# Corner Detection Using Multi-directional Structure Tensor with Multiple Scales

Weichuan Zhang[1] · Changming Sun[2]

## Abstract

Corners are important features for image analysis and computer vision tasks. Local structure tensors with multiple scales are widely used in intensity-based corner detectors. In this paper, the properties of intensity variations of a step edge, L-type corner, Y- or T-type corner, X-type corner, and star-type corner are investigated. The properties that we obtained indicate that the image intensity variations of a corner are not always large in all directions. The properties also demonstrate that existing structure tensor-based corner detection methods cannot depict the differences of intensity variations well between edges and corners which result in wrong corner detections. We present a new technique to extract the intensity variations from input images using anisotropic Gaussian directional derivative filters with multiple scales. We prove that the new extraction technique on image intensity variation has the ability to accurately depict the characteristics of edges and corners in the continuous domain. Furthermore, the properties of the intensity variations of step edges and corners enable us to derive a new multi-directional structure tensor with multiple scales, which has the ability to depict the intensity variation differences well between edges and corners in the discrete domain. The eigenvalues of the multi-directional structure tensor with multiple scales are used to develop a new corner detection method. Finally, the criteria on average repeatability (under affine image transformation, JPEG compression, and noise degradation), region repeatability based on the Oxford dataset, repeatability metric based on the DTU dataset, detection accuracy, and localization accuracy are used to evaluate the proposed detector against ten state-of-the-art methods. The experimental results show that our proposed detector outperforms all the other tested detectors.

## 1 Introduction

Corners have been shown to be well suited for a variety of image processing and computer vision tasks such as object tracking, stereo matching, and 3D reconstruction. Various corner detection methods have been reported in the literature. The existing corner detection methods can be broadly classified into three categories: contour-based methods (Rattarangsi and Chin 1990; Teh and Chin 1989; Mokhtarian and Suomela 1998; Zhong and Liao 2007; Zhang et al. 2014; Zhang and Shui 2015; Olson 2000; Zhang et al. 2015, 2019), template-based methods (Deriche and Giraudon 1993; Smith and Brady 1997; Rosten et al. 2010; Shui and Zhang 2013; Xia et al. 2014), and intensity based methods (Moravec 1979; Harris and Stephens 1988; Noble 1988; Gårding and Lindeberg 1996; Lindeberg 1998; Schmid et al. 2000; Kenney et al. 2003; Mikolajczyk and Schmid 2004; Laptev 2005; Lowe 2004; Bay et al. 2006; Marimon et al. 2010; Maver 2010; Su et al. 2012; Verdie et al. 2015; Yi et al. 2016; Lenc and Vedaldi 2016; Zhang et al. 2017). Contour-based methods detect corners by analyzing the shape changes on the edge contours which are extracted from an input image by an edge detector. The contour-based methods rely on the results of a

✉ Weichuan Zhang
zwc2003@163.com

Changming Sun
changming.sun@csiro.au

[1] College of Electrical and Information, Xi'an Polytechnic University, Xi'an 710048, China

[2] CSIRO Data61, PO Box 76, Epping, NSW 1710, Australia

preceding step on image edge detection which affects their applications.

Template-based methods find corners by fitting a small patch of an image with predefined corner templates. Deriche and Giraudon (1993) analyzed the behaviors of wedge and Y-type corners by using the Gaussian filter. In Smith and Brady (1997), every pixel inside a circular mask is compared with a center pixel and the intensity difference is recorded. Corners are defined as the smallest univalue segment assimilating nucleus (SUSAN) points. In Ruzon and Tomasi (2001), junctions are defined as points in an image where two or more piecewise constant wedges meet at the central point. Shui and Zhang (2013) applied the anisotropic Gaussian directional derivative filters (Shui and Zhang 2012) to derive the representations of L-type, Y-type, X-type, and star-type corners and detect corners from edge pixels. Xia et al. (2014) presented a junction detector based on the intensity variations of edge pixels. Pham et al. (2014) presented a junction detection method in which junctions are obtained by searching for optimal meeting points of median lines in line-drawing images. In recent years, machine learning algorithms are used in template-based corner detection methods. Trujillo and Olague (2006) used a genetic programming based learning approach to extract corners from input images. Rosten et al. (2010) extended the SUSAN detector (Smith and Brady 1997) and presented the features from accelerated segment test (FAST) detector.

Intensity-based methods detect corners directly from an input image by analyzing the information on local intensity variations. Following Moravec's observation (Moravec 1979) that the intensity variations of corners are large in all directions, Harris and Stephens (1988) developed the famous Harris detector. The isotropic Gaussian filter was used to smooth the input image and the first-order image derivatives along the horizontal and vertical directions were obtained to construct a $2 \times 2$ structure tensor and detect corners. The aim of the Harris detector is to find corners which have significant changes of image intensities in both directions. The Harris detector is one of the most successful detectors and has been widely used. However, the Harris detector is a single scale detector which may miss significant corners or detect false corners (Lee et al. 1995). The reason is that most objects consist of a wide range of scale features. Meanwhile, it is indicated in Bay et al. (2006) that the most valuable property of a corner detector is its repeatability in affine image transformations. A large number of detectors (Gårding and Lindeberg 1996; Lindeberg 1998; Schmid et al. 2000; Mikolajczyk and Schmid 2004; Laptev 2005; Lowe 2004; Bay et al. 2006; Marimon et al. 2010) have been presented to enhance the repeatability performance of corner detectors in a scale-space representation (Witkin 1984; Koenderink 1984).

Lindeberg (1998) presented a corner detection method with automatic scale selection. Mikolajczyk and Schmid (2004) presented a scale invariant Harris–Laplace detector where corners were detected by the Harris detector in multi-scales, and Laplace operator was used to depict corners' characteristic scales. Lowe (2004) approximated the normalized Laplacian of Gaussian filter by a difference of Gaussian (DoG) filter and presented the scale-invariant feature transform (SIFT) detector. Bay et al. (2006) proposed the speeded up robust features (SURF) detector which uses box filters to approximate the determinant of a Hessian matrix and extracts feature points. Brox et al. (2006) applied anisotropic nonlinear diffusions to construct a nonlinear structure tensor for detecting corners. Lepetit and Fua (2006) used Laplace of Gaussian filter with multiple scales to smooth the input image and decision trees technique (Quinlan 1986) was used to extract corners. Alcantarilla et al. (2012) presented the KAZE operator which detects interest points in a nonlinear scale space by using additive operator splitting techniques (Weickert et al. 1998) to approximate the Perona and Malik diffusion equation (Perona and Malik 1990). Miao and Jiang (2013) employed the rank order Laplacian of Gaussian to smooth the input image and construct $2 \times 2$ Hessian matrix for detecting corners. Duval-Poo et al. (2015) replaced the Log-Gabor wavelet smoothing (Gao et al. 2007) by multi-scale shearlet filters and constructed a nonlinear $2 \times 2$ structure tensor for detecting corners. Verdie et al. (2015) presented a temporally invariant learned detector (TILDE) which learned from images with the same scene under drastic illumination changes. In Yi et al. (2016), SIFT method (Lowe 2004) was used to extract interest points from the input images and train an interest point detector. In Lenc and Vedaldi (2016), the local covariant constraint was used to train a feature detector. The approach of Lenc and Vedaldi (2016) was also extended by using TILDE (Verdie et al. 2015) as guidance (Zhang et al. 2017). DeTone et al. (2018) presented a self-supervised framework for training an interest point detector.

Our research indicates that the intensity variations of a corner are not significant in all directions. Our research also shows that no one has explained why the intensity variation based methods (Harris and Stephens 1988; Gårding and Lindeberg 1996; Kenney et al. 2003; Mikolajczyk and Schmid 2004; Gao et al. 2007; Duval-Poo et al. 2015) which used the first-order derivatives along the horizontal and vertical directions to construct the $2 \times 2$ structure tensor cannot detect corners well. Up to now, a large number of filters [e.g., Zhang et al. (2014), log-Gabor filters (Field 1987), shearlet filters (Duval-Poo et al. 2015), anisotropic nonlinear diffusion filters (Brox et al. 2006), and anisotropic Gaussian filters (Shui and Zhang 2012)] have been used to smooth the input image and extract intensity variations. However, within the scope of our investigations, no one has presented methods on how

to accurately extract the local intensity variations to depict the differences between edges and corners.

In this paper, the properties of the isotropic and anisotropic Gaussian directional derivative representations (Shui and Zhang 2013) of a step edge and several general corners (such as L-type, Y- or T-type, X-type, and star-type corners) are investigated to explain why the existing $2 \times 2$ structure tensor based algorithms (Noble 1988; Gårding and Lindeberg 1996; Kenney et al. 2003; Mikolajczyk and Schmid 2004; Gao et al. 2007; Duval-Poo et al. 2015) cannot detect corners well. The properties indicate that the first-order derivatives along the horizontal and vertical directions cannot depict the differences between edges and corners well. In fact, the intensity variation around a corner is not large in all directions. All the existing $2 \times 2$ structure tensor based algorithms (Noble 1988; Gårding and Lindeberg 1996; Kenney et al. 2003; Mikolajczyk and Schmid 2004; Gao et al. 2007; Duval-Poo et al. 2015) are based on Moravec's theory (Moravec 1979) that the intensity variation around corners are large in all directions, which result in false corner detections. Some corners may be detected as edges, while some edge pixels may be judged as corners. Furthermore, for a corner, it may be detected by using the two orthogonal directional derivatives. However, if the image is rotated by a certain angle, the horizontal and vertical directional derivatives of the corner may become smaller. Then the corner may not be detected.

We present a new technique to obtain the local intensity variations from the input image. We proved that the new intensity variation extraction technique has the ability to accurately depict the intensity variation differences between edges and corners in the continuous domain. The properties of the intensity variations of step edges and corners and the new intensity variation extraction technique enable us to derive a new multi-directional structure tensor with multiple scales, which has the ability to depict the differences between edges and corners well in the discrete domain. The eigenvalues of the multi-directional structure tensor with multiple scales are used in our new corner detection method. The proposed corner detector is compared with ten state-of-the-art feature detectors (Harris (Harris and Stephens 1988), Harris–Laplace (Mikolajczyk and Schmid 2004), FAST (Rosten et al. 2010), DoG (Lowe 2004), SURF (Bay et al. 2006), KAZE (Alcantarilla et al. 2012), ANDD (Shui and Zhang 2013), ACJ (Xia et al. 2014), LIFT (Yi et al. 2016), and Superpoint (DeTone et al. 2018)). Thirty images with various scenes without ground truth are used to evaluate the detectors' average repeatabilities under affine transformation, JPEG compression, and noise degradation. The Oxford dataset is used to assess the performance of the detectors on region repeatability (Mikolajczyk et al. 2005). The DTU-Robots dataset (Aanæs et al. 2012) is used to assess the performance of the detectors on repeatability metric. Two test images with ground truths are used to

assess the detection accuracy and localization accuracy of these methods. The experimental results show that the proposed method is of very high quality. This is impossible for the other tested detectors (Harris and Stephens 1988; Mikolajczyk and Schmid 2004; Rosten et al. 2010; Lowe 2004; Bay et al. 2006; Alcantarilla et al. 2012; Shui and Zhang 2013; Xia et al. 2014; Yi et al. 2016; DeTone et al. 2018).

The rest of the paper is presented as follows. In Sect. 2, the Harris detector and the representations of a step edge, L-type corner, Y- or T-type corner, X-type corner, and star-type corner are introduced. In Sect. 3, the weakness of the existing structure tensor based corner detection techniques is identified. Several edges and corner properties are summarized. Then, a new corner detection algorithm based on a multi-directional structure tensor with multiple scales is presented and a new intensity variation extraction technique is introduced. Extensive experimental results are presented in Sect. 4, and conclusions are given in Sect. 5.

## 2 Related Work

In this section, the standard Harris detection algorithm is introduced first. Then, the isotropic and anisotropic Gaussian directional derivative representations of a step edge and several general corner models are presented.

### 2.1 Harris Corner Detector

The Harris corner detector employs a $2 \times 2$ structure tensor to measure the local intensity variations of the input image along the horizontal and vertical directions. For a given 2D input image $I(x, y)$, the weighted sum of squared difference $\Im(m_x, m_y)$ is defined as

$$\Im(m_x, m_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_\sigma(x, y) \\ \left( I(x + m_x, y + m_y) - I(x, y) \right)^2 \mathrm{d}x\mathrm{d}y, \quad (1)$$

where $h_\sigma(x, y)$ is an isotropic Gaussian filter, $\sigma$ is the scale factor ($\sigma > 0$), $(x, y)$ is a point location in the image, and $(m_x, m_y)$ is a local shift. The shifted image patch $I(x + m_x, y + m_y)$ is approximated by a Taylor expansion truncated to the first order terms

$$I(x + m_x, y + m_y) \approx I(x, y) + m_x I_x(x, y) + m_y I_y(x, y), \quad (2)$$

where $I_x(x, y)$ and $I_y(x, y)$ denote the partial derivatives of the input image $I$ with respect to the horizontal and vertical

directions. Substituting approximation Eq. (2) into Equation (1) yields

$$\Im(m_x, m_y) \approx \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_\sigma(x, y)$$
$$\left(m_x I_x(x, y) + m_y I_y(x, y)\right)^2 dx dy \quad (3)$$
$$= (m_x \ m_y) A \begin{pmatrix} m_x \\ m_y \end{pmatrix},$$

where $A$ is the structure tensor

$$A = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_\sigma(x, y)$$
$$\begin{bmatrix} I_x^2(x, y) & I_x(x, y) I_y(x, y) \\ I_x(x, y) I_y(x, y) & I_y^2(x, y) \end{bmatrix} dx dy. \quad (4)$$

Typically, a corner is characterized by a large variation of $\Im$ in all directions at $(x, y)$. Let $\lambda_1$ and $\lambda_2$ ($\lambda_1 < \lambda_2$) be the eigenvalues of structure tensor $A$. There are three cases to be considered. (1) If both $\lambda_1$ and $\lambda_2$ are small, then there is no feature at pixel $(x, y)$. (2) If $\lambda_1 \approx 0$ and $\lambda_2$ is a large positive value, then an edge is found. (3) If both $\lambda_1$ and $\lambda_2$ are large positive values, then a corner is found.

