A Multi-scale Bilateral Structure Tensor Based Corner Detector

Lin Zhang, Lei Zhang*, and David Zhang

Biometrics Research Center, Department of Computing
The Hong Kong Polytechnic University
Hong Kong, China
Tel.: 852-27667355
{cslinzhang,cslzhang,csdzhang}@comp.polyu.edu.hk

Abstract. In this paper, a novel multi-scale nonlinear structure tensor based corner detection algorithm is proposed to improve effectively the classical Harris corner detector. By considering both the spatial and gradient distances of neighboring pixels, a nonlinear bilateral structure tensor is constructed to examine the image local pattern. It can be seen that the linear structure tensor used in the original Harris corner detector is a special case of the proposed bilateral one by considering only the spatial distance. Moreover, a multi-scale filtering scheme is developed to tell the trivial structures from true corners based on their different characteristics in multiple scales. The comparison between the proposed approach and four representative and state-of-the-art corner detectors shows that our method has much better performance in terms of both detection rate and localization accuracy.

Keywords: Harris, corner detector, bilateral structure tensor.

1 Introduction

Corner detection is a critical task in various machine vision and image processing systems because corners play an important role in describing object unique features for recognition and identification. Applications that rely on corners include motion tracking, object recognition, 3D object modeling, and stereo matching, etc.

Considerable research has been carried out on corner detection. One of the earliest successful corner detectors can be Harris corner detector [1]. Harris et al. [1] calculated the first-order derivatives of the image along horizontal and vertical directions, with which a 2×2 structure tensor was formed. The corner detection was accomplished by analyzing the eigenvalues of the structure tensor at each pixel. However, computing derivatives is sensitive to noise, and the Harris corner detector has poor localization performance because it needs to smooth the derivatives for noise reduction. Thus, several methods [2-3] have been proposed to improve its performance.

Apart from Harris corner detector and its variants, many other corner detectors have also been proposed by researchers. Kitchen and Rosenfeld [4] proposed a cornerness measure based on the change of gradient direction along an edge contour

* Corresponding author.

H. Zha, R.-i. Taniguchi, and S. Maybank (Eds.): ACCV 2009, Part II, LNCS 5995, pp. 618–627, 2010.

multiplied by the local gradient magnitude. Smith and Brady [5] proposed the SUSAN scheme. In SUSAN, a circular mask is taken around the examined pixel and this pixel is considered as the nucleus of the mask. Then "USAN" (Univalue Segment Assimilating Nucleus) is defined as an area of the mask which has the similar brightness as the nucleus. Smith et al. [5] assumed that the USAN would reach a minimum when the nucleus lies on a corner point. Wang and Brady [6] proposed a corner detection algorithm based on the measurement of surface curvature. In [7] and [8], Mokhtarian et al. proposed two CSS (Curvature Scale Space) based corner detectors. In these two algorithms, edge contours are first extracted and then corners are detected as the positions with high curvatures on edge contours. Zheng et al.'s [9] cornerness measure was simply the gradient module of the image gradient direction.

This paper presents a novel effective evolution of the classical Harris corner detector. In the original Harris corner detector, an isotropic Gaussian kernel is used to smooth each of the four elements in the 2×2 structure tensor over a local window before calculating the eigenvalues. Such a smoothing operation will have two disadvantages. First, some weak corners will be smoothed out. Second, the localization accuracy is much degraded. Inspired by the success of bilateral filters [10] in image denoising, which consider both the spatial and the intensity similarities in averaging neighboring pixels for noise removal, in this paper we construct a nonlinear bilateral structure tensor and use it to detect corner points.

The basic idea of the proposed method lies in that both the spatial and gradient distances should be involved in smoothing the structure tensor elements. The neighboring pixels that have shorter spatial and gradient distances to the given one should have higher weights in the averaging. In this way, a nonlinear structure tensor, which is adaptive to image local structures, could be constructed and hence the image local pattern could be better distinguished. It can be seen that the classical Harris corner detector is a special case of the proposed method by exploiting only the spatial distance in the structure tensor smoothing. However, the proposed nonlinear structure tensor has much higher sensitivity to corner-like fine structures than the linear structure tensor. Therefore, it may respond strongly to some trivial feature points in the image. In order to get rid of the possible false corners detected at fine image scales, we propose a multi-scale filtering scheme based on the different characteristics of true corners and trivial structures in multiple scales.

