the four descriptors is unsatisfactory. To achieve photometric variations totally, we develop a new color SIFT descriptor by normalizing each channel independently. The descriptor combines both color and geometrical invariant in a way that is robust to both photometric and geometric changes in imaging condition. Experimental comparison shows that our descriptor outperforms the other color based SIFT, especially under light color changes and light color changes and shift conditions.

Feature detection is essential for feature description. A variety of feature detectors and evaluations of some of them have been proposed[16−18].

Moravec developed one of the earliest corner detectors[19]. It seeked the local maximum of the minimum intensity changes of an image by shifting a binary rectangle window over the image. This detector was anisotropic, noisy and sensitive to edges. Harris *et al.* improved Moravec's descriptor using the second moment matrix[20]. Harris detector was invariant to rotation but failed to deal with scaling. With the development of the

scale-space theory[21,22], Lowe[4] pioneered a scale invariant local feature detection, named as scale invariant feature transform (SIFT). SIFT detector detected a set of local feature vectors through scale extremes and restricted the keypoints' localization based on the measurements of their stability. Meanwhile, Mikolajczyk *et al.*[17,18] extended Harris-Laplace and Hessian-Laplace detectors to obtain invariance to affine adaptation process based on second moment matrix. More recently, Rosten *et al.*[23] proposed FAST detector which compares pixels only on the circle of a fixed radius centered at a key point.

Given the interest points detected by the feature detectors, the remaining task is to describe them for matching and recognition. Usually, the neighbor of a point is used as a support region to generate a descriptor. A good descriptor should be distinctive and invariant to as many variations as possible such as geometric and photometric variations. Many different techniques for describing local regions of an image have been developed[4−8,26−28] and the evaluations of feature descriptors have been carried out[6,24−25].

Mikolajczyk *et al.*[24] categorized the descriptors into three types, namely, distribution-based descriptors, differential descriptors, and some other descriptors[6]. For the distribution-based descriptors, histograms were used to represent different feature points. The most typical one is Lowe's SIFT descriptor[4] which has been proved to be the most robust among the local invariant feature descriptors. Ke *et al.*[5] simplified the SIFT descriptor by utilizing principal component analysis (PCA) to normalized gradient patches to reduce the size of the descriptor. Lazebnik *et al.*[8] put forward the rotation invariant feature transform (RIFT) which divided each circular normalized patch into concentric rings. Bay *et al.*[7] used the integral images to reduce the computation time and proposed the speeded up robust feature (SURF) descriptor. For the differential descriptors, a set of image derivatives with a given order approximates a point neighborhood. Freeman *et al.*[26] proposed a "steerable filter", which steered derivatives in a particular direction for orientation and scale selection to achieve rotation invariant. Other descriptors used local edges[24], generalized moment invariants[27] and wavelet coefficient[28]. However, almost all the preceding descriptors ignore color information which is important in visual representations.

Color information can make significant contributions to feature detection and matching[29−31]. Because of the additional discriminative power, color invariants have been studied theoretically and experimentally. Funt *et al.*[29] used the Lambertian assumption to achieve photometric invariant indexing of images. Facilitated by Koenderink's Gaussian framework, Geusebroek *et al.*[32] extended the photometric invariance to Gaussian-based derivatives and developed various physical-based color invariants for invariant color representations under different imaging conditions. Stokman *et al.*[33] proposed a generic selection model to select and weight color invariant models for discriminatory and robust image feature detection, yielding an optimal balance between color invariance and distinctiveness. Abdel-Hakim *et al.*[14] proposed CSIFT descriptor using the color invariance approach which was proposed by Geusebroek *et al.*[32]. Weijer *et al.*[15] proposed the *rg*SIFT and HueSIFT based on the rg-histogram and Hue-histogram, respectively. However, these descriptors are not fully invariant to photometric transformation.

In this letter, SIFT, CSIFT, *rg*SIFT and HueSIFT are analyzed and a new color descriptor based on SIFT are proposed.