## 2.2 Isotropic and Anisotropic Gaussian Directional Derivative Representations

In the spatial domain, the anisotropic Gaussian kernel (AGK) $g_{\sigma,\rho,\theta}(x, y)$ can be represented as (Zhang and Shui 2015; Shui and Zhang 2013, 2012; Zhang et al. 2017)

$$g_{\sigma,\rho,\theta}(x, y) =$$
$$\frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}[x, y]\mathbf{R}_{-\theta} \begin{bmatrix} \rho^{-2} & 0 \\ 0 & \rho^2 \end{bmatrix} \mathbf{R}_\theta[x, y]^\top\right), \quad (5)$$

with

$$\mathbf{R}_\theta = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix},$$

where $\top$ represents matrix transpose, $\rho$ is the anisotropic factor ($\rho > 1$), and $\mathbf{R}_\theta$ is the rotation matrix with angle $\theta$. From Eq. (5), the anisotropic Gaussian directional derivative (AGDD) filter $\psi_{\sigma,\rho,\theta}(x, y)$ at orientation $\theta + \pi/2$ is derived as

$$\phi_{\sigma,\rho,\theta}(x, y) = \frac{\partial g_{\sigma,\rho}}{\partial y} (\mathbf{R}_\theta[x, y]^\top). \quad (6)$$

It is worth noting that the directional derivative obtained by Eq. (6) has a $\pi/2$ shift with the directional derivative obtained by deriving the partial derivative of orientation $\theta$. If

the anisotropic factor $\rho$ is 1, $g_{\sigma,\rho,\theta}(x, y)$ and $\phi_{\sigma,\rho,\theta}(x, y)$ in Eqs. (5) and (6) represent isotropic Gaussian kernel and isotropic Gaussian directional derivative (IGDD) filter respectively.

The anisotropic Gaussian directional derivative of the input image $I(x, y)$ along direction $\theta + \pi/2$ is computed by the convolution operator

$$\nabla_{\sigma,\rho,\theta} I(x, y) = \frac{\partial}{\partial(\theta + \pi/2)}(I(x, y) \otimes g_{\sigma,\rho,\theta}(x, y))$$
$$= I(x, y) \otimes \phi_{\sigma,\rho,\theta}(x, y), \quad (7)$$

where $\otimes$ represents a convolution operation. The AGDD reflects the gray-scale intensity variation of the input image along direction $\theta + \pi/2$. It is easy to verify that

$$\nabla_{\sigma,\rho,\theta} I(x, y) = -\nabla_{\sigma,\rho,\theta+\pi} I(x, y). \quad (8)$$

It means that the interval $[0, \pi)$ for AGDDs is enough to describe the intensity variation of the input image.

In the polar coordinate system, a point function in a wedge-shaped region can be defined as (Shui and Zhang 2013)

$$\zeta_{\beta_1,\beta_2}(r, \beta)$$
$$= \begin{cases} T, & \text{if } 0 \leq r < +\infty, \ \beta_1 \leq \beta \leq \beta_2, \ \beta_2 - \beta_1 \neq \pi \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where $r$ is the radius, $\beta$ is the polar angle, $T$ is the gray value, and $\beta_1$ and $\beta_2$ are the lower and upper bounds of angle $\beta$ as shown in Fig. 1. It can be easily found that a corner point is located at the tip $o$ of the wedge-shaped region. In this paper, the point function in a wedge-shaped region is named as a basic corner model. A similar corner model is also presented in Deriche and Giraudon (1993).
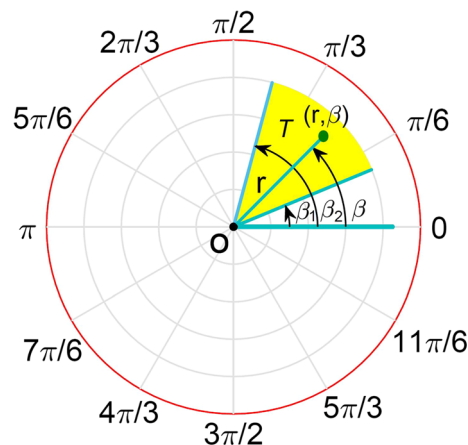


**Fig. 1** Examples of a basic corner model in the polar coordinate system

The general corner model (e.g., L-type corner, Y- or T-type corner, X-type corner, and star-type corner) can be derived by the several basic corner models as follows

$$\hbar_{(T_i,\beta_i)}(r,\beta) = \sum_{i=1}^{s} T_i \zeta_{\beta_i,\beta_{i+1}}(r,\beta), \qquad (10)$$

where $T_i$ represents the gray value of the $i$-th wedge-shaped region. $s$ is the number of wedge-shaped regions. It is noted that $\beta_{s+1} = \beta_1$. If $s = 2$ and $\beta_2 - \beta_1 = \pi$, Eq. (10) represents a step edge. If $s = 2$ and $\beta_2 - \beta_1 \neq \pi$, Eq. (10) corresponds to an L-type corner. If $s = 3$, Eq. (10) represents a Y- or T-type corner. If $s = 4$, Eq. (10) represents an X-type corner. If $s = 5$, Eq. (10) corresponds to a star-type corner.

The AGDD representation of the basic corner model is (Shui and Zhang 2013)

$$\xi_{\sigma,\rho}(\theta) = \iint_{\mathbb{R}^2} \zeta_{\beta_1,\beta_2}(r,\beta)\phi_{\sigma,\rho,\theta}(-r,-\beta)rdrd\beta$$
$$= \frac{T\rho}{2\sqrt{2\pi}\sigma}\left(\frac{\cos(\beta_1-\theta)}{(\rho^4\sin^2(\beta_1-\theta)+\cos^2(\beta_1-\theta))^{\frac{1}{2}}}\right.$$
$$\left. - \frac{\cos(\beta_2-\theta)}{(\rho^4\sin^2(\beta_2-\theta)+\cos^2(\beta_2-\theta))^{\frac{1}{2}}}\right), \qquad (11)$$

where $\mathbb{R}^2$ represents the 2D real space and $\psi_{\sigma,\rho,\theta}(r,\beta)$ represents the AGDD filter in the polar coordinate system. Then, the AGDD representation of the general corner model is

$$\Lambda_{\sigma,\rho}(\theta)$$
$$= \frac{\rho}{2\sqrt{2\pi}\sigma}\sum_{i=1}^{s} T_i\left(\frac{\cos(\beta_i-\theta)}{(\rho^4\sin^2(\beta_i-\theta)+\cos^2(\beta_i-\theta))^{\frac{1}{2}}}\right.$$
$$\left. - \frac{\cos(\beta_{i+1}-\theta)}{(\rho^4\sin^2(\beta_{i+1}-\theta)+\cos^2(\beta_{i+1}-\theta))^{\frac{1}{2}}}\right). \qquad (12)$$

With $\rho = 1$, Eq. (12) is the IGDD representation of the general corner model

$$\kappa_{\sigma,\rho}(\theta) = \frac{1}{2\sqrt{2\pi}\sigma}\sum_{i=1}^{s} T_i\left(\cos(\beta_{i+1}-\theta)-\cos(\beta_i-\theta)\right)$$
$$= \frac{1}{\sqrt{2\pi}\sigma}\sum_{i=1}^{s} T_i\sin\left(\theta-\frac{\beta_i+\beta_{i+1}}{2}\right)\sin\left(\frac{\beta_{i+1}-\beta_i}{2}\right), \qquad (13)$$

which means that all the IGDD representations of a step edge and general corners are sine functions.

Examples of the AGDD and IGDD representations of the step edge and general corners are shown in Fig. 2. The step edge, L-type, Y- or T-type, X-type, and star-type corner

models are illustrated in Fig. 2a–e respectively. Their corresponding intensity variations of the AGDD representations are shown in the second column. Their corresponding intensity variations of the IGDD representations are shown in the third column.

# 3 Proposed Method

In this section, the problems of the existing $2 \times 2$ structure tensor based corner detection methods are demonstrated and several corner intensity variation properties are summarized. Then, a new multi-directional structure tensor with multiple scales is derived for corner detection. Finally, a new image intensity variation extraction technique is presented.

## 3.1 Corner Properties

As shown in Fig. 2b-e, for the L-type, the Y- or T-type, the X-type, and the star-type corner models, it is obvious from the second and third columns of Fig. 2 that the directional derivatives are large in most directions at a corner while their directional derivatives of the AGDD or the IGDD representations are very small or even near zero along the horizontal (0) or vertical ($\pi/2$) directions. Then, these corners may not be correctly detected by the Harris detector. This phenomenon does not satisfy the definition for a corner (Moravec 1979) that the directional derivatives are large in all directions. Furthermore, along the horizontal and vertical directions, the directional derivatives of the input image cannot accurately depict the intensity variation differences between edges and corners. Take a step edge, an L-type corner, and an X-type corner as examples as shown in the first column of Fig. 2, their corresponding directional derivatives are zero in the horizontal direction. However, in the vertical direction, the absolute magnitude of the directional derivative on the edge is larger than that of the L-type and X-type corners. Then, edges may be detected as corners by Eq. (4), while the real corners may be marked as edges. The reason is that the existing intensity variation based methods (Harris and Stephens 1988; Noble 1988; Gårding and Lindeberg 1996; Kenney et al. 2003; Mikolajczyk and Schmid 2004; Gao et al. 2007; Duval-Poo et al. 2015) do not take the directional derivative differences in different filtering orientations between edges and corners fully into account.

We found from Eq. (13) that the isotropic Gaussian representations cannot depict the intensity variation differences between a step edge and general corners as shown in the third column of Fig. 2. However, we found from Eq. (12) that the anisotropic Gaussian representations have the ability to accurately depict the intensity variation differences between a step edge and general corners. As shown in the second column of Fig. 2, a step edge with $\beta_1 = \pi/2$, $\beta_2 = 3\pi/2$,
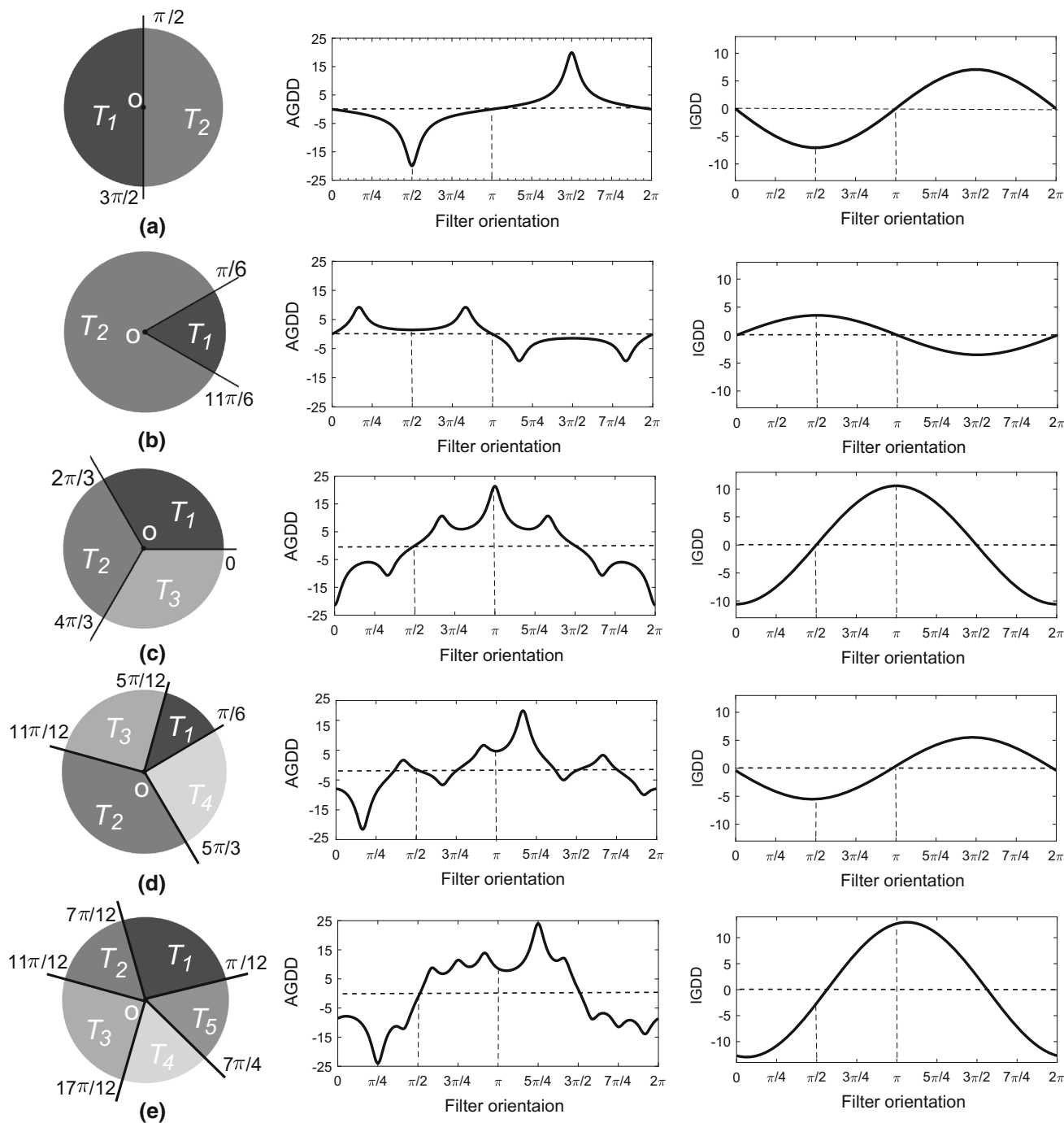
**Fig. 2** A step edge, L-type, Y-type, X-type, and star-type corner models are shown in **a**–**e** in the first column (gray values $T_1 = 50$, $T_2 = 100$, $T_3 = 150$, $T_4 = 200$, and $T_5 = 120$). Their corresponding directional derivatives of the AGDD representations ($\rho^2 = 8$, $\sigma^2 = 8$) and IGDD representations ($\rho^2 = 1$, $\sigma^2 = 8$) are shown in the second and third columns respectively