The rest of the paper is organized as follows. Section 2 briefly reviews the Harris corner detector. Section 3 presents the new corner detector in detail. Experimental results are presented in section 4 and the conclusion is made in section 5.

2 Harris Corner Detector

Harris corner detector [1] has been very widely used in machine vision applications. Consider a 2D gray-scale image I. Denote by $W \in I$ an image patch centered on (x_0, y_0). The sum of square differences between W and a shifted window $W_{(\Delta x, \Delta y)}$ is calculated as

$$S = \sum_{(x_i, y_i) \in W} (I(x_i, y_i) - I(x_i - \Delta x, y_i - \Delta y))^2 \tag{1}$$

By approximating the shifted patch using a Taylor expansion truncated to the first order terms, we have:

$$S = [\Delta x, \Delta y] A \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \tag{2}$$

where $A = \begin{bmatrix} \sum_{(x_i, y_i) \in W} (\nabla_i^h)^2 & \sum_{(x_i, y_i) \in W} \nabla_i^h \nabla_i^v \\ \sum_{(x_i, y_i) \in W} \nabla_i^v \nabla_i^h & \sum_{(x_i, y_i) \in W} (\nabla_i^v)^2 \end{bmatrix}$ and ∇_i^h and ∇_i^v represent the first order partial derivatives of image I along horizontal and vertical directions at pixel (x_i, y_i).

In practice matrix A is computed by averaging the tensor product $\nabla I \cdot \nabla I^{\mathrm{T}}$ (∇I denotes the gradient image of I) over the window W with a weighting function K_ρ, i.e.

$$A_\rho = \begin{bmatrix} \sum_{(x_i, y_i) \in W} K_\rho(i)(\nabla_i^h)^2 & \sum_{(x_i, y_i) \in W} K_\rho(i) \nabla_i^h \nabla_i^v \\ \sum_{(x_i, y_i) \in W} K_\rho(i) \nabla_i^v \nabla_i^h & \sum_{(x_i, y_i) \in W} K_\rho(i)(\nabla_i^v)^2 \end{bmatrix} \tag{3}$$

Usually K_ρ is set as a Gaussian function $K_\rho(i) = \frac{1}{\sqrt{2\pi}\rho} \exp\left(-\frac{d_i^2}{2\rho^2}\right)$, where $d_i^2 = (x_i - x_0)^2 + (y_i - y_0)^2$ and ρ is the standard deviation of the Gaussian kernel.

A_ρ is symmetric and positive semi-definite. Its main modes of variation correspond to the partial derivatives in orthogonal directions and they are reflected by the eigenvalues λ_1 and λ_2 of A_ρ. The two eigenvalues can form a rotation-invariant description of the local pattern. Under the situation of corner detection, three distinct cases are considered. 1) Both the eigenvalues are small. This means that the local area is flat around the examined pixel. 2) One eigenvalue is large and the other one is small. The local neighborhood is ridge-shaped. 3) Both the eigenvalues are rather large. This indicates that a small shift in any direction can cause significant change of the image at the examined pixel. Thus a corner is detected at this pixel.

Harris suggested that the exact eigenvalue computation can be avoided by calculating the response function

$$R(A_\rho) = det(A_\rho) - k \cdot trace^2(A_\rho) \tag{4}$$

where $det(A_\rho)$ is the determinant of A_ρ, $trace(A_\rho)$ is the trace of A_ρ, and k is a tunable parameter.

3 Bilateral Structure Tensor Based Corner Detection

This section presents the proposed multi-scale nonlinear bilateral structure tensor based corner detector in detail. Our algorithm differs from the original Harris corner detector mainly in two aspects. First, a nonlinear structure tensor is constructed to substitute for the linear one used in the Harris corner detector; second, a multi-scale filtering scheme is proposed to filter out the false and trivial corners detected at small scales.

