Scale- and Affine-Invariant Fan Feature

Chunhui Cui and King Ngi Ngan, *Fellow, IEEE*

***Abstract*—Most existing feature detectors assume no surface discontinuity within the keypoints' support regions and, hence, have little chance to match the keypoints located on or near the surface boundaries. These keypoints, though not many, are salient and representative. In this paper, we show that they can be successfully matched by using the proposed scale- and affine-invariant Fan features. Specifically, the image neighborhood of a keypoint is depicted by multiple fan subregions, namely Fan features, to provide robustness to surface discontinuity and background change. These Fan features are made scale-invariant by using the automatic scale selection method based on the Fan Laplacian of Gaussian (FLOG). Affine invariance is further introduced to the Fan features based on the affine shape diagnosis of the mirror-predicted surface patch. The Fan features are then described by Fan-SIFT, which is an extension of the famous scale-invariant feature transform (SIFT) descriptor. Experimental results of quantitative comparisons show that the proposed Fan feature has good repeatability that is comparable to the state-of-the-art features for general structured scenes. Moreover, by using Fan features, we can successfully match image structures near surface discontinuities despite significant scale, viewpoint, and background changes. These structures are complementary to those found by the traditional methods and are especially useful for describing weakly textured scenes, which is demonstrated in our experiments on image matching and object rendering.**

***Index Terms*—Affine invariance, background invariance, feature description, feature detection, scale invariance, viewpoint invariance.**

I. Introduction

LOCAL image features have proven to be very successful in wide baseline matching and object recognition [1] as well as many other applications. Their robustness to partial visibility allows for successful matching even in severe cluttered scenes. Their good discriminative property provides high confidence in recognition. Basically, the feature-based schemes consist of two steps. First, the keypoints and their associated *support regions* are extracted from the image. Together they are referred to as features. Second, the descriptors are composed to summarize the features' appearance such as the shape and texture. For extensive investigation and comparison of feature detectors and descriptors, one can refer to [2] and [3]. The major problem of designing local features is how to obtain their invariance under different viewing conditions.

A. Related Work

There is a considerable body of previous research on scale-invariant features. In the early 1980s, Crowley *et al.* [4], [5] proposed to search for local extrema in the 3-D scale-space representation. A local 3-D extremum, (x, y, σ), in the scale space indicates a local feature with the keypoint located on (x, y) and the region extent (window size) determined by the scale parameter σ. In [6], Lindeberg proposed a systematic methodology for automatic scale selection. The basic idea is to select the characteristic scales, for which a given function attains extrema over scales. The scale is characteristic in the sense that it responds to some salient signal change in the image and consequently can be repeatably detected under different viewing conditions. Lindeberg proved that the scale normalized Gaussian derivatives are good choices to compute the multiscale function. Specifically, he suggested to use the scale-normalized Laplacian-of-Gaussian (LoG) to detect blob-like features. Later, Lowe [7] proposed the Difference of Gaussian (DoG) as the approximation of scale normalized LoG to accelerate the computation of scale-space representation. In detailed experimental comparisons, Mikolajczyk [8] found that the scale-normalized LoG produces the most stable features compared with a range of other Gaussian derivative functions, such as squared gradient, Hessian, and Harris corner function. Actually, a number of feature detectors [9]–[11] have adopted the scale-normalized LoG to select the characteristic scales. Other methods like maximally stable extreme regions (MSER) [12], edge based region (EBR) and intensity based region (IBR) [13] use different approaches to achieve scale invariance, yet the similar idea is using the salient intensity changes to indicate the characteristic local structures. Kadir *et al.* [14] proposed a different scale selection method, where local complexity is used instead as a measure of saliency and the salient scale is selected at the entropy extremum of the local descriptors.

To achieve rotation invariance, the common method is to describe the features using some rotationally invariant image measures, such as the generalized moments [15], the local jets [16], and RIFT [18]. In Lowe's SIFT [11], the free rotation is determined by estimating the dominant gradient orientation.