$T_1 = 50$, and $T_2 = 100$ has only one local maximum for a directional derivative at $\theta = 3\pi/2$ and one local minimum for a directional derivative at $\theta = \pi/2$. For an L-type corner with $\beta_1 = 11\pi/6$, $\beta_2 = \pi/6$, $T_1 = 50$, and $T_2 = 100$, it has two local maxima for directional derivatives at $\theta = \pi/6$ and $\theta = 5\pi/6$ and two local minima for directional derivatives

at $\theta = 7\pi/6$ and $\theta = 11\pi/6$. For a Y- or T-type corner with $\beta_1 = 0$, $\beta_2 = 2\pi/3$, $\beta_3 = 4\pi/3$, $T_1 = 50$, $T_2 = 100$, and $T_3 = 150$, it has three local maxima for directional derivatives at $\theta = 2\pi/3$, $\theta = \pi$, and $\theta = 4\pi/3$ and three local minima for directional derivatives at $\theta = 0$, $\theta = \pi/3$, and $\theta = 5\pi/3$. For an X-type corner, it has four local maxima
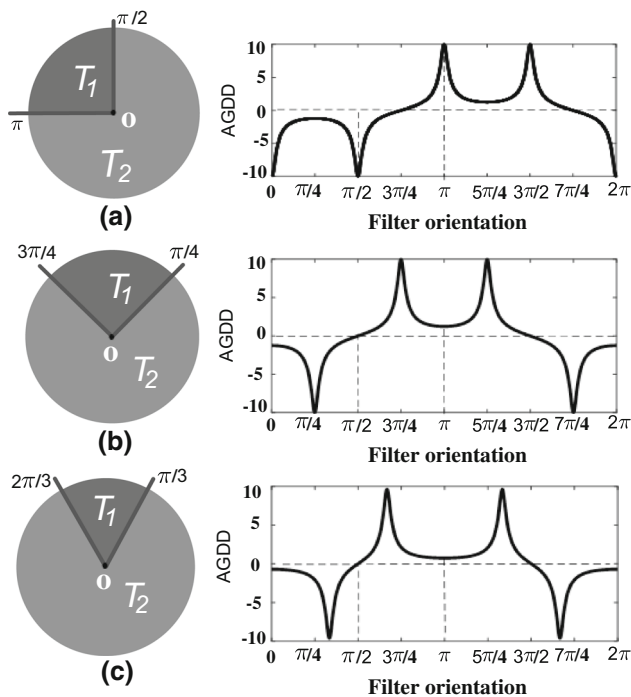
**Fig. 3** Examples of directional derivative changes by image rotation or affine image transformation with $\rho^2 = 16$ and $\sigma^2 = 4$. **a** An L-type corner with gray values $T_1 = 50$ and $T_2 = 100$, **b** The L-type corner is rotated by $\pi/4$ clockwise, **c** the L-type corner undergoes an affine image transformation

and four local minima for directional derivatives. For a star-type corner, it has five local maxima and five local minima for directional derivatives.

Furthermore, our researches indicate that two orthogonal directional derivatives along the horizontal and vertical directions cannot accurately detect corners on an affine transformed image. Take an L-type corner as an example as shown in Fig. 3a, its corresponding two orthogonal directional derivatives are large from Eq. (12) as shown in the second column of Fig. 3. According to the criteria of Harris corner detection, it can be detected as a corner. After the L-type corner is rotated by $\pi/4$ clockwise as shown in Fig. 3b, its corresponding two orthogonal directional derivatives are small from Eq. (12) as shown in the second column of Fig. 3. Then the corner may not be detected with such or similar image rotation transformations. The reason is that the two orthogonal directional derivatives do not contain enough local structure information. The existing multi-scale filtering techniques (Gårding and Lindeberg 1996; Lindeberg 1998; Schmid et al. 2000; Mikolajczyk and Schmid 2004; Laptev 2005; Lowe 2004; Bay et al. 2006; Brox et al. 2006; Gao et al. 2007; Marimon et al. 2010; Alcantarilla et al. 2012; Miao and Jiang 2013; Duval-Poo et al. 2015; Perona and Malik 1990; Wang 1999; Widynski and Mignotte 2014) cannot solve the aforementioned problem because the multi-scale filtering

technique only efficiently enhance the local intensity variation extraction along the horizontal and vertical directions. Another example is when the image is rotated and squeezed, which means that the shape of the corner is changed. If the L-type corner undergoes an affine image transformation as shown in Fig. 3c, its corresponding two orthogonal directional derivatives are also small from Eq. (12) as shown in the second column of Fig. 3. Then the corner may not be detected with such or similar affine image transformations. The existing multi-scale filtering techniques cannot solve the aforementioned problem either.

Based on the above analysis, several properties of corners are summarized as follows:

**Property 1** The intensity variation of a corner is large in most directions, not necessarily in all directions.

**Property 2** The first-order derivatives along the horizontal and vertical directions cannot depict the intensity variations of step edges and corners well.

**Property 3** The isotropic Gaussian filter cannot depict the intensity variation differences between step edges and corners accurately.

**Property 4** The anisotropic Gaussian filters have the ability to depict the intensity variation differences between step edges and corners.

**Property 5** The existing $2 \times 2$ structure tensor based techniques may not depict the differences between step edges and corners accurately.

The above properties will help us propose a new corner measure, a new corner detection algorithm, and a new image intensity variation extraction technique which will be presented in the following section.

### 3.2 Multi-directional Structure Tensor with Multiple Scales

Based on the aforementioned analysis, it can be concluded that the intensity variation based methods (Harris and Stephens 1988; Gårding and Lindeberg 1996; Kenney et al. 2003; Mikolajczyk and Schmid 2004; Miao and Jiang 2013; Gao et al. 2007) which used the first-order derivatives along the horizontal and vertical directions to construct the $2 \times 2$ structure tensor cannot detect corners well. In this section, the multi-scale and multi-directional anisotropic Gaussian filters are used as an example to explain how to detect corners using multi-scale and multi-directional intensity variation information.

Images are 2D discrete signals in the integer lattice $\mathbb{Z}^2$, and the continuous AGKs and AGDD filters in Eqs. (5) and (6) need to be discretized in $\mathbb{Z}^2$. Given multi-scales $\sigma_s$ (e.g.,

$s = 1, 2, 3$), an anisotropic factor $\rho$, and $K$ oriented angles $\theta_k = (k-1)\pi/K$ ($k = 1, 2, \ldots, K$), the discrete version of the functions for multi-directional AGKs $g_{\sigma_s,\rho,k}(x, y)$ and AGDD $\phi_{\sigma_s,\rho,k}(x, y)$ with multiple scales are below

$$g_{\sigma_s,\rho,k}(\mathbf{n}) = \frac{1}{2\pi\sigma_s^2} \exp\left(-\frac{1}{2\sigma_s^2}\mathbf{n}^\top \mathbf{R}_{-k} \begin{bmatrix} \rho^{-2} & 0 \\ 0 & \rho^2 \end{bmatrix} \mathbf{R}_k \mathbf{n}\right),$$

$$\phi_{\sigma_s,\rho,k}(\mathbf{n}) = \frac{-\rho^2[-\sin\theta_k \ \cos\theta_k]\mathbf{n}}{\sigma_s^2} g_{\sigma_s,\rho,k}(\mathbf{n}), \tag{14}$$

with

$$\mathbf{R}_k = \begin{bmatrix} \cos\theta_k & \sin\theta_k \\ -\sin\theta_k & \cos\theta_k \end{bmatrix}, \quad \mathbf{n} = \begin{bmatrix} n_x \\ n_y \end{bmatrix} \in \mathbb{Z}^2,$$

where $(n_x, n_y)$ represents the pixel coordinate in the integer lattice $\mathbb{Z}^2$.

Given the multi-directional anisotropic Gaussian filters $g_{\sigma_s,\rho,k}(n_x, n_y)$ with multi-scales $\sigma_s$, the discrete weighted sum of square differences $\Im_s(n_x, n_y)$ of point $(n_x, n_y)$ is redefined as Eq. (15), where $(n_x + i, n_y + j)$ is a point in an image patch over an area with width $u + 1$ and height $v + 1$ centered on $(n_x, n_y)$, $\triangle t$ is a shift at point $I(n_x+i, n_y+j)$, $\theta_k$ is the angle between the horizontal axis and the $k$-th oriented vector. In this paper, the size of $(u + 1) \times (v + 1)$ is set to $7 \times 7$.

$$\Im_s(n_x, n_y) = \frac{\pi}{K(u+1)(v+1)} \sum_{i=-\frac{u}{2}}^{\frac{u}{2}} \sum_{j=-\frac{v}{2}}^{\frac{v}{2}} \sum_{k=1}^{K}$$
$$g_{\sigma_s,\rho,k}(n_x + i, n_y + j)\otimes$$
$$\left(I(n_x + i + \triangle t\cos\theta_k, n_y + j + \triangle t\sin\theta_k)\right.$$
$$\left. - I(n_x + i, n_y + j)\right)^2, \tag{15}$$

$I(n_x + i + \triangle t\cos\theta_k, n_y + j + \triangle t\sin\theta_k)$ can be approximated by a Taylor expansion as

$$I(n_x + i + \triangle t\cos\theta_k, n_y + j + \triangle t\sin\theta_k)$$
$$\approx I(n_x + i, n_y + j) + \triangle t I_k(n_x + i, n_y + j), \tag{16}$$

where $I_k(n_x+i, n_y+j)$ is the directional derivative of $I(n_x + i, n_y + j)$ in the direction of $\theta_k$. Substituting approximation Eq. (16) into Eq. (15) yields Eq. (17).

$$\Im_s(n_x, n_y) \approx \frac{\pi}{K(u+1)(v+1)} \sum_{i=-\frac{u}{2}}^{\frac{u}{2}} \sum_{j=-\frac{v}{2}}^{\frac{v}{2}} \sum_{k=1}^{K}$$
$$g_{\sigma_s,\rho,k}(n_x + i, n_y + j)$$
$$\otimes (\triangle t I_k(n_x + i, n_y + j))^2. \tag{17}$$

It is worth to note that

$$\nabla_{\sigma_s,\rho,k} I(n_x + i, n_y + j)$$
$$= g_{\sigma_s,\rho,k}(n_x + i, n_y + j) \otimes I_k(n_x + i, n_y + j). \tag{18}$$

As a result, Eq. (17) can be rewritten as Eq. (19), where $M$ is a multi-directional structure tensor at multiple scales which is a symmetric $K \times K$ matrix as given in Eq. (20). From Eq. (20), it can be easily concluded that the eigenvalues of matrix $M$ are determined by scale $\sigma_s$, the anisotropic factor $\rho$, and the number of orientations $K$ of the anisotropic Gaussian filters.

$$\Im_s(n_x, n_y) \approx \frac{\pi}{K(u+1)(v+1)} \sum_{i=-\frac{u}{2}}^{\frac{u}{2}} \sum_{j=-\frac{v}{2}}^{\frac{v}{2}}$$
$$\left([\nabla_{\sigma_s,\rho,1}I(n_x+i, n_y+j), \nabla_{\sigma_s,\rho,2}I(n_x+i, n_y+j), \ldots,\right.$$
$$\left.\nabla_{\sigma_s,\rho,K}I(n_x + i, n_y + j)][\triangle t, \triangle t, \ldots, \triangle t]^\top\right)^2$$
$$= \frac{\pi}{K(u+1)(v+1)}(\triangle t \ \triangle t \ \ldots \ \triangle t)M \begin{pmatrix} \triangle t \\ \triangle t \\ \vdots \\ \triangle t \end{pmatrix}, \tag{19}$$

$$M = \begin{bmatrix} \sum\limits_{i=-\frac{u}{2}}^{\frac{u}{2}} \sum\limits_{j=-\frac{v}{2}}^{\frac{v}{2}} \nabla_{\sigma_s,\rho,1}^2 I(n_x + i, n_y + j) & \cdots & \sum\limits_{i=-\frac{u}{2}}^{\frac{u}{2}} \sum\limits_{j=-\frac{v}{2}}^{\frac{v}{2}} \nabla_{\sigma_s,\rho,1} I(n_x + i, n_y + j)\nabla_{\sigma_s,\rho,K} I(n_x + i, n_y + j) \\ \vdots & \ddots & \vdots \\ \sum\limits_{i=-\frac{u}{2}}^{\frac{u}{2}} \sum\limits_{j=-\frac{v}{2}}^{\frac{v}{2}} \nabla_{\sigma_s,\rho,K} I(n_x + i, n_y + j)\nabla_{\sigma_s,\rho,1} I(n_x + i, n_y + j) & \cdots & \sum\limits_{i=-\frac{u}{2}}^{\frac{u}{2}} \sum\limits_{j=-\frac{v}{2}}^{\frac{v}{2}} \nabla_{\sigma_s,\rho,K}^2 I(n_x + i, n_y + j) \end{bmatrix} \tag{20}$$

## 3.3 Corner Measure and Corner Detection Algorithm

In this section, a new corner measure and a new corner detection algorithm are presented as follows.

In this paper, $K$ eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_K\}$ of the $K \times K$ multi-directional structure tensor at each scale are used to form a new corner measure to distinguish corners from other points in the input image. The new corner measure is defined as

$$\wp_s(n_x, n_y) = \frac{\prod\limits_{k=1}^{K} \lambda_k}{\sum\limits_{k=1}^{K} \lambda_k + \tau}, \tag{21}$$

where $\tau$ is a small constant ($\tau = 2.22 \times 10^{-16}$) which is used to avoid a singular denominator in the case of a rank zero structure tensor. For each image pixel $(n_x, n_y)$, it is marked as a corner if its corresponding $\wp_s(n_x, n_y)$ is a local maximum within a $7 \times 7$ window and is larger than a threshold $T_h$ at each scale $\sigma_s$ ($s = 1, 2, 3$).