3.1 Construction of the Bilateral Structure Tensor

The structure tensor for a gray level image I is a 2×2 symmetric matrix that contains in each element the orientation and intensity information in a local area. Denote by ∇I the gradient image of I. The initial matrix field can be computed as the tensor product $J_0 = \nabla I \cdot \nabla I^{\mathrm{T}}$. To incorporate the neighboring structural information into the given position, an averaging kernel could be used to smooth each element of J_0. Usually a Gaussian kernel K_ρ with standard deviation ρ is employed for this purpose:

$$J_\rho = K_\rho * J_0 \tag{5}$$

where symbol "*" means convolution. Since convolution is a linear operator, the structure tensor J_ρ is referred to as linear structure tensor [11]. It is a symmetric, positive semi-definite matrix. Comparing Eq. (3) with Eq. (5), we see that the matrix A_ρ in Harris corner detector is actually the linear structure tensor J_ρ at pixel (x_0, y_0).

In Harris corner detector [1], the "cornerness" of a pixel (x,y) is totally determined by its local structure tensor $J_\rho(x,y)$. However, the smoothing kernel K_ρ has two problems. First, the isotropic smoothing operation will smooth some weak corner features out so that the detection capability is decreased. Second, the localization accuracy of detected corner points will be reduced, which is a well-known problem of the Harris corner detector. Intuitively, if the local structure tensor can better preserve the local structural information at (x,y), the cornerness measured from it should be more reliable and accurate.

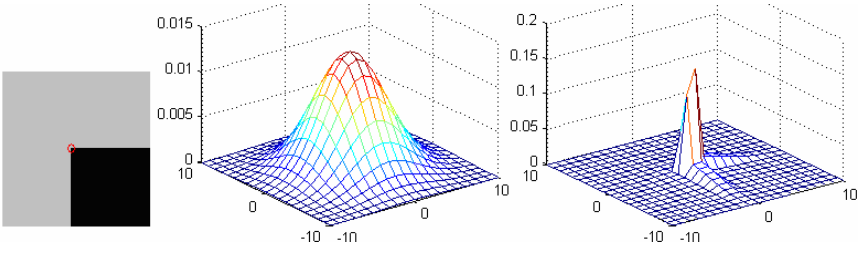


Fig. 1. Weight distributions in a neighborhood of a corner pixel. (a) An artificial image with an ideal corner (red circle); (b) weights distribution by using the Gaussian kernel K_ρ; (c) weights distribution by using the proposed bilateral weighting function $N_{\rho,\sigma}$.

As an early denoising technique, Gaussian smoothing is simple but it will over-blur the image details. The Gaussian weighting kernel only uses the notation of spatial location in the weights assignment. The greater the spatial distance from a neighboring pixel to the central pixel, the smaller the averaging weight will be assigned. The intensity similarity between the pixels is not exploited in Gaussian smoothing. In [10], the bilateral filter was proposed, which employs both the spatial and intensity similarities between pixels in averaging weight design. It has been shown that bilateral filtering could significantly improve the edge structure preservation while removing noise [10].

Inspired by the success of bilateral filters in image denoising, in this paper we construct a bilateral structure tensor for better corner detection performance. There are two basic factors in the formation of a local pattern: the relative positions between neighboring pixels and the intensity variations between them. Therefore, in the smoothing of J_0, we should consider both the spatial distance and the gradient distance in the averaging weight assignment. In the original Harris corner detector, only the spatial distance is considered by applying a Gaussian smoothing kernel K_ρ to $\nabla I \nabla I^{\mathrm{T}}$. In this paper, we will also involve the gradient distance in the smoothing of $\nabla I \nabla I^{\mathrm{T}}$.

Here, the gradient distance from the position (x_i, y_i) to the central position (x_0, y_0) is defined as:

$$d_i^g = \sqrt{\left(\nabla_i^h - \nabla_0^h\right)^2 + \left(\nabla_i^v - \nabla_0^v\right)^2} \tag{6}$$

The spatial distance from (x_i, y_i) to (x_0, y_0) is the same as in the original Harris corner detector:

$$d_i^s = \sqrt{\left(x_i - x_0\right)^2 + \left(y_i - y_0\right)^2} \tag{7}$$

By considering both the spatial and gradient distances into the assignment of averaging weight, we define the following bilateral weighting function for each pixel $(x_i, y_i) \in W$:

$$N_{\rho,\sigma}(i) = \frac{1}{C_{\rho,\sigma}} \exp\left(-\frac{(d_i^s)^2}{2\rho^2}\right) \cdot \exp\left(-\frac{(d_i^g)^2}{2\sigma^2}\right) \tag{8}$$

where ρ and σ are the parameters to control the decaying speeds over spatial and gradient distances, and

$$C_{\rho,\sigma} = \sum_W \exp\left(-\frac{(d_i^s)^2}{2\rho^2}\right) \cdot \exp\left(-\frac{(d_i^g)^2}{2\sigma^2}\right) \tag{9}$$

is the normalization factor.