SIFT consists of four major stages: (1) scale-space extrema detection; (2) feature point locating; (3) feature point orientation assignment; (4) descriptor generation.

In the first stage, the potential feature points are detected by searching over all scales and locations. This can be implemented efficiently by constructing a Gaussian pyramid and searching for local peaks in a series of difference-of-Gaussian (DoG).

$$\begin{aligned} D(x,\, y,\, \sigma) &= [G(x,\, y,\, k\sigma) - G(x,\, y,\, \sigma)]^* I(x,y) \\ &= L(x,\, y,\, k\sigma) - L(x,\, y,\, \sigma), \end{aligned} \tag{1}$$

where $G(x,\, y,\, \sigma)$ is variable scale Gaussian, $I(x,\, y)$ is the input image and $*$ is the convolution operation in x and y of the $G(x,\, y,\, \sigma)$. DoG scale space is computed from the difference of two nearby scales separated by a constant multiplicative factor k.

In the second stage, the candidate feature points with low contrast and localized along an edge are rejected. Sub-sample accurate position and scale is computed for each candidate feature point giving

$$\widehat{X} = -\frac{\partial^2 D^{-1}}{\partial X^2}\frac{\partial D}{\partial X}, \tag{2}$$

where $\widehat{X} = (x,\, y,\, \delta)^{\mathrm{T}}$ is the extremum position providing accurate position and scale.

In the third stage, a consistent orientation is assigned to the keypoints based on local image gradient directions. Dominant orientation is determined as follows. Firstly, the histogram of the gradient orientations are built in the neighborhood of a feature point. Then, every peak in the histogram greater than 80% of the highest peak are selected as a feature point with the corresponding orientation. Lastly, a parabola is fit to interpolate the peaks' position.

The fourth stage builds a local image descriptor for each keypoint based on the magnitudes and orientations of the image gradients in the neighborhood of the key point. Each region is rotated based on its dominant orientation and partitioned into 16 sub-regions of 4×4 pixels. For each pixel within a sub-region, SIFT accumulates the pixel's gradient to orientation histograms with 8 bins by weighting the contribution of each gradient according to its magnitude which forms a 128-element vector. After normalized to unit, the 128-element vector constructs the feature descriptor.

An RGB image recorded by a camera depends on three factors: the physical content of the scene, the illumination incident on the scene, and the characteristics of the camera. Identifying and removing this illumination incident is called color constancy and it is important in computer vision.

Many studies have been carried out to solve the color constancy problem[11−13]. Forsyth[12] modeled the illumination change using three scale factors $[d_1, d_2, d_3]$ such that an observed RGB $[R^{\mathrm{o}}, G^{\mathrm{o}}, B^{\mathrm{o}}]$ is mapped to its cor-

responding RGB under a reference light according to

$$\begin{bmatrix} R^{\mathrm{c}} \\ G^{\mathrm{c}} \\ B^{\mathrm{c}} \end{bmatrix} = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} \begin{bmatrix} R^{\mathrm{o}} \\ G^{\mathrm{o}} \\ B^{\mathrm{o}} \end{bmatrix}. \quad (3)$$

This model is referred to diagonal model. A diagonal model of illumination change which maps the image taken under one illuminant to the image under another illuminant by scaling each channel independently. However, the diagonal model has deficiency to model illumination change for all surfaces in the given image accurately. To overcome this shortcoming, Finlayson *et al.*[13] extended the diagonal model with an offset $[O_1, O_2, O_3]^{\mathrm{T}}$ resulting in the color diagonal-offset model:

$$\begin{bmatrix} R^{\mathrm{c}} \\ G^{\mathrm{c}} \\ B^{\mathrm{c}} \end{bmatrix} = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} \begin{bmatrix} R^{\mathrm{o}} \\ G^{\mathrm{o}} \\ B^{\mathrm{o}} \end{bmatrix} + \begin{bmatrix} O_1 \\ O_2 \\ O_3 \end{bmatrix}. \quad (4)$$

Five types of common photometric changes in image values can be categorized based on the color diagonal-offset model, i.e. light intensity changes, light intensity shifts, light intensity changes and shifts, light color changes and light color changes and shifts.