In general, the new corner measure has the following advantages over the existing $2 \times 2$ structure tensor based methods (Noble 1988; Gårding and Lindeberg 1996; Kenney et al. 2003; Mikolajczyk and Schmid 2004; Gao et al. 2007; Duval-Poo et al. 2015). The proposed corner measure has the ability to accurately detect corners, and it is also robust for corner detection with image affine transformations. This is impossible for the existing corner detectors (Deriche and Giraudon 1993; Smith and Brady 1997; Rosten et al. 2010; Shui and Zhang 2013; Xia et al. 2014; Moravec 1979; Harris and Stephens 1988; Noble 1988; Gårding and Lindeberg 1996; Lindeberg 1998; Schmid et al. 2000; Kenney et al. 2003; Mikolajczyk and Schmid 2004; Laptev 2005; Lowe 2004; Bay et al. 2006; Marimon et al. 2010; Maver 2010; Su et al. 2012).

The proposed corner detection method is described as follows:

1. Use the multi-directional anisotropic Gaussian filters at multi-scales to smooth the input image, and derive the multi-direction directional derivatives at multi-scales as in Eq. (7).
2. For each image pixel, construct the multi-directional structure tensor at multi-scales as in Eq. (20).
3. Obtain the eigenvalues at each scale based on Eq. (21).
4. Mark the pixel as a candidate corner if its corresponding corner measure is the local maximum within a window ($7 \times 7$) and is larger than the threshold $T_h$ at the lowest scale.

5. Mark the candidate corner as a corner if its corresponding corner measure is larger than the threshold $T_h$ at all the scales.

## 3.4 A New Image Intensity Variation Extraction Technique

In this subsection, our aim is to present a new image intensity variation extraction technique to accurately depict the intensity variation differences between edges and corners.

From Eq. (10), the step edge, the L-type, the Y- or T-type, the X-type, and the star-type corners can be represented by the sum of several basic corner models as described by Eq. (9). From Eq. (12), we derived that step edge, L-type, Y- or T-type, X-type, and star-type corners have only one, two, three, four, and five local maxima for the first-order anisotropic Gaussian directional derivative respectively. Noting that each local maximum of the first-order derivatives corresponds to a local minimum of the first-order derivatives. It means that if the extracted image intensity variations have the ability to describe the number of the maximum points of the first-order anisotropic Gaussian directional derivatives, the extracted image intensity variations have the ability to depict the characteristics of edges and corners. Then for each AGDD representation of the basic corner model as given in Eq. (11), if its corresponding two local maxima on the directional derivatives can be identified, the extracted local intensity variation information has the ability to depict the intensity variation differences between edges and corners. In what follows, we discuss how to design the anisotropic Gaussian filters to exactly identify the two local maxima on the directional derivatives of the AGDD representation of the basic corner model.

The AGDD representation of the basic corner model is shown in Eq. (11). Without loss of generality, let $\beta_2 - \beta_1 \in (0, \pi)$. A basic corner with $\beta_1 = -\pi/6$, $\beta_2 = \pi/3$, and $T = 50$ is selected as an example, and the directional derivatives of the basic corner model is shown in Fig. 4. For the basic corner model, it is easy to verify that the two local maxima on the directional derivative curve are at $\beta_2 + \pi$ and $\beta_1 + 2\pi$ as shown in Fig. 4. The angle difference between the two local maxima is $\pi - (\beta_2 - \beta_1)$. Then, only if $\theta$ equals $(\beta_1 + \beta_2 + 3\pi)/2$, there exists a local minimum on the directional derivatives when the two local maxima can be distinguished.

The first-order AGDD representation of the basic corner model is

$$\xi'_{\sigma,\rho}(\theta) = \frac{T}{2\sqrt{2\pi}\sigma} \left( \frac{\rho^2 \sin(\beta_1 - \theta)}{(\rho^2 \sin^2(\beta_1-\theta) + \rho^{-2}\cos^2(\beta_1-\theta))^{\frac{3}{2}}} - \frac{\rho^2 \sin(\beta_2 - \theta)}{(\rho^2 \sin^2(\beta_2 - \theta) + \rho^{-2}\cos^2(\beta_2 - \theta))^{\frac{3}{2}}} \right). \tag{22}$$
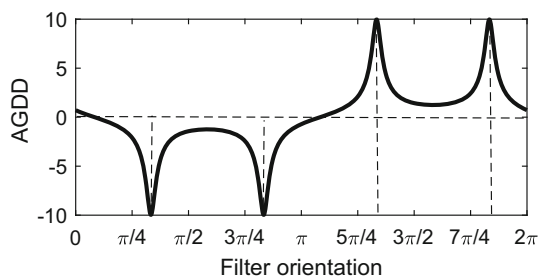
**Fig. 4** Examples of directional derivatives of the basic corner model with $\beta_1 = -\pi/6$, $\beta_2 = \pi/3$, $T = 50$, $\rho^2 = 4$, and $\sigma^2 = 4$

The second-order AGDD representation of the basic corner model is

$$\xi''_{\sigma,\rho}(\theta) = \frac{T}{2\sqrt{2\pi}\sigma} \left( \frac{(2(\rho^4-1)\sin^2(\beta_1-\theta)-1)\cos(\beta_1-\theta)}{(\rho^2\sin^2(\beta_1-\theta)+\rho^{-2}\cos^2(\beta_1-\theta))^{\frac{5}{2}}} \right.$$
$$\left. - \frac{(2(\rho^4-1)\sin^2(\beta_2-\theta)-1)\cos(\beta_2-\theta)}{(\rho^2\sin^2(\beta_2-\theta)+\rho^{-2}\cos^2(\beta_2-\theta))^{\frac{5}{2}}} \right). \quad (23)$$

If $\xi_{\sigma,\rho}(\frac{\beta_1+\beta_2+3\pi}{2})$ is a local minimum on the directional derivatives, its corresponding first-order and second-order derivatives should satisfy

$$\xi'_{\sigma,\rho}\left(\frac{\beta_1+\beta_2+3\pi}{2}\right) = 0,$$
$$\xi''_{\sigma,\rho}\left(\frac{\beta_1+\beta_2+3\pi}{2}\right) > 0. \quad (24)$$

When $\theta$ equals $\frac{\beta_1+\beta_2+3\pi}{2}$, we can conclude from Eq. (22) that $\xi'_{\sigma,\rho}(\theta)$ is 0, and its corresponding second-order derivative $\xi''_{\sigma,\rho}(\theta)$ is

$$\xi''_{\sigma,\rho}\left(\frac{\beta_1+\beta_2+3\pi}{2}\right)$$
$$= \frac{T}{\sqrt{2\pi}\sigma} \frac{\left(2(\rho^4-1)\cos^2(\frac{\beta_2-\beta_1}{2})-1\right)\sin(\frac{\beta_2-\beta_1}{2})}{\left(\rho^2\cos^2(\frac{\beta_2-\beta_1}{2})+\rho^{-2}\sin^2(\frac{\beta_2-\beta_1}{2})\right)^{\frac{5}{2}}}. \quad (25)$$

From Eq. (25), it can be derived that $\xi_{\sigma,\rho}(\frac{\beta_1+\beta_2+3\pi}{2})$ is the local minimum on the gradient magnitude responses if it satisfies

$$2(\rho^4-1)\cos^2\left(\frac{\beta_2-\beta_1}{2}\right) - 1 > 0. \quad (26)$$

Inequality (26) holds if the following is satisfied

$$\rho^4 > 1 + \frac{1}{2\cos^2\left(\frac{\beta_2-\beta_1}{2}\right)}. \quad (27)$$

When $\beta_2 - \beta_1 = 0$, the right-hand side of inequality (27) gives the minimum $\frac{3}{2}$. For a given anisotropic factor $\rho^2 > \frac{\sqrt{6}}{2}$, the two local maxima on the directional derivatives can be resolved only when the angle $\beta_2 - \beta_1$ satisfies

$$0 < \beta_2 - \beta_1 < 2\arccos\left(\frac{1}{\sqrt{2(\rho^4-1)}}\right). \quad (28)$$

Inequality (28) can be further written as inequality (29)

$$\pi - 2\arccos\left(\frac{1}{\sqrt{2(\rho^4-1)}}\right) < \pi - (\beta_2-\beta_1) < \pi. \quad (29)$$

It is worth to note that $\beta_2-\beta_1$ is the range of $\beta$ for the basic corner model (9) and $\pi - (\beta_2-\beta_1)$ is the angle difference between the two local maxima on the directional derivative of the basic corner. From inequality (28) and inequality (29), it can be concluded that the larger the anisotropic factor, the more the local intensity variation information that can be extracted by the anisotropic Gaussian filters which have a stronger ability to distinguish adjacent local maxima on the directional derivatives. We note that the L-type, Y- or T-type, X-type, and star-type corners can be represented by the sum of several basic corner models. Then, if all the angles of the basic corner models satisfy with inequality (28), it means that the obtained intensity variation information has the ability to describe the number of the maximum points of the first-order anisotropic Gaussian directional derivatives, and it also means that the extracted local intensity variation information can accurately depict the intensity variation differences between edges and corners.

## 4 Experimental Results and Performance Evaluation

The proposed corner detector is compared with ten state-of-the-art detectors [Harris (Harris and Stephens 1988), Harris–Laplace (Mikolajczyk and Schmid 2004), FAST (Rosten et al. 2010), DoG (Lowe 2004), SURF (Bay et al. 2006), KAZE (Alcantarilla et al. 2012), ANDD (Shui and Zhang 2013), ACJ (Xia et al. 2014), LIFT (Yi et al. 2016), and Superpoint (DeTone et al. 2018)]. Thirty images (Bowyer et al. 1999) are used to evaluate the average repeatabilities (Awrangjeb and Lu 2008) of these detectors. The Oxford dataset is used to assess the region repeatability (Mikolajczyk et al. 2005) of these detectors. The DTU-Robots dataset (Aanæs et al. 2012) that contains 3D objects under changing viewpoints is used to evaluate the repeatability metric of these detectors. Furthermore, two test images with ground truths are used to assess the detection capability and localization accuracy of these methods. Execution time,

**Fig. 5** Test images

memory usage, and 3D reconstruction from large scale structure from motion dataset are also investigated.

The original codes for seven of these detectors in Rosten et al. (2010), Bay et al. (2006), Alcantarilla et al. (2012), Shui and Zhang (2013), Xia et al. (2014), Yi et al. (2016), DeTone et al. (2018) are from the authors. The codes for the Harris–Laplace (Mikolajczyk and Schmid 2004) and DoG (Lowe 2004) detectors are from http://www.robots.ox.ac.uk/vgg/affine/. The code for the Harris detector (Harris and Stephens 1988) is from http://peterkovesi.com/matlabfns/. The parameter settings for the proposed detector are: $\rho^2 = 1.5$, $\sigma_1^2 = 1.5$, $\sigma_2^2 = 3$, $\sigma_3^2 = 4.5$, $K = 8$, $(u+1) \times (v+1) = 7 \times 7$, and $T_h = 1.0 \times 10^7$. The program or web demos of the proposed method can be accessed at http://vision-cdc.csiro.au/corner1st/. The selection of the parameters for the proposed method will be discussed in Sect. 4.1.

### 4.1 Repeatability Under Affine Transformation

In Awrangjeb and Lu (2008), the average repeatability $R_{\mathrm{avg}}$ measures the average number of the repeated corners between the original and affine transformed images. It is defined as

$$R_{\mathrm{avg}} = \frac{N_r}{2}\left(\frac{1}{N_o} + \frac{1}{N_t}\right), \tag{30}$$

where $N_o$ and $N_t$ are the numbers of detected corners from the original and transformed images by a detector, and $N_r$ is the number of repeated corners between them. If a corner is detected in a geometrically transformed image, and it is in the neighbourhood of the ground truth location (say within 4 pixels), then a repeated corner is detected. A higher

average repeatability means a better performance. Thirty images (Bowyer et al. 1999) with different scenes as shown in Fig. 5 are used for measuring the average repeatability for the detectors.

We followed the criteria standard (Awrangjeb and Lu 2008) that a total of 6,510 transformed test images were obtained by applying the following six different types of transformations on each original image:

- Rotation: The original image was rotated at $10°$ apart within $[-\pi/2, \pi/2]$, excluding $0°$.
- Uniform scaling: The scale factors $s_x = s_y$ are in $[0.5, 2]$ with 0.1 apart, excluding 1.
- Non-uniform scaling: The scale $s_x$ is in $[0.7, 1.5]$ and $s_y$ is in $[0.5, 1.8]$ with 0.1 apart, excluding the case when $s_x = s_y$.
- Shear transformations: The shear factor $c$ was chosen by sampling the range $[-1, 1]$ with a 0.1 interval, excluding 0, with the following formula

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

- Lossy JPEG compression: A compression factor is in $[5, 100]$ at 5 apart.
- Gaussian noise: Zero mean white Gaussian noise was added to the original image at 15 standard deviations in $[1, 15]$ with an interval 1.

From inequality (28) and inequality (29), it is concluded that the larger the anisotropic factor, the higher the potential to extract the intensity variation information to depict the intensity variation differences between step edges and cor-

**Table 1** Average repeatability of the proposed method

| Proposed method with different parameter settings | Repeatability (percentage) |
| --- | --- |
| $K = 2$, $\rho^2 = 1.5$ | 0.771 |
| $K = 4$, $\rho^2 = 1.5$ | 0.807 |
| $K = 8$, $\rho^2 = 1.5$ | 0.827 |
| $K = 16$, $\rho^2 = 1.5$ | 0.821 |
| $K = 8$, $\rho^2 = 1.2$ | 0.813 |
| $K = 8$, $\rho^2 = 2$ | 0.802 |
| $K = 8$, $\rho^2 = 2.5$ | 0.781 |

ners. Meanwhile, it is proved in Shui and Zhang (2012) that the variance $\varepsilon_w^2$ of the image Gaussian noise smoothed by the AGDD filters is $\varepsilon_w^2 = \frac{\rho^2 \epsilon^2}{8\pi\sigma^4}$. It means that the noise response of an AGDD filter is proportional to the noise variance and to the square of the anisotropic factor and inversely proportional to the power of four of the scale factor. Considering the use of the extracted intensity variation information to depict the intensity variation differences between step edges and corners and the capability on noise suppression, the scale factors with $\sigma_1^2 = 1.5$, $\sigma_2^2 = 3$, and $\sigma_3^2 = 4.5$ are used in the proposed detector. The next step is to discuss the selection of the number of directions and the anisotropic factor.