As an important step towards viewpoint invariance, affine invariance is highly desired for local features. Actually affine transformation is sufficient to locally model the image distortion arising from viewpoint changes, provided that: 1) small surface patches can be thought of as being comprised of coplanar points and 2) perspective effect can be ignored at a local scale. In the mid-1990s, Lindeberg *et al.* [19] developed a method to detect blob-like affine features in the context of shape from texture. It explores the properties of the second moment matrix and iteratively estimates the affine transformation of local patterns. This shape estimation method was later used for matching and recognition by Baumberg [20]. He used a multiscale Harris detector to

Manuscript received December 04, 2009; revised June 29, 2010, November 24, 2010; accepted December 14, 2010. Date of publication January 10, 2011; date of current version May 18, 2011. This work was supported in part by a grant from the Research Grants Council of the Hong Kong SAR, China, under Project CUHK 415707. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Min Wu.

The authors are with the Department of Electronic Engineering, the Chinese University of Hong Kong, Shatin N.T., Hong Kong (e-mail: chcui@ee.cuhk.edu.hk; knngan@ee.cuhk.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2010.2103948

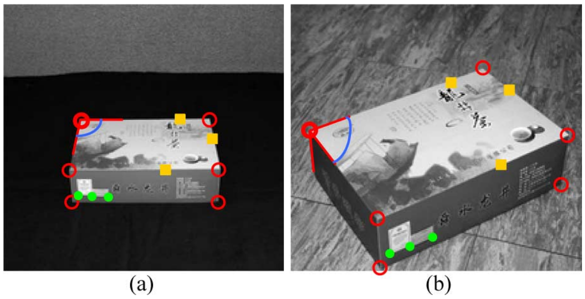


Fig. 1. What kind of extra keypoints can be matched by using Fan features?

extract the keypoints and then employed the iterative procedure proposed by Lindeberg to adapt the shape of the point neighborhood to the local image structure. Mikolajczyk and Schmid [10] went a step further by iteratively modifying the location, scale, and the neighborhood of a keypoint, such that both the keypoint and its associated support region are extracted in an affine-invariant way. Apart from the second moment matrix, the covariance matrix (or region moments) is also widely used for affine-invariant image normalization [21], as is employed by [12], [22], and [23] to cope with geometric deformation introduced by viewpoint change.

However, the basic assumption for affine-invariant features [2] does not hold for the keypoints located on or near the object boundaries. Conventional methods such as SIFT [11], Harris & Hessian Affine [10], and MSER [12] probably fail to match these keypoints because the point neighborhood cannot be modeled by a single planar surface due to depth discontinuity. Fig. 1 gives an example of these keypoints such as the 3-D corners and junctions (red circle, online version), the keypoints along the boundaries (golden square, online version) and the keypoints close to the boundaries (green dots, online version). Note that, for those "green dots," conventional methods may adapt their support regions to small or highly deformed ones that do not cross the surface boundaries. However, small regions are usually not sufficiently distinctive for reliable matching, and the highly deformed regions basically have low repeatability of detection under significant viewpoint changes. Therefore, it is difficult for conventional methods to match these green dots. In [26], SIFT has been improved by incorporating the object boundary information to guide anisotropic smoothing. The green dots can now be saved since the background clutter can be eliminated providing accurate object boundaries. However, in practice, it is nontrivial to obtain the object boundaries, especially from a single image. This is why in [26] stereo disparity map is employed. However, it cannot be obtained as the prior knowledge in applications such as wide-baseline matching.