For the light intensity changes, the $d_1 = d_2 = d_3 = d$ and $O_1 = O_2 = O_3 = 0$. Light intensity changes include the intensity of the light source and no-colored shadows and shading.

For the light intensity shifts, $d_1 = d_2 = d_3 = 1$ and $O_1 = O_2 = O_3 = O$. Light intensity shifts are mainly due to the scattering of a white light source, object highlights under a white light source, inter reflections and infrared sensitivity of the camera sensor.

For the light intensity changes and shifts, $d_1 = d_2 = d_3 = d$ and $O_1 = O_2 = O_3 = O$, which are the combination of the first two classes of changes.

For the light color change, the image channels scale independently, i.e. $d_1 \neq d_2 \neq d_3$ and $O_1 = O_2 = O_3 = O$. The changes in the illuminant color and light scattering can be modeled by these changes.

For the light color changes and shifts which are the full diagonal-offset model with $O_1 \neq O_2 \neq O_3$ and $d_1 \neq d_2 \neq d_3$. These changes are the combination of the second and the forth class of changes. For an illustration of images recorded under these photometric changes, reference image and photometric changes are shown in Fig. 1.

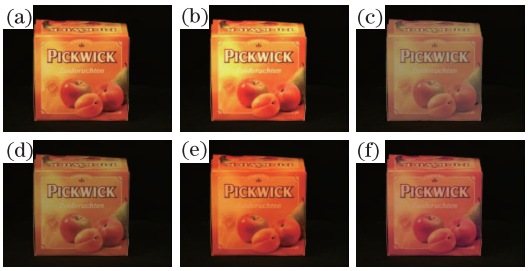

Fig. 1. Reference images and photometric changes. (a) Reference image, (b) light intensity change, (c) light intensity shift, (d) light intensity change and shift, (e) light color change, and (f) light color change and shift.

In Lowe's SIFT, the vector is obtained by first computing the gradient magnitude and orientation at each image sample point around a key point location. So a constant added to each image pixel will not affect the gradient value. Hence the gradient detector makes the descriptor light shift-invariant. Then, the vector is normalized to unit length. A change in image contrast in which each pixel value is multiplied by a constant will be canceled by the normalization. So the normalization makes the descriptor light intensity invariant. However, the SIFT descriptor uses gray gradients which is a linear combination of R, G and B channels. Therefore, the descriptor is not invariant to light color changes.

Weijer *et al.*[15] introduced a concatenation of hue histogram with the SIFT descriptor which was named Hue-SIFT. Hue $= \arctan(O_1/O_1)$, where $O_1 = (R-G)/\sqrt{2}$ and $O_2 = [(R+G-2B)/\sqrt{6})]$ were opponent colors. The Hue color model was invariant to light shifts and light changes.

In *rg*SIFT, the descriptor was normalized as $r = R/(R+G+B)$ and $g = G(R+G+B)$. It can be derived that the descriptor is invariant to light intensity changes due to the normalization.

CSIFT, introduced by Abdel-Hakim *et al.*[14], used the color information based on the color invariance model[32]. In this descriptor, the color invariant H was derived by the Kubelka-Munk theory[34] assuming equal energy illumination,

$$H = \frac{E_\lambda}{E_{\lambda\lambda}}. \quad (5)$$

The color invariant can be obtained by substitution of E, E_λ and $E_{\lambda\lambda}$ by $\widehat{E}$, $\widehat{E}_\lambda$ and $\widehat{E}_{\lambda\lambda}$ using the Gaussian color model[35] as

$$\begin{pmatrix} \widehat{E} \\ \widehat{E}_\lambda \\ \widehat{E}_{\lambda\lambda} \end{pmatrix} = \begin{pmatrix} 0.06 & 0.63 & 0.27 \\ 0.3 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}. \quad (6)$$

It can be derived that only light intensity invariance can be modeled by H.