In this evaluation criteria, we firstly fix the anisotropic factor with $\rho^2 = 1.5$ to check the average repeatability of the proposed methods with different number of directions. It can be observed from Table 1 that the proposed method achieves the best performance when $K$ is 8. Secondly, we fix the direction number $K = 8$ to check the average repeatability of the proposed method with different anisotropic factors. It can be observed from Table 1 that with $K = 8$, the proposed method achieves the best performance when $\rho^2$ is 1.5. From this experiment, we found that the different numbers of directions have a great influence on the performance under image rota-

tion transformations, as shown in Fig. 6a. With $K = 2$, the performance of the proposed method drops dramatically in the case of image rotation transformation. The reason is that the AGDD filters with two directions cannot extract enough intensity variation information and cannot accurately detect corners with image rotation transformations. Meanwhile, we can also found that the anisotropic factor has a great influence on the performance of the proposed method under image lossy JPEG compression and additive white Gaussian noises as shown in Fig. 6b, c. With anisotropic factor $\rho^2 = 2.5$, the performances of the proposed method drop dramatically in the cases of image lossy JPEG compression and additive white Gaussian noises. The reason is that the large anisotropic factor will reduce the ability of AGDD filters to suppress the Gaussian noise. Based on the aforementioned analysis, the direction number with $K = 8$ and the anisotropic factor with $\rho^2 = 1.5$ are used in the proposed detector.

Then, the proposed approach with the fixed parameter setting has been compared with the ten other detectors (Harris and Stephens 1988; Mikolajczyk and Schmid 2004; Rosten et al. 2010; Lowe 2004; Bay et al. 2006; Alcantarilla et al. 2012; Shui and Zhang 2013; Xia et al. 2014; Yi et al. 2016; DeTone et al. 2018. The results with different rotations, uniform scalings, non-uniform scalings, shear transformation, lossy JPEG compression, and Gaussian noises are shown in Fig. 7. It can be observed that the proposed detector achieves the best performance under this evaluation criteria.

## 4.2 Repeatability Score Under Region Repeatability Evaluation

In http://www.robots.ox.ac.uk/vgg/affine/, each of the image sequences used in the evaluation contains six images of naturally textured scenes with increasingly geometric and photometric transformations. The images in a sequence are related by a homography which is provided with the image
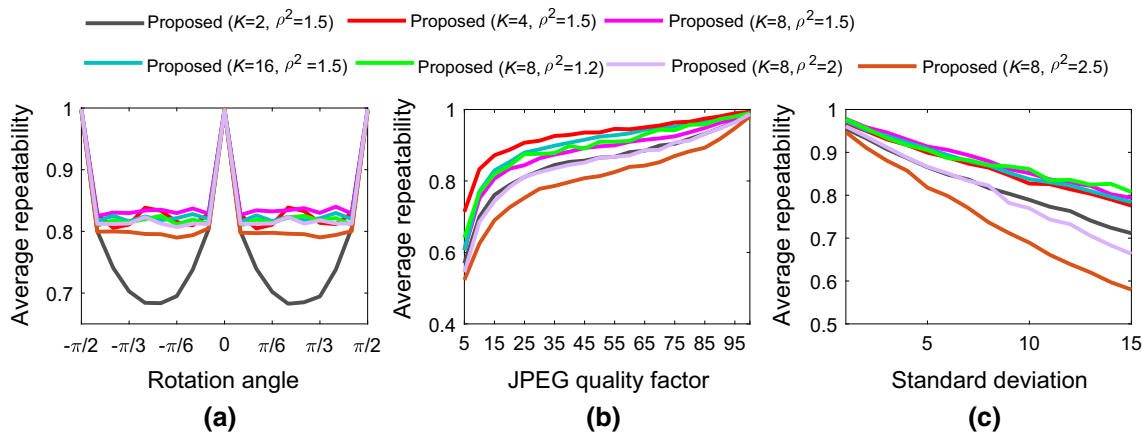


**Fig. 6** Average repeatability of the proposed method under rotation, lossy JPEG compression, and additive white Gaussian noises
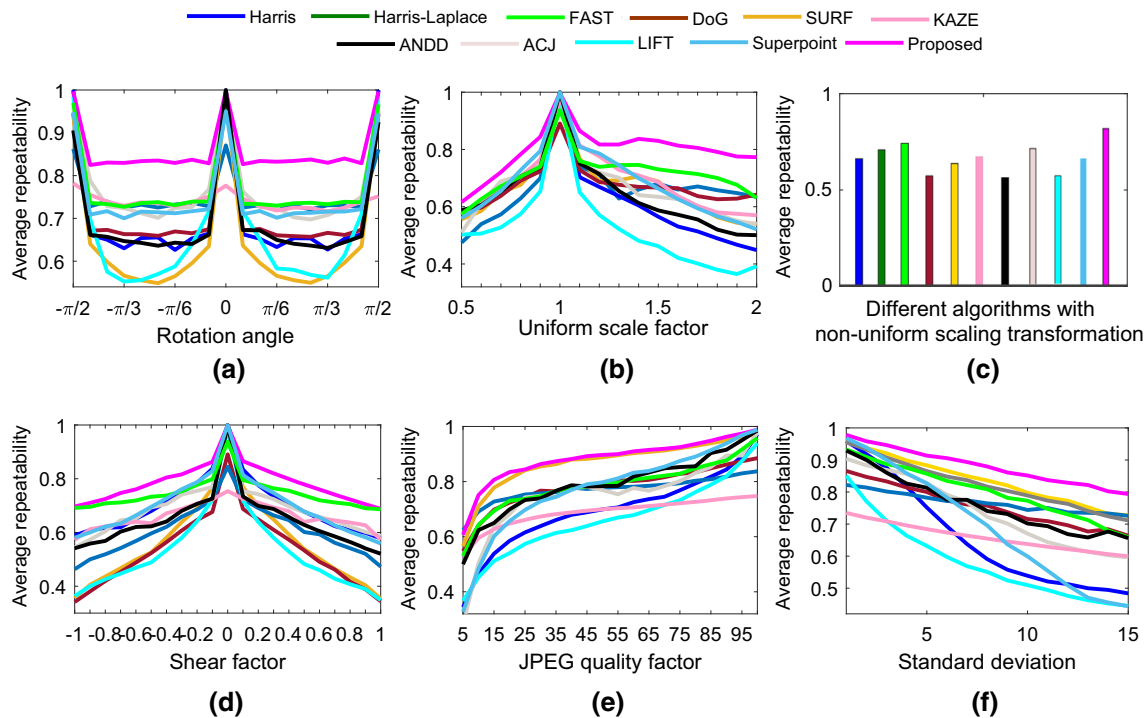
**Fig. 7** Average repeatability of the eleven detectors under rotation, uniform scaling, non-uniform scaling, shear transforms, lossy JPEG compression, and additive white Gaussian noises

data (http://www.robots.ox.ac.uk/vgg/affine/). The repeatability score for a given pair of images is computed as the ratio between the number of region-to-region correspondences and the minimum number of regions in one of the images. Two regions are deemed to correspond if the overlap error $\epsilon$ is sufficiently small. For region repeatability evaluation (Mikolajczyk et al. 2005), the overlap error is defined as one minus the ratio between the intersection of regions, $A \cap H^\top B H$, and the union of the regions, $A \cup H^\top B H$,

$$\epsilon = 1 - \frac{A \cap H^\top B H}{A \cup H^\top B H}, \tag{31}$$

where $A$ represents a region in the original image, $B$ represents the corresponding region in the transformed image, and $H$ is the corresponding homography between the original and the transformed image. When the overlap error between two regions is less than 40%, a correspondence is detected. The repeatability score is defined as

$$RS_i = \frac{CR_{1i}}{\min(C_1, C_i)}, \tag{32}$$

where $CR_{1i}$ is the number of correspondences between the original image and the $i$-th transformed image ($i = 1, \ldots, 6$), $C_1$ is the number of the detected corners from the original image, and $C_i$ is the number of the detected corners from the $i$-th transformed image.

In this experiment, six image sequences from http://www.robots.ox.ac.uk/vgg/affine/ are selected for performance evaluation and two image sequences (large zooming and rotations) are discarded. The reason is that it usually needs an appropriate descriptor to handle large image zooming and rotations (Duval-Poo et al. 2015). The threshold for each method is tuned to extracts about 1,000 corners from each input image. The repeatability scores for the six image sequences are illustrated in Fig. 8. Compared with the other ten methods, the proposed method achieves the best performance for the 'Trees', 'Bikes', 'Ubc', and 'Leuven' images. For the 'Wall' and 'Graffiti' images, the proposed method obtains a moderate performance. It is worth to note that the performance of the other methods vary greatly for different image sequences. The main reason is that the issue on how to effectively obtain the intensity variation information from the input images has not been considered in other ten methods. In conclusion, the proposed method achieves the best overall detection performance on region repeatability evaluation.

### 4.3 Repeatability Metric Under the DTU-Robots Dataset

In the DTU-Robots dataset (Aanæs et al. 2012), the performances of the detectors are evaluated under viewpoint, scale, and light changes using a large database of images with repeatability metric as a performance measure. The camera is

**Fig. 8** Comparison of different corner detectors on six image sequences, **a** Trees (image blur), **b** Bikes (image blur), **c** Ubc (image compression), **d** Leuven (light change), **e** Wall (viewpoint change), and **f** Graffiti (viewpoint change)
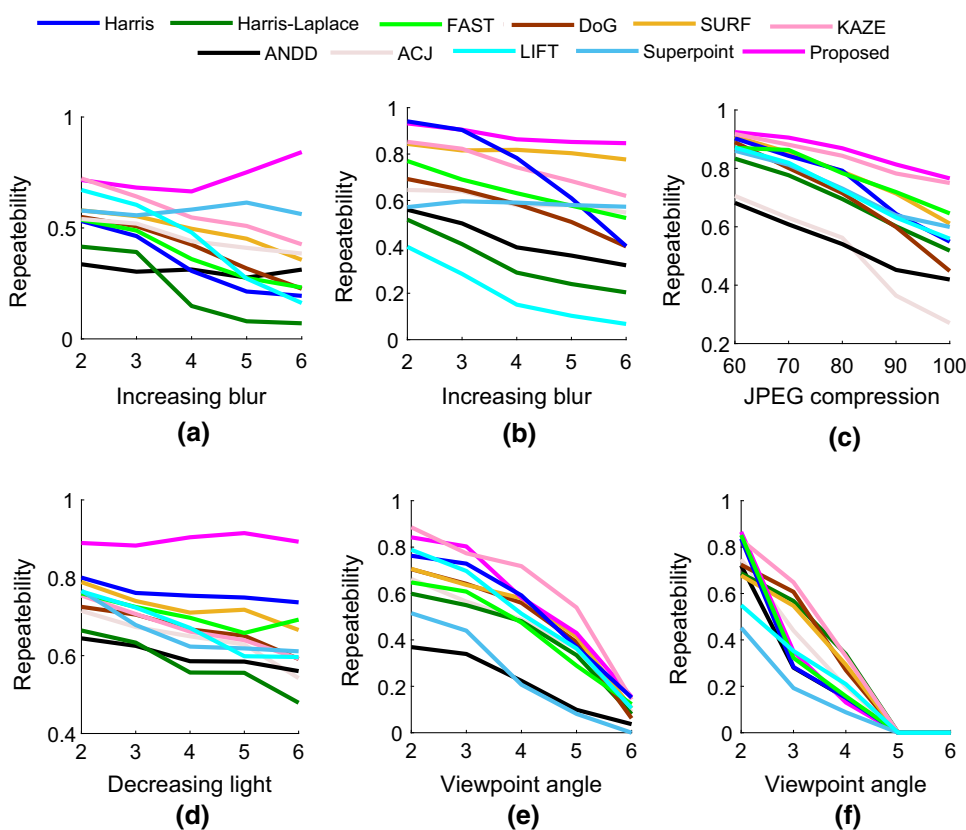
placed at 119 positions in three horizontal paths (Arc 1, Arc 2, and Arc 3) and along a linear path (Linear path) in front of 60 scenes. For each scene, 119 images of $1{,}200 \times 1{,}600$ pixels are acquired from the 119 camera positions. The center image which is the closest to the scene is chosen as the reference image. In the first evaluation setting, all feature points found in each image are compared with the points extracted from the reference image. Meanwhile, to simulate natural scenes, light varies from being diffuse on an overcast day to highly directional in sunshine and the scene is illuminated by 18 individually controlled light emitting diodes, which can be combined to provide a highly controlled and flexible light setting. In the second evaluation setting, the scene relighting has been carried out both from right to left and from back to front to investigate the sensitivity of the feature detectors to changes of lightings. At a camera position, ten different illumination settings are configured by changing the lighting directions. Then, ten different images are obtained from ten different illumination settings. All feature points found in each image are compared with the points extracted from the reference image (the tenth image is chosen as the reference image in this evaluation setting).

In this experiment, the repeatability metric for one pair of images is used as a performance measure which is defined as

$$R_{\text{metric}} = \frac{M_{\text{corresp}}}{M_{\text{total}}}, \tag{33}$$

where $M_{\text{corresp}}$ is the number of correspondences between the reference image and each image, and $M_{\text{total}}$ is the number of the detected corners from the reference image. A point in the reference image is marked as a correspondence point if it meets the following three criteria.