B. Our Method

The basic idea to address the problem of surface discontinuity is straightforward. This idea is to divide the keypoint neighborhood into multiple subregions, each of which can now be reasonably assumed to represent a planar surface or just background. The subregions are described separately and all attached to the keypoint as independent signatures. As long as one of them exists in both views and can be matched successfully, the correspondence of the keypoints can be established accordingly. This is illustrated in Fig. 1, where the two upper-left box corners in images (a) and (b) can be matched according to the corresponding upper box surfaces bounded by the red lines (online version) and the blue arcs (online version). Similar ideas are shared by a few works, including 3-D singularity [17], [36], EBR [13], Edge-based feature [9], and Edge descriptor [33].

Both 3-D singularity [17], [36] and EBR [13] only aim at the well-formed edge junctions like the red circles (online version) in Fig. 1 and try to extract the support regions in a scale and affine-invariant manner. However, the extraction of 3-D singularity relies too much on the detection of complete and straight edges. The EBR feature [13] is more practically designed since only continuity of edges is required, and the intensity function is further introduced to detect salient and invariant regions. On the other hand, both Edge-based feature [9] and Edge descriptor [33] focus on extracting keypoints along the edges such as the golden squares (online version) in Fig. 1, yet only extract scale-invariant half-regions without taking into account affine deformation. The Edge-based feature [9] selects edge points as keypoints as long as the LoG filter detects salient scales. The features are usually duplicated and not distinctive enough, and selecting a single scale for both half-regions is not reasonable because they are supposed to represent different surfaces with independent extents (or one of them is background). The scale selection is improved in Edge descriptor [33], where the two sides divided by the edge can have different LoG scales. However, the computation of scale envelope highly relies on the continuity of edges which is difficult to guarantee in different images, and hence may hinder the features' repeatability. This is also true for EBR features.

In this paper, we propose a unified framework to extract and match both the edge junctions and the salient points along the edges for general structured scenes. The keypoints are selected from the edges that are efficiently and carefully detected to favor accurate surface boundaries. The repeatability of keypoints is guaranteed by a multiscale selection scheme. The point neighborhood is divided into multiple fan-shaped subregions, namely Fan features, by a method of edge association which does not rely on the continuity and completeness of edges. To achieve scale invariance for each Fan feature, we propose the Fan Laplacian of Gaussian (FLOG) filter to select its characteristic scales. To cope with geometric deformation, affine normalization is further applied by diagnosing the elliptical shape from the mirror-predicted surface patch. This in general gives us a better shape estimation than the traditional way. Note that both the scale selection and the affine normalization are based on textures, rather than edges. Finally, the scale- and affine-invariant Fan features are described by the Fan-SIFT, which is an extension of the well-known SIFT descriptor. Fan grids are carefully designed to replace the square grids used in SIFT. Strong gradients arising from the region boundaries are efficiently suppressed by a boundary mask.

The remainder of this paper is organized as follows. Section II describes the FLOG-based scale selection method. Section III presents the method to extract scale- and affine-invariant Fan features. Section IV introduces the Fan-SIFT descriptor, and

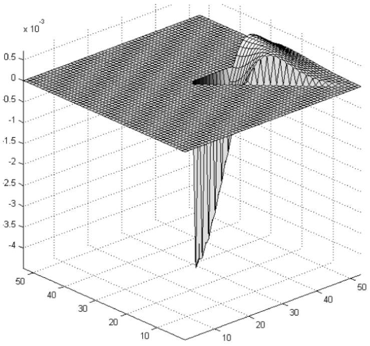


Fig. 2. FLOG kernel with included angle equal to 45° degree.

Section V discusses the matching strategy based on Fan features. The experimental results are given in Section VI, and Section VII concludes the paper.

II. Automatic Scale Selection by FLOG

In order to achieve scale invariance for fan subregions, a novel automatic scale selection method is proposed based on FLOG. Here, we first give the definition of the FLOG kernel and prove its transformation property under uniform scaling, which is emphasized in [6] as the fundamental requirement on a scale selection mechanism. We then describe the FLOG-based scale selection method and demonstrate its feasibility using some simple image patterns.