From the analysis above, we find that all the proposed descriptors are not fully invariant to photometric transformation especially to the color changes and shifts. To solve this problem, we propose a new color invariant SIFT here.

Theoretical studies of the above four descriptors reveal that when all the color channels of the pixes are added (or multiplied) by a constant simultaneously, konwn as the light intensity shift (or light intensity change) , the gradient detector (or normalization) can make these changes invariant. However, once the color channels changes inconsistently, the perfomance of the four descriptors will decline. Therefore, to make the descriptor invariant to these changes, each color channel will be normalized independently.

Like the SIFT detector, in our framework the potential interest points are localized at the local maxima of the difference-of-Gaussian space with local scale which makes each component of our framework for feature point descriptor invariant to scale changes. Then the candidate keypoints are localized to sub-pixel accuracy and the points with low contrast and at the edge are rejected. In

the third stage, instead of calculating the dominant orientations on the intensity channel, we calculate the dominant orientation on R, G and B channels respectively. After detecting the interest points, the local image descriptors are constructed as follows, taking R channel for example.

(1) The values R_i $(i = 1, 2, \cdots, N)$ of R channel in a region around a keypoint are obtained. Here N is the number of the pixels in the region.

(2) The maximum ($R_{\max}$) and the minimum ($R_{\min}$) of R_i in the region around the keypoint are found and each pixel of R_i is normalized, i.e.,

$$R'_i = 255 \times (R_i - R_{\min})/(R_{\max} - R_{\min}). \tag{7}$$

(3) The magnitude and orientation of the gradient of R'_i around the keypoint are calculated by

$$\begin{aligned} m(x, y) &= \sqrt{[R'_i(x+1, y) - R'_i(x-1, y)]^2 + [R'_i(x, y+1) - R'_i(x, y-1)]^2}, \\ \theta(x, y) &= \arctan \frac{R'_i(x, y+1) - R'_i(x, y-1)}{R'_i(x+1, y) - R'_i(x-1, y)}. \end{aligned} \tag{8}$$

(4) The coordinates of the descriptor and the gradient orientations are rotated with respect to the orientation of the keypoint to achieve orientation invariance.

(5) After being weighted by a Gaussian window, these samples are computed at a 8 orientation planes over 4×4 sub-regions. The length of each arrow corresponds to the sum of the gradient magnitudes near that direction within the region. Lastly, the descriptor is formed from a vector containing 128 elements.

Similarly, the pixels in G and B channels are normalized by:

$$\begin{aligned} G'_i &= 255 \times (G_i - G_{\min})/(G_{\max} - G_{\min}) \\ B'_i &= 255 \times (B_i - B_{\min})/(B_{\max} - B_{\min}). \end{aligned} \tag{9}$$

where $G_{\max}$, $G_{\min}$ are the maximum and the minimum of $G_{\rm i}$ in the region around the keypoint, respectively and $B_{\max}$, $B_{\min}$ are the maximum and the minimum of $B_{\rm i}$ in the region around the keypoint, respectively.

Once the descriptors of R, G and B channels are constuced respectively, we combine the three 128-dimension descriptors and get the descriptor vector: $V = [V_{\rm RSIFT}, V_{\rm GSIFT}, V_{\rm BSIFT}]$, where $V_{\rm RSIFT}$, $V_{\rm GSIFT}$ and $V_{\rm BSIFT}$ are the SIFT descriptor vectors that computed for the normalized R, G and B channel, respectively. Figure 2 shows the procedure of the descriptor vectors.