- Epipolar Geometry: Consistency with epipolar geometry is used as the first evaluation criterion. The camera positions provide the basis for the relationship between points in one image and associated epipolar lines in another. Points are eliminated if they are more than 2.5 pixels away from the epipolar line.
- Surface Geometry: 3D reconstruction is used as the second evaluation criterion. Two points are considered as a positive match if their 3D positions are within a window with a radius of 10 pixels to the scene surface obtained from the structured light reconstruction. On the contrary, points within a window with a radius of 10 pixels without reconstruction are removed.
- Absolute Scale: Scale consistency is used as the third evaluation criterion. The output scale of the point and the output scale of the corresponding point in another test image should be within a scale range of 2.

Fifty-four sets of images (a total of 122,094 test images) from the original sixty sets of images (Aanæs et al. 2012)

**Fig. 9** Average repeatability metric for different scene settings, **a** Arc 1, **b** Arc 2, **c** Arc 3, and **d** Linear path
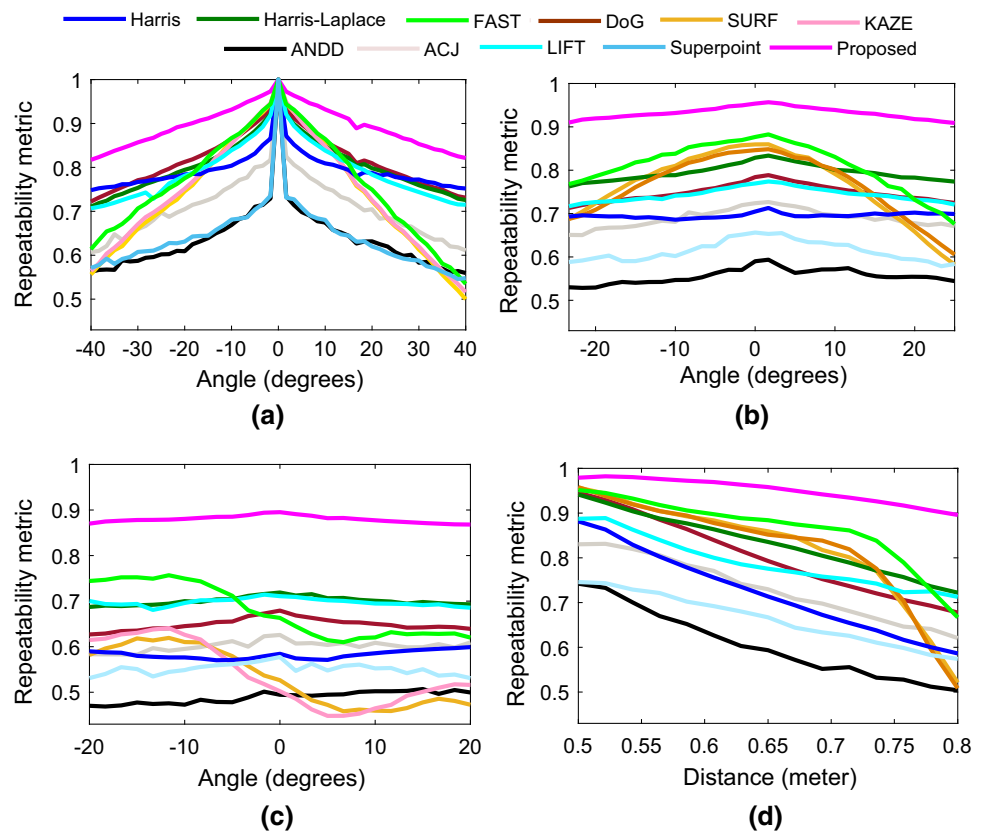


are obtained for our evaluation (the 31th–36th sets cannot be downloaded from Aanæs et al. (2012)). In this experiment, the threshold for each detector is adjusted so that each detector extracts about 2,000 corners from each input image. Figure 9 shows the average match percentage for 119 positions. The average repeatability metrics for the changes in light directions from right to left for four camera positions (1, 20, 64, and 65) are shown in the first row of Fig. 10. The average repeatability metrics for the changes in light direction from back to front for four camera positions (1, 20, 64, and 65) are shown in the second row of Fig. 10. It is worth to note that we follow the statement in Aanæs et al. (2012) and left out the FAST corner detector (Rosten et al. 2010) in the light change experiments because of the missing scale information and its, in general, unreliable performance. The average repeatability metric for each detector is summarized in Table 2. It can be observed that the proposed method outperforms all other methods by a large margin. The reason is that the proposed method has the ability to accurately extract image local intensity variation information to depict the differences between step edges and corners and accurately extract corners from the input images.

**Table 2** Average match percentage

| Detector | 119 positions | Light change from right to left | Light change from back to front |
|---|---|---|---|
| Harris | 0.715 | 0.644 | 0.549 |
| Harris–Laplace | 0.784 | 0.654 | 0.587 |
| FAST | 0.769 | – | – |
| DoG | 0.762 | 0.634 | 0.563 |
| SURF | 0.712 | 0.443 | 0.603 |
| KAZE | 0.715 | 0.438 | 0.561 |
| ANDD | 0.581 | 0.507 | 0.498 |
| ACJ | 0.684 | 0.559 | 0.540 |
| LIFT | 0.759 | 0.578 | 0.528 |
| Superpoint | 0.617 | 0.536 | 0.535 |
| Proposed | 0.907 | 0.788 | 0.705 |

## 4.4 Evaluation of Detection Performance Based on Ground Truth Images

Let $DC = \{(\hat{x}_i, \hat{y}_i), i = 1, 2, \ldots, M_1\}$ and $GT = \{(x_j, y_j), j = 1, 2, \ldots, M_2\}$ be the sets for detected corners by a corner detector and the true corners in the ground truth images respectively. For a corner $(x_j, y_j)$ in set $GT$, a corner is found from set $DC$ with the minimal distance. If
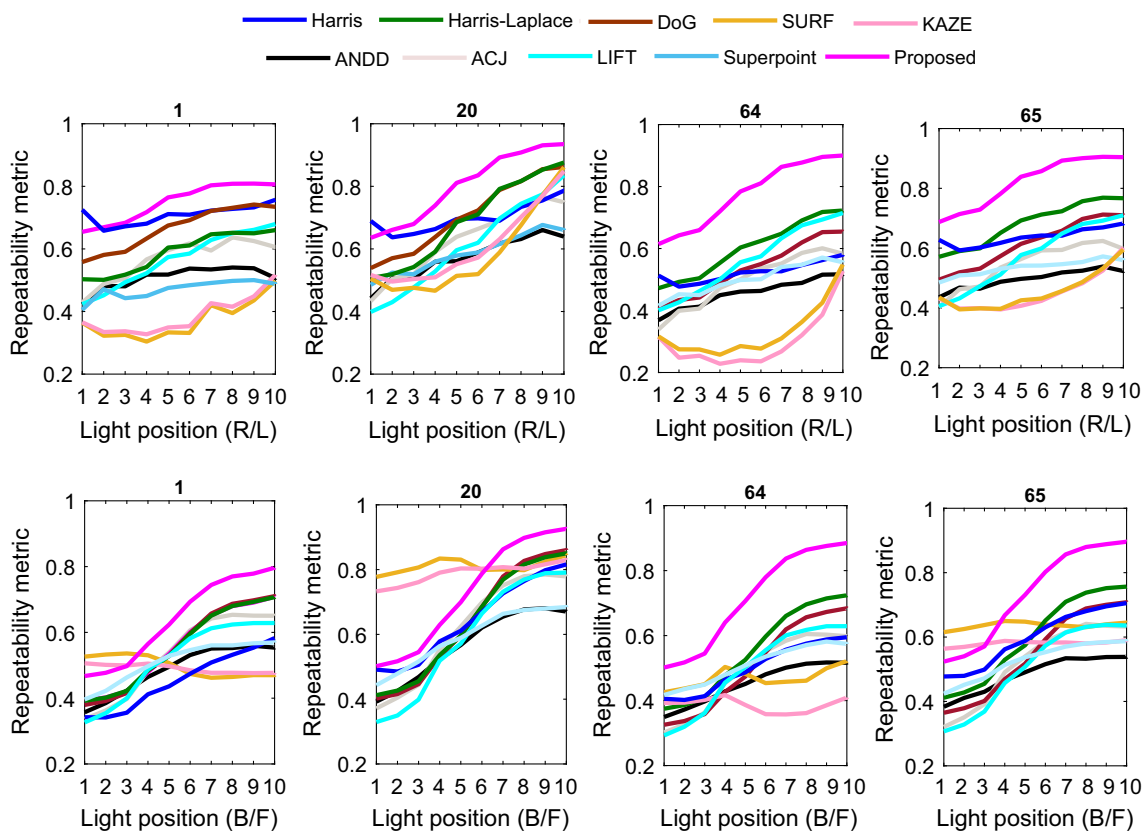
**Fig. 10** Average repeatability metric for the change in light direction for four camera positions (1, 20, 64, and 65). The first row is the average repeatability metric for the change in light direction from right to left (R/L). The second row is the average repeatability metric for the change in light direction from back to front (B/F)

the minimal distance is not more than a predefined threshold $\delta$ (here $\delta = 4$), corner $(\hat{x}_i, \hat{y}_i)$ is treated as correctly detected, and corner $(x_j, y_j)$ in set $GT$ and the detected corner in set $DC$ form a matched pair. Otherwise, the corner $(x_j, y_j)$ is counted as a missed corner. Similarly, for a corner $(\hat{x}_i, \hat{y}_i)$ in set $DC$, a corner is found from set $GT$ with the minimal distance. If the minimal distance is larger than threshold $\delta$, then corner $(\hat{x}_i, \hat{y}_i)$ is labelled as a false corner. The localization error is defined as the average distance for all the matched corner pairs. Let $\{(\hat{x}_l, \hat{y}_l), (x_l, y_l) : l = 1, 2, \ldots, N_m\}$ be the matched pairs in sets $GT$ and $DC$. The average localization error is calculated by

$$L_e = \sqrt{\frac{1}{N_m} \sum_{l=1}^{N_m} ((\hat{x}_l - x_l)^2 + (\hat{y}_l - y_l)^2)}. \tag{34}$$

The two commonly used images 'Geometric' and 'Lab' (Shui and Zhang 2013; Xia et al. 2014) are used for accuracy evaluations (Shui and Zhang 2013). The ground truths for the two test images are shown in Fig. 11. The image 'Geometric' contains 84 corners and the image 'Lab' contains 249 corners.
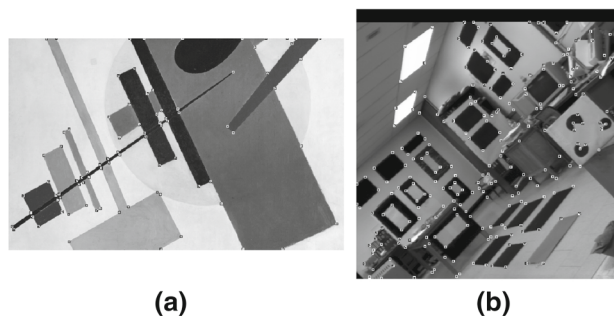


**Fig. 11** Test images **a** 'Geometric' and **b** 'Lab' and their ground truth corner positions

In this experiment, the proposed method is compared with five detectors (Harris (Harris and Stephens 1988), Harris–Laplace (Mikolajczyk and Schmid 2004), FAST (Rosten et al. 2010), ANDD (Shui and Zhang 2013), and ACJ (Xia et al. 2014)). The detection results of the six detectors are shown in Figs. 12 and 13. The number of missed corners, the number of false corners, and the localization error for each detector are listed in Table 3. For the two test images, different detectors show different detection characteristics.
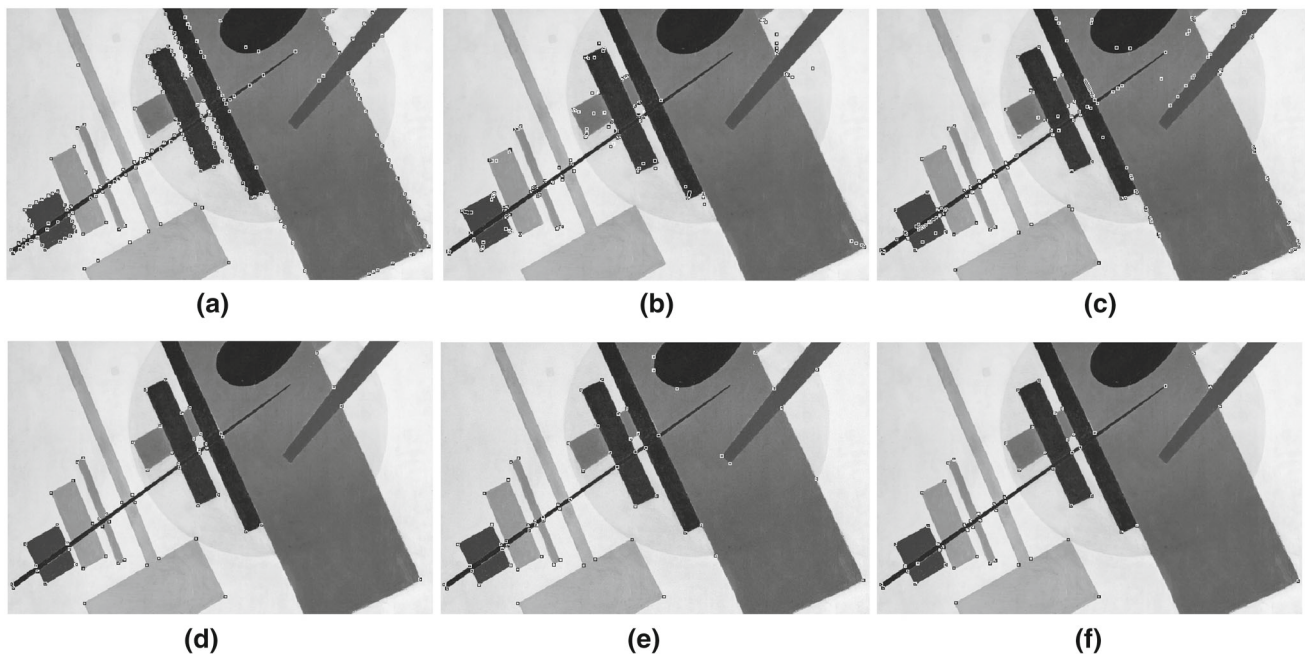
**Fig. 12** Detection results on the test image 'Geometric', **a** Harris (Harris and Stephens 1988), **b** Harris–Laplace (Mikolajczyk and Schmid 2004), **c** FAST (Rosten et al. 2010), **d** ANDD (Shui and Zhang 2013), **e** ACJ (Xia et al. 2014), and **f** Proposed detectors

Assume that missing a corner point and marking a false corner point incur the same loss in detection performance, the total number is used to assess the detection performance for a corner detector. The fewer the number of missed and false corner points, the better the detection performance. For the 'Geometric' image, the total number of missed and false corner points for the Harris, the Harris–Laplace, the FAST, the ANDD, the ACJ, and the proposed detectors are 144, 132, 112, 29, 28, and 28, respectively. For the 'Lab' image, the total number of missed and false corner points for the Harris, the Harris–Laplace, the FAST, the ANDD, the ACJ, and the proposed detectors are 215, 361, 255, 144, 225, and 169, respectively. It can be observed that the proposed detector and the ANDD detector attain the best corner detection.