Fig. 1 shows an example to illustrate the weight distributions by using the Gaussian kernel K_ρ and the proposed function $N_{\rho,\sigma}$. Fig. 1-a is an artificial image with an ideal corner in the center, which is marked by a red circle. The size of local window W for smoothing is set as 21×21. Figs. 1-b and 1-c illustrate the weight distributions for the pixels within W by using the Gaussian kernel K_ρ and the proposed bilateral weighting function $N_{\rho,\sigma}$, respectively. It is clearly seen that K_ρ is isotropic and is independent of the image local structure, while $N_{\rho,\sigma}$ is anisotropic and is adaptive to the image local pattern. In this example, the edge pixels have higher weights than the non-edge pixels because they are more similar to the examined corner pixel in terms of gradient. Meanwhile, for the pixels lying on the same edge, the ones near to the corner pixel have higher weights than the others because they have shorter spatial distances to the corner point.

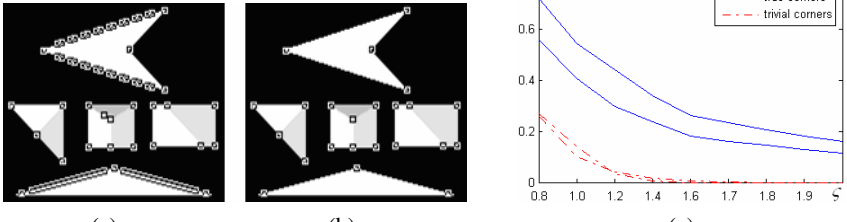
With the nonlinear bilateral weighting function $N_{\rho,\sigma}$, the nonlinear bilateral structure tensor is defined as:

$$A_{\rho,\sigma} = \begin{bmatrix} \sum_{(x_i,y_i)\in W} N_{\rho,\sigma}(i)(\nabla_i^h)^2 & \sum_{(x_i,y_i)\in W} N_{\rho,\sigma}\nabla_i^h\nabla_i^v \\ \sum_{(x_i,y_i)\in W} N_{\rho,\sigma}\nabla_i^v\nabla_i^h & \sum_{(x_i,y_i)\in W} N_{\rho,\sigma}(\nabla_i^v)^2 \end{bmatrix} \tag{10}$$

The corner detection is based on the analysis of the above defined nonlinear bilateral structure tensor $A_{\rho,\sigma}$. Similar to the original Harris corner detector, we calculate the response function $R(A_{\rho,\sigma}) = det(A_{\rho,\sigma}) - k \cdot trace^2(A_{\rho,\sigma})$ to determine if a corner point exists in the current position.

3.2 Multi-scale Filtering

Because the proposed nonlinear bilateral structure tensor $A_{\rho,\sigma}$ incorporates the local gradient information in the structure tensor construction, it could achieve much higher true detection and localization accuracies than the linear structure tensor used in the original Harris corner detector. However, it is also sensitive to some trivial structures. Due to digitization in the square grid, in discrete images often the ramp edges will show corner-like trivial structures in a fine scale. Those trivial structures will be enhanced by the proposed nonlinear structure tensor $A_{\rho,\sigma}$ and they may be falsely detected as true corners. Fig. 2-a shows an example. We can see many false detections along the ramp edge by using $A_{\rho,\sigma}$. To solve this problem, we propose a multi-scale filtering scheme to filter out those small scale trivial structures.