From the color normalization and the descriptor construction presented above, we can see that our method is coincident with the diagonal offset model. Because Eqs. (7) and (9) can be re-written in a matrix form:

$$\begin{bmatrix} R'_i \\ G'_i \\ B'_i \end{bmatrix} = \begin{bmatrix} \frac{255}{R_{\max} - R_{\min}} & 0 & 0 \\ 0 & \frac{255}{G_{\max} - G_{\min}} & 0 \\ 0 & 0 & \frac{255}{B_{\max} - B_{\min}} \end{bmatrix} \begin{bmatrix} R_i \\ G_i \\ B_i \end{bmatrix} - \begin{bmatrix} \frac{255^* R_{\min}}{R_{\max} - R_{\min}} \\ \frac{255^* G_{\min}}{G_{\max} - G_{\min}} \\ \frac{255^* B_{\min}}{B_{\max} - B_{\min}} \end{bmatrix}. \tag{10}$$

It is equivalent to the color diagonal-offset model mentioned in Eq. (4). The relationship between these two equations are

$$\begin{aligned} &d_1 = 255/(R_{\max} - R_{\min}), \quad d_2 = 255/(G_{\max} - G_{\min}), \\ &d_3 = 255/(B_{\max} - B_{\min}). \\ &O_2 = 255^* R_{\min}/(R_{\max} - R_{\min}), \\ &O_2 = 255^* G_{\min}/(G_{\max} - G_{\min}), \\ &O_2 = 255^* B_{\min}/(B_{\max} - B_{\min}). \end{aligned}$$

Generally, our method has stable color invariance. Because R, G and B are uncorrelated, the colors of an image are presented by vectors of (R, G, B). Different values of R, G or B of an object are caused by different sensitivities of different imaging systems, different intensities of different illumination systems, and different reflections of the different environments. Hence, the colors of the object are different. In our method, the different values of R, G and B are normalized, respectively. Therefore, our method is color invariant to the imaging systems and illumination systems.

We have done some experiments to compare the discriminative power of Lowe's SIFT, HueSIFT, *rg*SIFT, CSIFT and our proposed framework under photometric changes. We used Euclidean distance between two feature vectors and nearest neighbor-based ratio matching to determine whether the two vectors belong to the same keypoint in different images or not.

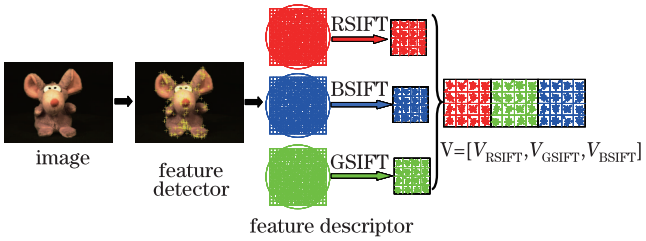

Fig. 2. Procedure of the descriptor generation.

In the experiments, two regions were matched when $||D_P - D_Q||/||D_P - D_Q|| < T$, where D_P and D_R were the first nearest and second nearest neighbor to D_Q respectively, and T was a threshold.

Each descriptor from the reference image was compared with each descriptor from the photometric changes and we calculated the number of the correct matches as well as the number of the mismatches. The results are presented below with recall versus precision[6] .

$$\text{recall} = \frac{\text{correctmatches}}{\text{correspondence}},$$
$$\text{precision} = \frac{\text{correct matches}}{\text{total matches(correct or false)}}. \quad (11)$$

Here, recall is the number of the correctly matched regions with respect to the number of corresponding regions between the two images with same scene. precision indicates the relative number of correct matches in all the returned matches. The curve of recall versus precision is obtained by varying the threshold T. We used Amsterdam Library of Object Images (ALOI)[36], which contains a large number of objects under various photometric changes. Light intensity changes and shifts are not presented in the dataset, so we have artificially added these two condition changes to the dataset. The discriminative power of Lowe's SIFT, HueSIFT, *rg*SIFT, CSIFT and our approach under the photometric changes are depicted in Fig. 3.