Besides, the corner localization accuracy is another important measure to evaluate corner detectors. For the 'Geometric' image, the proposed method attains the smallest localization error, and the ANDD detector attains the second smallest localization error. For the 'Lab' image, the ANDD detector attains the smallest localization error, and the proposed detector attains the second smallest localization error. In conclusion, the proposed detector and the ANDD detector attains the best detection performance.

It is worth to note that our research also indicates that the performance of the proposed method can be affected by threshold selection and the change of illumination. Take a house image as an example, the corner detection results of the proposed method under different illuminations are shown in Fig. 14. It can be seen that

some obvious corner points in the window area (marked by 'O') cannot be detected under different illuminations as shown in Fig. 14c, d. The reason is that the directional derivatives $\xi_{\sigma,\rho}(\theta)$ in the window area are very small.

### 4.5 Execution Time and Memory Usage

The proposed corner detector has been implemented in MATLAB (R2017b) using a 2.81 GHz CPU with 16 GB of memory. For different images (http://www.robots.ox.ac.uk/vgg/affine/), the thresholds for the Harris, DoG, KAZE, ANDD, ACJ, and the proposed methods are tuned for detecting around 2000 features and each detector was executed 100 times. The corresponding execution time and memory usage are shown in Table 4. The codes for the Harris, DoG, KAZE, ANDD, and ACJ methods are written in MATLAB. It can be found that the memory usage of the proposed method is in the middle range among all the compared methods. Meanwhile, it can be observed that the proposed method cannot meet the needs of real-time applications. The proposed method can be implemented using GPU (Cornelis and Van Gool 2008) or FPGA (Huang et al. 2012) to improve the speed performance.

### 4.6 Application for 3D Reconstruction

In order to verify the performance of the proposed corner detector in real tasks, 3D reconstruction based on the
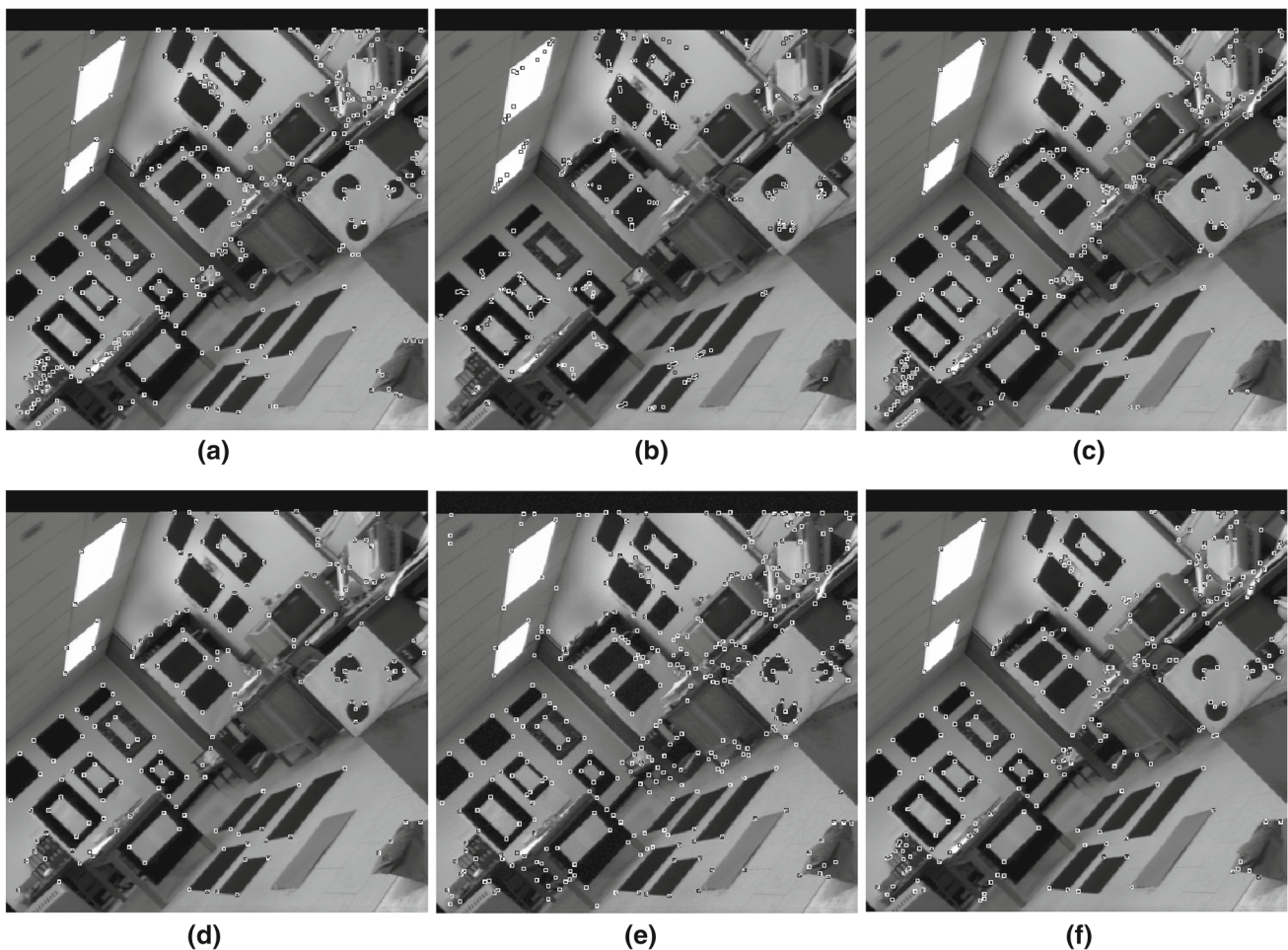
**Fig. 13** Detection results on the test image 'Lab', **a** Harris (Harris and Stephens 1988), **b** Harris–Laplace (Mikolajczyk and Schmid 2004), **c** FAST (Rosten et al. 2010), **d** ANDD (Shui and Zhang 2013), **e** ACJ (Xia et al. 2014), and **f** Proposed detectors

proposed corner detection method is carried out. Our 3D reconstruction process is based on the structure from motion technique in Hartley and Zisserman (2004) and Snavely et al. (2006) which aims to recover camera parameters, pose estimates, and sparse 3D reconstruction from image sequences. In this experiment, two datasets (Aanæs et al. 2012; Wilson and Snavely 2014) are used to perform 3D reconstruction. The first dataset (Aanæs et al. 2012) contains high resolution images which are captured from 49 fixed viewpoints. The second dataset (Wilson and Snavely 2014) contains unordered images and many are distorted images. These two datasets represent two typical image collection situations for 3D reconstruction applications. The first dataset (Aanæs et al. 2012) is widely used for applications about reconstructing a very specific scene or object. The second dataset (Wilson and Snavely 2014) is widely used for applications about reconstructing a very large scale place such as landmarks or cities.

We combined the proposed corner detector with the SURF descriptor (Bay et al. 2006) for sparse 3D reconstruction. The

proposed method is compared with the SURF method (Bay et al. 2006). The threshold for each detector is adjusted so that each detector extracts about 1,500 corners from each input image. The SURF descriptor (Bay et al. 2006) is with default scales. For each scene, forty images are selected for 3D reconstruction. The results of sparse 3D reconstruction are shown in Fig. 15. In this experiment, we use the number of reconstructed 3D points as the performance indicator for the two methods. For the 'Rabit' images, the SURF method and the proposed method used 7426 and 8562 points for 3D reconstruction respectively. For the 'Alamo' images, the SURF method and the proposed method used 21,322 and 25,680 points for 3D reconstruction respectively. It can be observed that the sparse 3D reconstruction from the proposed method contains more scene structure information. The reason is that the proposed method has the ability to accurately extract corners from the input images.

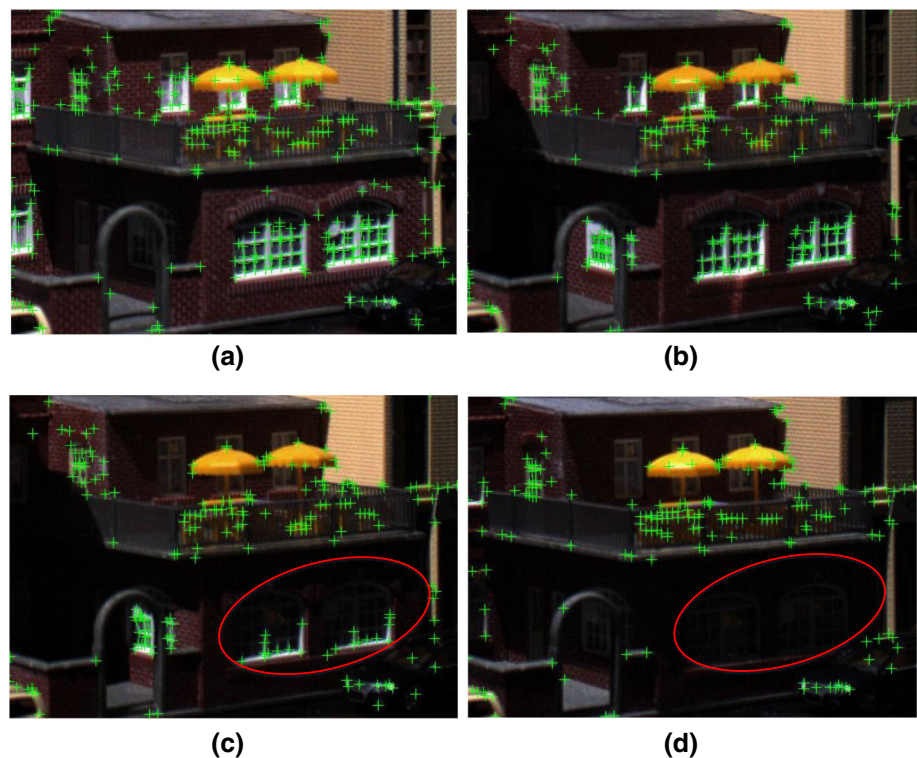**Fig. 14** Examples of the corner detection results of the proposed method under different illuminations



(a)

(b)

(c)

(d)

**Table 3** Performance comparison for the six detectors on two ground truth test images

| Detectors | Missed corners | False corners | Localization error (in pixels) |
|---|---|---|---|
| 'Geometric' image | | | |
| Harris | 15 | 129 | 1.279 |
| Harris–Laplace | 33 | 99 | 2.331 |
| FAST | 16 | 96 | 1.200 |
| ANDD | 28 | 1 | 1.264 |
| ACJ | 24 | 4 | 2.217 |
| Proposed | 27 | 1 | 0.889 |
| 'Lab' image | | | |
| Harris | 57 | 158 | 1.526 |
| Harris–Laplace | 155 | 206 | 3.032 |
| FAST | 75 | 180 | 1.572 |
| ANDD | 94 | 50 | 1.455 |
| ACJ | 45 | 180 | 2.273 |
| Proposed | 66 | 103 | 1.496 |

## 5 Conclusion

The contributions of the paper include six aspects. First, we proved for the first time that the existing intensity variation based corner detectors using the first-order derivatives along the horizontal and vertical directions cannot effectively detect corners. It is necessary to extract image local intensity variation information using the AGDD filters along multi-directions. Second, the properties of the anisotropic and isotropic Gaussian directional derivative representations of step edge, L-type corner, and other types of corners are investigated and discovered. Third, a new intensity variation extraction technique is presented which has the ability to accurately depict the intensity variation differences between step edges and corners. Fourth, a multi-directional structure tensor with multiple scales is derived for corner detection. Fifth, a new corner measure and a new corner detection algorithm are presented. Sixth, the proposed detector outperforms ten state-of-the-art corner detectors in terms of average repeatability (under affine image transformation, JPEG compression, and noise degradation), region repeatability, repeatability metric, detection accuracy, and localisation error. In our approach, the AGDD filters can also be replaced by other filters, such as shearlet, Gabor, or anisotropic diffusion filters for corner detection. The proposed corner detector also has a great potential to be applied in object tracking and many other fields. The program and demo on our corner detection can be accessed at http://vision-cdc.csiro.au/corner1st/.