(a) (b) (c)

Fig. 2. (a) Corner candidates before multi-scale filtering; (b) final corner detection result after multi-scale filtering; (c) Relative cornerness ratio (RCR) curves of two true corners (blue curves) and two trivial corners (red curves);

Suppose that we have obtained some corner candidates with the proposed nonlinear structure tensor. We will distinguish the trivial corner-like structures from the true corners by their different cornerness characteristics at multiple image scales. The images at different scales can be obtained by smoothing the original image I with a series of Gaussian kernels K_ς with different standard deviations ς. By increasing the values of ς, a fine to coarse scale space can be formed. The underlying principle for our multi-scale filtering scheme is as follows. If a trivial structure is detected as a corner at a fine scale, the cornerness of this point should decrease rapidly with the increase of scale ς because it will be smoothed out by K_ς. On the contrary, if a true

corner point is detected at a fine scale, the cornerness of it will decrease smoothly with the increase of ς because it will appear in a wide range of scales.

Denote by R_0 the cornerness of a corner candidate measured by Eq. (4) at the finest scale 0, and by R_ς its cornerness measured at scale ς. We define the relative cornerness ratio (RCR) as

$$r_\varsigma = R_\varsigma / R_0 \tag{11}$$

Fig. 2-c shows the RCR curves of two true corner points (blue curves) and two trivial corner points (red curves). From this figure we can clearly see that the RCR of false corners will decay much faster than the RCR of true corners.

Based on the different behaviors of true corners and trivial corners in the scale space, we are able to tell them to remove false and trivial corners. Suppose we use L scales in the multi-scale filtering. A candidate corner point is recognized as a true corner point if

$$\sum_{l=1}^{L} r_\varsigma(l) \geq T \tag{12}$$

where T is a threshold. Fig. 2-b shows the final corner detection result after multi-scale filtering (L=3). We see that many false corners detected in Fig. 2-a are removed in Fig. 2-b without affecting the true corners.

4 Experimental Results

Experiments were performed on 3 standard test images. The ground truth corner points were manually labeled. For the *artificial* test image (refer to Fig. 3-a_3), it is easy to identify these reference corners and the locations of corners can be accurately located. However, for real test images *blocks* (refer to Fig. 3-a_1) and *house* (refer to Fig. 3-a_2), it is nearly impossible to give absolutely accurate corner locations. Therefore, we only computed the localization accuracy for the artificial test image, while computed the detection accuracy for all the three test images. The code can be found at http://www.comp.polyu.edu.hk/~cslzhang/MBST_CD/.

The proposed corner detector was compared with four representative algorithms: Harris [1], SUSAN [5], Enhanced CSS [8] and the nonlinear structure tensor based method [11]. In [11], the authors proposed two different ways to construct a nonlinear structure tensor: one is by isotropic diffusion and the other is by anisotropic diffusion. In this paper, we compared the result given by the isotropic diffusion because it achieves similar result to that by anisotropic diffusion but has much less computational cost. We refer to it as INLST for short in the following. For the four methods used in comparison, we tuned the parameters so that the best corner detection results were obtained.

The proposed method has several parameters. The parameter ρ (referring to Eq. (8)) is adaptively determined based on the size of window W, i.e. the spatial range, according to the 3-sigma principle of Gaussian function. Similarly, the parameter σ (referring to Eq. (8)) is fixed by the range of d_i^g (referring to Eq. (6)), i.e. the gradient range, according to the 3-sigma principle. In the multi-scale filtering, we empirically

find that it is insensitive to the scale selection and usually 3~5 scales are enough. Thus, in our experiments we used 3 scales and the same threshold for all the test images: ς_1=0.6, ς_2=1.0, ς_3=1.4 and T=1.0 (referring to Eq. (11) and Eq. (12)). Finally, the parameters left to set are the window size W and coefficient k (referring to Eq. (4)). In this paper they were set as follows: for the artificial test image, W=5×5 and k=0.04; for the *blocks* test image, W=21×21 and k=0.02; and for the *house* test image, W=13×13 and k=0.02.

Denote by C_{ref} the set of reference (ground truth) corners and by C_{det} the set of detected corners by a particular detector. Denote by d_{max} the maximal acceptable distance between the reference corner and the detected corner. In this paper, we set d_{max}=4(pixels). For a pair of corner points $C_i \in C_{ref}$ and $C_j \in C_{det}$, if the distance $d_{i,j}$ between C_i and C_j is minimum for $\forall i, j$ and $d_{i,j} \le d_{max}$, then C_j is labeled as a "correct" detection of C_i. Otherwise, C_i is labeled as "missed". The corners labeled as "missed" in C_{ref} are considered as true corners but not detected, and the remaining corners in C_{det} are considered to be the "false" detections. The localization error is the average of all the distances $d_{i,j}$ for the corners detected correctly.