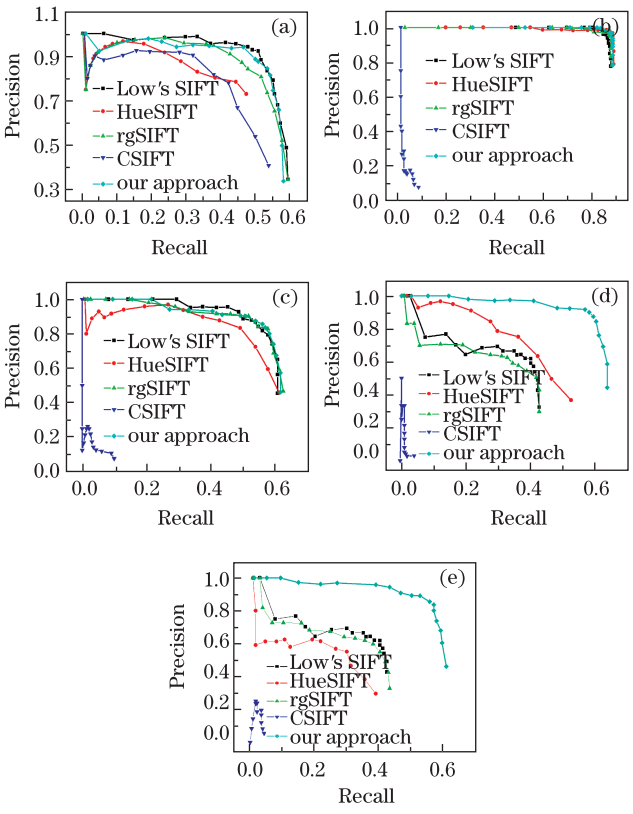

Fig. 3. Discriminative performance of Lowe's SIFT, HueSIFT, *rg*SIFT, CSIFT and our approach under photometric changes. (a) Light intensity changes, (b) light intensity shifts, (c) light intensity changes and shifts, (d) light color changes, and (e) light color changes and shifts.

It can be seen from Fig. 3(a) that the five descriptors perform equally well under light intensity changes. The reseason lies that normalization in these descriptors makes the descriptor light intensity invariant which is consistent with the theoretical analysis. Figures 3(b) and (c) show that CSIFT descriptor lacks of discriminative power under light intensity shifts and light intensity changes and shifts. The reason is that the color invariant $H = (E_\lambda/E_{\lambda\lambda})$ which is based on the color invariance model has no invariance to intensity shifts. It should be pointed out that theoretical studies indicate that the color invariant of *rg*SIFT, where $r = R/(R+G+B)$ and $g = G/(R+G+B)$, has no invariance to light intensity shifts, but the experimental results showed that the *rg*SIFT performs equally well under light intensity shifts. This can be attributed to the fact that SIFT is applied to r and g respectively in the *rg*SIFT descriptor and SIFT descriptor is shift invariant. This makes the *rg*SIFT descriptor invariant to light intensity shifts. Figs. 3(d) and Fig. (e) show clearly the advantage of our approach. It can be seen from Figs. 3(d) and (e) that when *RGB* channel changes independently, like the case of light color changes and shifts, the discriminative power of SIFT, HueSIFT, *rg*SIFT and CSIFT drop distinctly. It indicates that these four descriptors have no invariant to these changes. By constrat, our approach performace significantly better that others. This is due to the fact that normalization is implemented for every *RGB* channel independently in our descriptor, so it makes our descriptor invariant to these photometric changes and shifts.

In conclusion, the photometric invariant properties of Lowe's SIFT, HueSIFT, *rg*SIFT and CSIFT are analyzed based on color diagonal offset model. Theoretical and experimental results show that the four descriptors are not fully invariant to photometric transformation especially to the color changes and shifts. To solve this problem, we proposed a new color invariant framework. In our framework, SIFT descriptor is computed for every normalized *RGB* channel independently. Experimental comparison shows that our descriptor outperforms the other four descriptors, especially under light color changes and light color changes and shifts.

The authors are grateful to the Education Ministry's Key Research Project on Science and Technology of China (No. 107094) and the National Natural Science Foundation of China (No. 11176018).