**Table 4** Execution time and memory usage comparisons (image size is in pixels, the units of execution time and memory usage are in second and MB respectively)

| Detector | Bark (512 × 765) | Bikes (700 × 1000) | Boat (680 × 850) | Graffiti (640 × 800) | Leuven (600 × 900) |
|---|---|---|---|---|---|
| Execution time (s) | | | | | |
| Harris | 0.473 | 0.717 | 0.681 | 0.542 | 0.551 |
| DoG | 0.242 | 0.334 | 0.261 | 0.145 | 0.214 |
| KAZE | 0.515 | 0.993 | 0.778 | 0.657 | 0.709 |
| ANDD | 15.601 | 31.642 | 33.572 | 35.234 | 31.632 |
| ACJ | 5.215 | 9.452 | 7.245 | 6.436 | 6.421 |
| Proposed | 4.815 | 8.982 | 6.682 | 5.372 | 5.892 |
| Memory usage (MB) | | | | | |
| Harris | 25.7 | 45.9 | 38.7 | 28.3 | 35.4 |
| DoG | 155.6 | 205.1 | 188.1 | 160.2 | 171.3 |
| KAZE | 140.1 | 298.5 | 222.6 | 217.2 | 203.6 |
| ANDD | 155.6 | 219.6 | 201.5 | 181.2 | 192.3 |
| ACJ | 51.2 | 78.5 | 66.8 | 68.7 | 76.9 |
| Proposed | 71.9 | 168.6 | 153.8 | 140.7 | 143.6 |
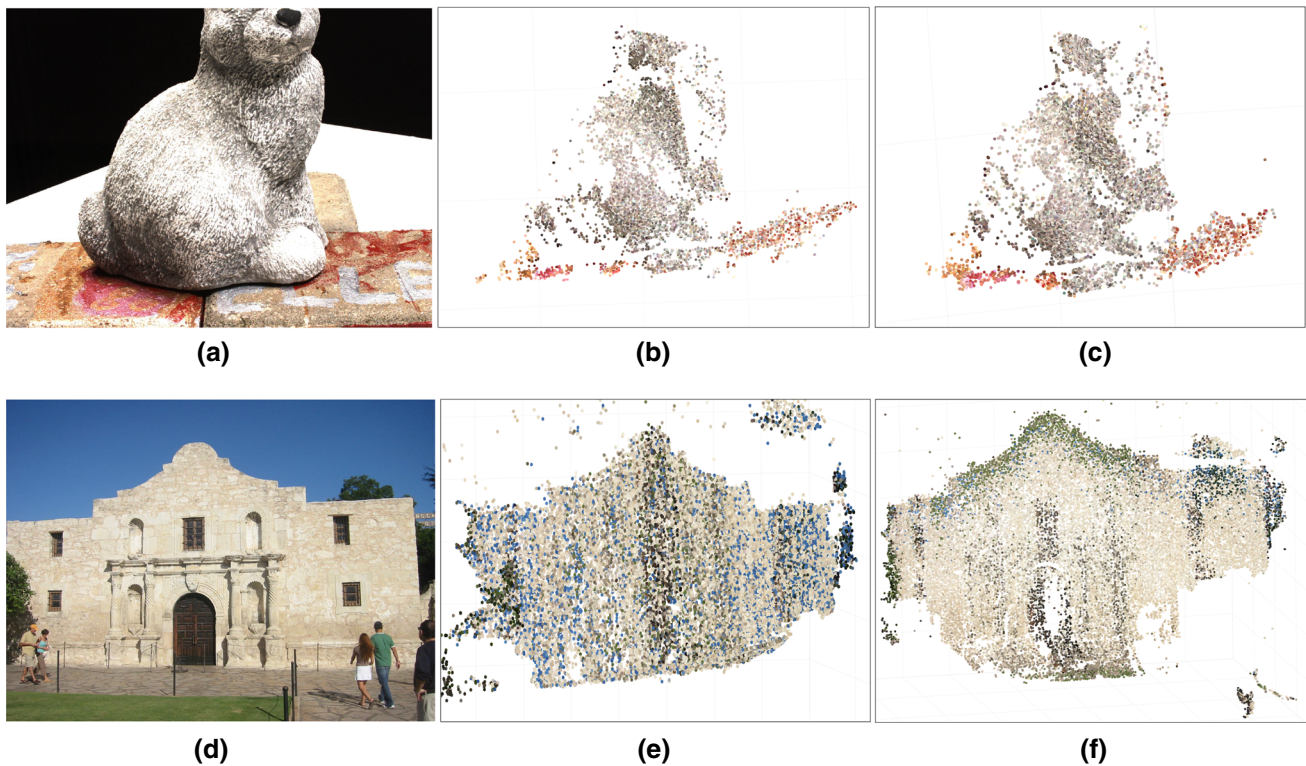


**Fig. 15** Test image 'Rabit' (Aanæs et al. 2012) and test image 'Alamo' (Wilson and Snavely 2014) are shown in (**a**) and (**d**) in the first column. Their corresponding sparse 3D reconstruction results of the SURF method (Bay et al. 2006) and the proposed method are shown in the second and the third columns respectively

## References

Aanæs, H., Dahl, A. L., & Pedersen, K. S. (2012). Interesting interest points. *International Journal of Computer Vision*, *97*(1), 18–35.

Alcantarilla, P., Bartoli, A., Davison, A. (2012). KAZE features. In *European conference on computer vision* (pp. 214–227). Springer.

Awrangjeb, M., & Lu, G. (2008). Robust image corner detection based on the chord-to-point distance accumulation technique. *IEEE Transactions on Multimedia*, *10*(6), 1059–1072.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded up robust features. In *European conference on computer vision* (pp. 404–417). Springer.

Bowyer, K., Kranenburg, C., & Dougherty, S. (1999). Edge detector evaluation using empirical ROC curves. In *IEEE conference on computer vision and pattern recognition* (Vol. 1, pp. 354–359).

Brox, T., Weickert, J., Burgeth, B., & Mrázek, P. (2006). Nonlinear structure tensors. *Image and Vision Computing*, *24*(1), 41–55.

Cornelis, N., & Van Gool, L. (2008). Fast scale invariant feature detection and matching on programmable graphics hardware. In *Computer vision and pattern recognition workshops* (pp. 1–8).

Deriche, R., & Giraudon, G. (1993). A computational approach for corner and vertex detection. *International Journal of Computer Vision*, *10*(2), 101–124.

DeTone, D., Malisiewicz, T., Rabinovich, A. (2018). Superpoint: Self-supervised interest point detection and description. In *IEEE conference on computer vision and pattern recognition* (pp. 224–236).

Duval-Poo, M. A., Odone, F., & De Vito, E. (2015). Edges and corners with shearlets. *IEEE Transactions on Image Processing*, *24*(11), 3768–3780.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A-Optics Image Science and Vision*, *4*(12), 2379–2394.

Gao, X., Sattar, F., & Venkateswarlu, R. (2007). Multiscale corner detection of gray level images based on LoG-Gabor wavelet transform. *IEEE Transactions on Circuits and Systems for Video Technology*, *17*(7), 868–875.

Gårding, J., & Lindeberg, T. (1996). Direct computation of shape cues using scale-adapted spatial derivative operators. *International Journal of Computer Vision*, *17*(2), 163–191.

Harris, C., & Stephens, M. (1988). A combined corner and edge detector. In *Alvey vision conference* (pp. 147–151).

Hartley, R. I., & Zisserman, A. (2004). *Multiple view geometry in computer vision*. Cambridge: Cambridge University Press.

Huang, F. C., Huang, S. Y., Ker, J. W., & Chen, Y. C. (2012). High-performance SIFT hardware accelerator for real-time image feature extraction. *IEEE Transactions on Circuits and Systems for Video Technology*, *22*(3), 340–351.

Kenney, C. S., Manjunath, B., Zuliani, M., Hewer, G. A., & Van Nevel, A. (2003). A condition number for point matching with application to registration and postregistration error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(11), 1437–1454.

Koenderink, J. J. (1984). The structure of images. *Biological Cybernetics*, *50*(5), 363–370.

Laptev, I. (2005). On space–time interest points. *International Journal of Computer Vision*, *64*(2–3), 107–123.

Lee, J. S., Sun, Y. N., & Chen, C. H. (1995). Multiscale corner detection by using wavelet transform. *IEEE Transactions on Image Processing*, *4*(1), 100–104.

Lenc, K., & Vedaldi, A. (2016). Learning covariant feature detectors. In *European conference on computer vision* (pp. 100–117). Springer.

Lepetit, V., & Fua, P. (2006). Keypoint recognition using randomized trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(9), 1465–1479.

Lindeberg, T. (1998). Feature detection with automatic scale selection. *International Journal of Computer Vision*, *30*(2), 79–116.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

Marimon, D., Bonnin, A., Adamek, T., & Gimeno, R. (2010). DARTs: Efficient scale-space extraction of DAISY keypoints. In *IEEE conference on computer vision and pattern recognition* (pp. 2416–2423).

Maver, J. (2010). Self-similarity and points of interest. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(7), 1211–1226.

Miao, Z., & Jiang, X. (2013). Interest point detection using rank order LoG filter. *Pattern Recognition*, *46*(11), 2890–2901.

Mikolajczyk, K., & Schmid, C. (2004). Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, *60*(1), 63–86.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., et al. (2005). A comparison of affine region detectors. *International Journal of Computer Vision*, *65*(1–2), 43–72.

Mokhtarian, F., & Suomela, R. (1998). Robust image corner detection through curvature scale space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(12), 1376–1381.

Moravec, H. P. (1979). Visual mapping by a robot rover. In *Proceedings of the 6th international joint conference on artificial intelligence* (Vol. 1, pp. 598–600).

Noble, J. A. (1988). Finding corners. *Image and Vision Computing*, *6*(2), 121–128.

Olson, C. F. (2000). Adaptive-scale filtering and feature detection using range data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(9), 983–991.

Perona, P., & Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*(7), 629–639.

Pham, T. A., Delalandre, M., Barrat, S., & Ramel, J. Y. (2014). Accurate junction detection and characterization in line-drawing images. *Pattern Recognition*, *47*(1), 282–295.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106.

Rattarangsi, A., & Chin, R. T. (1990). Scale-based detection of corners of planar curves. In *10th international conference on pattern recognition* (Vol. 1, pp. 923–930).

Rosten, E., Porter, R., & Drummond, T. (2010). Faster and better: A machine learning approach to corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(1), 105–119.

Ruzon, M. A., & Tomasi, C. (2001). Edge, junction, and corner detection using color distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(11), 1281–1295.

Schmid, C., Mohr, R., & Bauckhage, C. (2000). Evaluation of interest point detectors. *International Journal of Computer Vision*, *37*(2), 151–172.

Shui, P., & Zhang, W. (2012). Noise-robust edge detector combining isotropic and anisotropic Gaussian kernels. *Pattern Recognition*, *45*(2), 806–820.

Shui, P., & Zhang, W. (2013). Corner detection and classification using anisotropic directional derivative representations. *IEEE Transactions on Image Processing*, *22*(8), 3204–3218.

Smith, S. M., & Brady, J. M. (1997). SUSAN—A new approach to low level image processing. *International Journal of Computer Vision*, *23*(1), 45–78.

Snavely, N., Seitz, S. M., & Szeliski, R. (2006). Photo tourism: Exploring photo collections in 3D. *ACM Transactions on Graphics*, *26*(3), 835–846.

Su, R., Sun, C., & Pham, T. D. (2012). Junction detection for linear structures based on Hessian, correlation and shape information. *Pattern Recognition*, *45*(10), 3695–3706.

Teh, C. H., & Chin, R. T. (1989). On the detection of dominant points on digital curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(8), 859–872.

Trujillo, L., Olague, G. (2006). Synthesis of interest point detectors through genetic programming. In *Proceedings of the 8th annual*

*conference on genetic and evolutionary computation* (pp. 887–894).

Verdie, Y., Yi, V., Fua, P., & Lepetit, V. (2015). TILDE: A temporally invariant learned detector. In *IEEE conference on computer vision and pattern recognition* (pp. 5279–5288).

Wang, Y.-P. (1999). Image representations using multiscale differential operators. *IEEE Transactions on Image Processing*, *8*(12), 1757–1771.

Weickert, J., Romeny, B. T. H., & Viergever, M. A. (1998). Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Transactions on Image Processing*, *7*(3), 398–410.

Widynski, N., & Mignotte, M. (2014). A multiscale particle filter framework for contour detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(10), 1922–1935.

Wilson, K., & Snavely, N. (2014). Robust global translations with 1DSfM. In *European conference on computer vision* (pp. 61–75). Springer.

Witkin, A. (1984). Scale-space filtering: A new approach to multiscale description. In *IEEE international conference on acoustics, speech, and signal processing* (Vol. 9, pp. 150–153).

Xia, G., Delon, J., & Gousseau, Y. (2014). Accurate junction detection and characterization in natural images. *International Journal of Computer Vision*, *106*(1), 31–56.

Yi, K. M., Trulls, E., Lepetit, V., & Fua, P. (2016). LIFT: Learned invariant feature transform. In *European conference on computer vision* (pp. 467–483).

Zhang, X., Qu, Y., Yang, D., Wang, H., & Kymer, J. (2015). Laplacian scale-space behavior of planar curve corners. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(11), 2207–2217.

Zhang, W., & Shui, P. (2015). Contour-based corner detection via angle difference of principal directions of anisotropic Gaussian directional derivatives. *Pattern Recognition*, *48*(9), 2785–2797.

Zhang, W., Sun, C., Breckon, T., & Alshammari, N. (2019). Discrete curvature representations for noise robust image corner detection. *IEEE Transactions on Image Processing*, *28*(9), 4444–4459.

Zhang, W., Wang, F., Zhu, L., & Zhou, Z. (2014). Corner detection using Gabor filters. *IET Image Processing*, *8*(11), 639–646.

Zhang, W.-C., Zhao, Y., Breckon, T. P., & Chen, L. (2017). Noise robust image edge detection based upon the automatic anisotropic Gaussian kernels. *Pattern Recognition*, *63*(2), 193–205.

Zhang, X., Yu, F. X., Karaman, S., Chang, S.-F. (2017). Learning discriminative and transformation covariant local feature detectors. In *IEEE conference on computer vision and pattern recognition* (pp. 6818–6826).

Zhong, B., & Liao, W. (2007). Direct curvature scale space: Theory and corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(3), 100–108.