A. Scaling Property of FLOG Response

The standard LoG kernel in the polar coordinate system is defined in

$$\mathrm{LoG}(r;\sigma)=\frac{r^2-2\sigma^2}{2\pi\sigma^6}\exp\left(-\frac{r^2}{2\sigma^2}\right) \tag{1}$$

where σ is the standard deviation of Gaussian. The FLOG kernel can be interpreted as the standard LoG normalized by a factor of σ^2 and bounded within a fan domain $D=\{(r,\theta)\mid r\geq 0,\theta_1\leq\theta\leq\theta_2\}$. Formally, it is defined as

$$\mathrm{FLOG}_{\theta_1\theta_2}(r,\theta;\sigma)=\begin{cases}\sigma^2\mathrm{LoG}(r;\sigma), & \theta_1\leq\theta\leq\theta_2\\ 0, & \text{otherwise.}\end{cases} \tag{2}$$

Fig. 2 shows an example of the FLOG kernel. We can see that FLOG preserves the isotropy of LoG within the fan domain. When the fan domain expands to the circle domain, the FLOG kernel turns out to be exactly the same with the scale normalized LoG [6]. In this sense, FLOG is an extension of the scale normalized LoG. Next, we investigate the behavior of the integration of FLOG and input signal under uniform scaling.

Consider two 2-D signals f and f', where f' is obtained by uniformly scaling the spatial variables of f, i.e.,

$$f'(x',y')=f(x,y), \tag{3}$$

$$[x'\ y']^{\mathrm{T}}=s[x\ \mathrm{y}]^{\mathrm{T}}. \tag{4}$$

Accordingly, in the polar coordinate system, it holds that

$$f'(r',\theta')=f(r,\theta),\quad r'=sr,\quad \theta'=\theta. \tag{5}$$

Suppose that the scale parameters are transformed by the same factor in the two domains, i.e.,

$$\sigma'=s\sigma. \tag{6}$$

According to (1) and (2), we then have

$$\mathrm{FLOG}_{\theta_1\theta_2}(r',\theta';\sigma')=\frac{1}{s^2}\mathrm{FLOG}_{\theta_1\theta_2}(r,\theta;\sigma). \tag{7}$$

By (5)–(7), we can derive that

$$\begin{aligned}&\iint_{D'} f'(x',y')\cdot\mathrm{FLOG}_{\theta_1\theta_2}(x',y';\sigma')\,dx'dy'\\ &\quad=\int_{\theta_1}^{\theta_2}\int_0^\infty f'(r',\theta')\cdot\mathrm{FLOG}_{\theta_1\theta_2}(r',\theta';\sigma')r'\,dr'd\theta'\\ &\quad=s^2\int_{\theta_1}^{\theta_2}\int_0^\infty f'(r',\theta')\cdot\mathrm{FLOG}_{\theta_1\theta_2}(r',\theta';\sigma')r\,drd\theta\\ &\quad=\iint_{D} f(x,y)\cdot\mathrm{FLOG}_{\theta_1\theta_2}(x,y;\sigma)\,dxdy\end{aligned}$$

where

$$\begin{aligned}D&=\{(r,\theta)\mid r\geq 0,\theta_1\leq\theta\leq\theta_2\},\\ D'&=\{(r',\theta')\mid r'\geq 0,\theta_1\leq\theta'\leq\theta_2\}.\end{aligned}$$

It is rewritten as

$$\begin{aligned}&\iint_{D'} f'(x',y')\cdot\mathrm{FLOG}_{\theta_1\theta_2}(x',y';\sigma')\,dx'dy'\\ &\qquad\qquad=\iint_{D} f(x,y)\cdot\mathrm{FLOG}_{\theta_1\theta_2}(x,y;\sigma)\,dxdy.\end{aligned} \tag{8}$$

This means that the integration of FLOG and the input signal, called *FLOG response*, is equal in the two domains, provided that the spatial positions and the scale parameters are related according to (4) and (6). Let us look at the FLOG response as a function of the scale parameter σ. Based on above derivation, if the image pattern is rescaled by a constant scaling factor s, then the scale at which the FLOG response assumes its extrema will be multiplied by the same factor. Here, to guarantee the scale invariance, σ^2 is introduced to normalize the FLOG kernel, which is consistent with the scale normalized LoG [6].