References

1. X. Liu, Z. Tian, W. Yan, and X. Duan, Chin. Opt. Lett. **9,** 061001 (2011).
2. A. Bosch, A. Zisserman, and X. Munoz, IEEE Trans. Pattern Anal. Mach. Intell. **30,** 712 (2008).
3. J. Bai, Y. Ma, J. Li, F. Fan, and H. Wang, Chin. Opt. Lett. **9,** 081002 (2011).
4. D. Lowe, Int. J. Comput. Vis. **60,** 91 (2004).
5. Y. Ke and R. Sukthankar, in *Proceedings of Computer Vision and Pattern Recognition* 50 (2004).
6. K. Mikolajczyk and C. Schmid, IEEE Trans. Pattern Anal. Mach. Intell. **27,** 1615 (2005).
7. H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, Comput. Vis. Image Underst. **110,** 346 (2008).

8. S. Lazebnik, C. Schmid and J. Ponce, IEEE Trans. Pattern Anal. Mach. Intell. **27,** 1265 (2005).
9. P. Chang and J. Krumm, in *Proceedings of IEEE Computer Vision and Pattern Recognition* 498 (1999).
10. T. Gevers and A. W. M. Smeulders, Pattern Recognit. **32,** 453 (1999).
11. K. Barnard, V. Cardei, and B. Funt, IEEE Trans. Image Process. **11,** 972 (2002).
12. D. A. Forsyth, Int. J. Comput. Vision **5,** 5 (1990).
13. G. D. Finlayson, S. D. Hordley, and R. Xu, in *Proceedings of IEEE International Conference on Image Processing* **3,** 948 (2005).
14. A. E. Abdel-Hakim and A. A. Farag, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* **2,** 1978 (2006).
15. J. van de Weijer, T. Gevers, and A. Bagdanov, IEEE Trans. Pattern Anal. Mach. Intell. **28,** 150 (2006).
16. T. Tuytelaars and K. Mikolajczyk, Computer Graphics and Vision **3,** 177 (2008).
17. K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, Int. J. Comput. Vision **65,** 43 (2005).
18. K. Mikolajczyk and C. Schmid, Int. J. Comput. Vision **60,** 63 (2004).
19. H. Moravec, in *Proceedings of the International Joint Conference on Artif icial Intelligence* 584 (1977).
20. C. Harris and M. Stephens, in *Proceedings of Alvey Vision Conference* 147 (1988).
21. T. Lindeberg, Int. J. Comput. Vision **11,** 283 (1993).
22. T. Lindeberg, Int. J. Comput. Vision **30,** 79 (1998).
23. E. Rosten and T. Drummond, in *Proceedings of the International Conference on Computer Vision* 1508 (2005).
24. K. Mikolajczyk and C. Schmid, IEEE Trans. Pattern Anal. Mach. Intell. **27,** 1615 (2005).
25. J. Li and N. M. Allinson, Neurocomputing **71,** 1771 (2008).
26. W. Freeman and E. Adelson, IEEE Trans. Pattern Anal. Mach. Intell. **13,** 891 (1991).
27. L. Van Gool, T. Moons, and D. Ungureanu, in *Proceedings of Fourth European Conference on Computer Vision* 642 (1996).
28. A. Shokoufandeh, I. Marsic, and S. J. Dickinson, Image Vision Comput. **17,** 445 (1999).
29. B. V. Funt and G. D. Finlayson, IEEE Trans. Pattern Anal. Mach. Intell. **17,** 522 (1995).
30. G. J. Burghouts and J. M. Geusebroek, Comput. Vision Image Underst. **113,** 48 (2009).
31. C. Li and L. Ma, Pattern Recognit. Lett. **30,** 544 (2009).
32. J. M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders, and H. Geerts, IEEE Trans. Pattern Anal. Mach. Intell. **23,** 1338 (2001).
33. H. Stokman and T. Gevers, IEEE Trans. Pattern Anal. Machine Intell. **29,** 371 (2007).
34. L. Yang and B. Kruse, J. Opt. Soc. Am. **21,** 1933 (2004).
35. J. M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders, and A. Dev, in *Proceedings of Sixth European Conference on Computer Vision* **1,** 331 (2000).
36. J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, Int. J. Comput. Vision **61,** 103 (2005).