The experimental results are summarized in Table 1 and Fig. 3. The classical Harris corner detector performs moderately well with respect to the true detection rate. However, it loses some weak corners, which can be clearly seen in Fig. 3-b_1 and Fig. 3-b_2. SUSAN performs very well on the *artificial* test image whereas its performance on the natural images is not so good. For the enhanced CSS method, its detection rate and localization accuracy heavily depend on the output of the contour extraction. If an actual connected contour is broken up by the contour extraction step, more false corner points would be detected since the algorithm regards the line endings as corner points. Table 1 shows that INLST has better localization performance than Harris, SUSAN and Enhanced CSS. However, it is sensitive to noise and trivial structures and has much false detection. The proposed method performs the best in terms of both detection rate and localization accuracy.

Table 1. Evaluation results on test images

Method	artificial				blocks			house		
	correct	missed	false	location error	correct	missed	false	correct	missed	false
Harris	78	0	0	1.1347	52	8	3	57	20	46
SUSAN	78	0	0	1.0982	48	12	15	62	15	27
Enhanced CSS	76	2	3	1.6992	55	5	8	50	27	11
INLST	78	1	52	0.6235	55	5	5	57	20	12
Proposed	**78**	**0**	**0**	**0.4187**	**57**	**3**	**0**	**64**	**13**	**4**

Among the tested detectors, SUSAN is the fastest one. The proposed method is slower than the other ones because it needs to compute the weight function $N_{\rho,\sigma}$ for each pixel. In the future we will investigate how to reduce the computational cost without sacrificing much the accuracy.

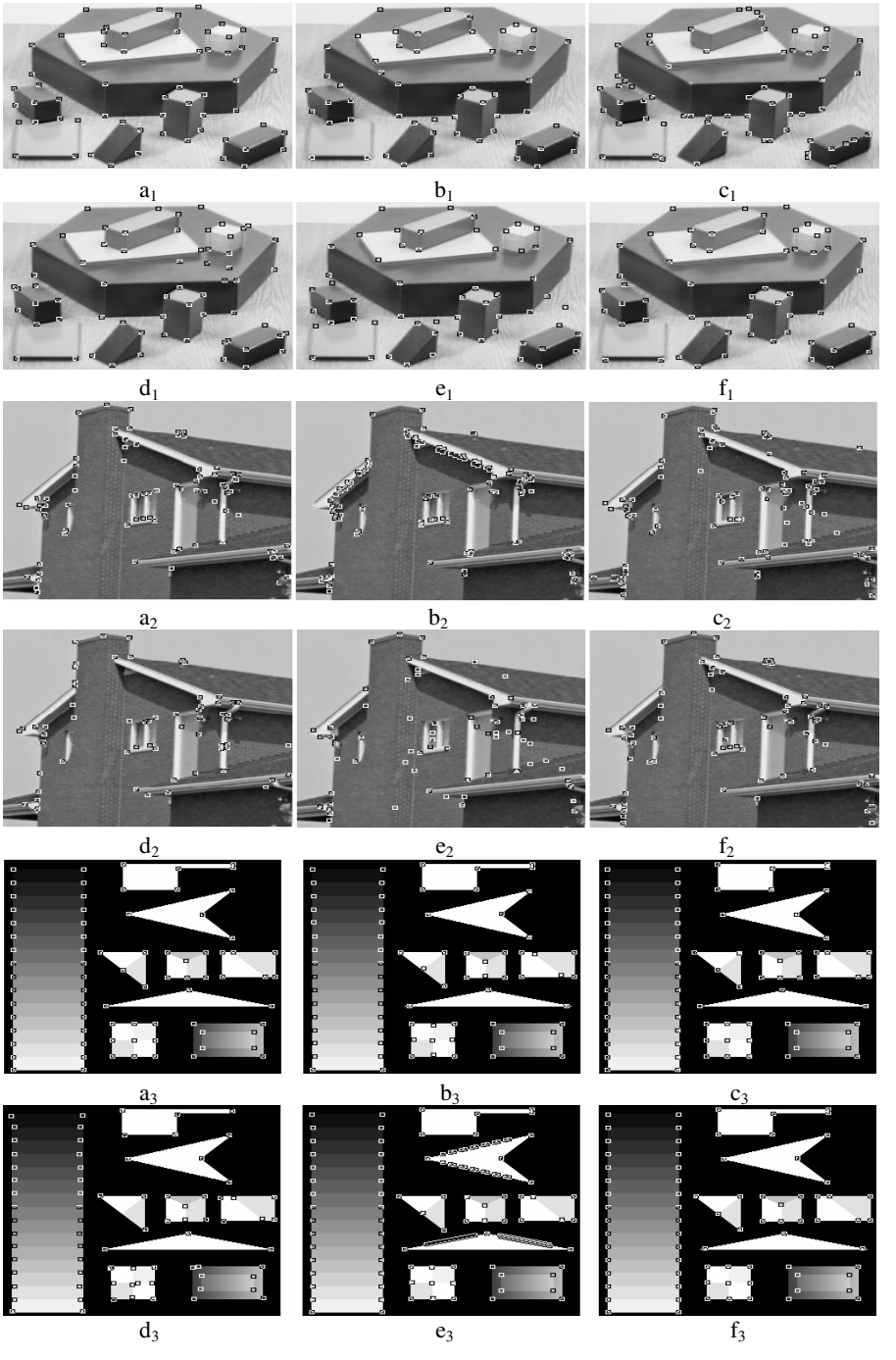


Fig. 3. Experimental results on 3 test images. (a_i) ground truth; (b_i) Harris; (c_i) SUSAN; (d_i) enhanced CSS; (e_i) INLST; (f_i) the proposed method.

5 Conclusions

In this paper, we proposed a corner detection algorithm by constructing a nonlinear bilateral structure tensor, which exploits both the spatial distances and the gradient distances from the neighboring pixels to the central pixel to be examined. Moreover, in order to remove the trivial corner-like structures, a multi-scale filtering scheme was developed. Experimental results on some standard test images show the effectiveness of the proposed corner detector in terms of both detection rate and localization accuracy. However, it should be noted that the computational cost of the proposed algorithm is higher than the other detectors. It can be a choice when the speed of corner detection is not a great concern but the accuracy is of the most importance.

Acknowledgement

The work is supported by the Hong Kong RGC General Research Fund (PolyU 5351/08E), Edward Sai Kim Hotung Fund (5-ZH52), and the HK-PolyU Internal Competitive Research Grant (G-YH54).

References

1. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. 4th Alvey Vision Conference, pp. 147–151 (1988)
2. Pei, S., Ding, J.: Improved Harris' Algorithm for Corner and Edge Detections. In: Proc. ICIP, pp. 57–60 (2007)
3. Liu, Y., Hou, M., Rao, X., Zhang, Y.: A Steady Corner Detection of Gray Level Images Based on Improved Harris Algorithm. In: Proc. Int. Conf. on Networking, Sensing and Control, pp. 708–713 (2008)
4. Kitchen, L., Rosenfeld, A.: Gray-level corner detection. Pattern Recognition Letters 1(2), 95–102 (1982)
5. Smith, S.M., Brady, J.M.: SUSAN—A New Approach to Low Level Image Processing. International Journal of Computer Vision 23(1), 45–78 (1997)
6. Wang, H., Brady, J.M.: Real-time corner detection algorithm for motion estimation. Image and Vision Computing 13(9), 695–703 (1995)
7. Mokhtarian, F., Suomela, R.: Robust image corner detection through curvature scale space. IEEE Trans. PAMI 20(12), 1376–1381 (1998)
8. Mokhtarian, F., Mohanna, F.: Enhancing the curvature scale space corner detector. In: Proc. Scandinavian Conf. on Image Analysis, pp. 145–152 (2001)
9. Zheng, Z., Wang, H.E.K., Teoh, E.K.: Analysis of gray level corner detection. Pattern Recognition Letters 20(2), 149–162 (1999)
10. Tomasi, C., Manduchi, R.: Bilateral Filtering for Gray and Color Images. In: Proc. ICCV, pp. 839–846 (1998)
11. Brox, T., Weickert, J., Burgeth, B., Mrazek, P.: Nonlinear structure tensors. Image and Vision Computing 24(1), 41–55 (2006)