B. FLOG-Based Scale Selection

As suggested in (8), the FLOG response can commute with the size change. This gives us a solution to detect the characteristic scales for a given subregion attached to a keypoint. First, according to the fan shape of the subregion, we choose a FLOG kernel with two appropriate directions θ_1 and θ_2. We then compute the multiscale FLOG response centered on the keypoint, i.e., the corner of the fan subregion. Finally, the extrema of FLOG response are detected and the corresponding scale parameters are chosen as the characteristic scales of the fan subregion. Ideally, if the image pattern within the subregion undergoes uniform scaling, the characteristic scales selected by this method before and after the scaling will indicate consistent image contents.

Intuitively, the characteristic scales can be repeatedly detected because they respond to salient signal changes. To see

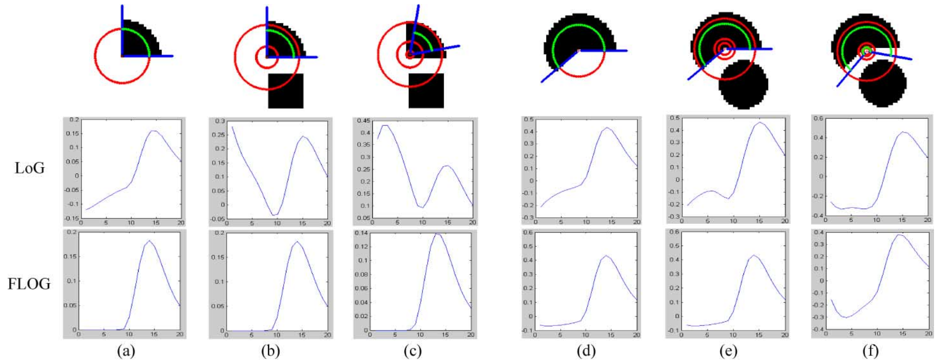


Fig. 3. Automatic scale selection for fan image patterns. The first row shows the scales detected by the scale normalized LoG (red circle, online version) and the FLOG (green arc, online version). The second and third rows present the corresponding multiscale responses computed using the scale normalized LoG kernel and the FLOG kernel, respectively. The horizontal axis is the parameter $k(\sigma = 1.2^{k-1})$. The vertical axis is the integration response.

how the extrema of FLOG response capture the salient signal changes, let us consider a simple fan step signal

$$f_s(r,\theta) = \begin{cases} 1, & 0 \le r \le r_0, \quad \theta_1 \le \theta \le \theta_2 \\ 0, & \text{otherwise.} \end{cases}$$

The extrema of its FLOG response can be found as follows:

$$\frac{\partial}{\partial\sigma}\int_{\theta_1}^{\theta_2}\int_0^{\infty} f_s(r,\theta)\cdot \text{FLOG}_{\theta_1\theta_2}(r,\theta;\sigma) r\, dr d\theta = 0 \Rightarrow \sigma = r_0/\sqrt{2}. \quad (9)$$

We can see that the scale parameter σ that makes the FLOG response attain its extremum is related to the distance from the step signal change, i.e., r_0, by a factor of $1/\sqrt{2}$.

In Fig. 3, the scale selection method is tested using some fan image patterns. For comparison, both the FLOG kernel and the scale normalized LoG kernel are applied. Suppose that we are only concerned with the extents of the fan patterns. Thus, we compute the multi-scale responses using the two kernels centered in the fan corner. The scale parameter is set as $\sigma = 1.2^{k-1}$ $(k = 1, 2, \ldots 20)$. The scales detected by LoG and FLOG are represented by the red circles and green arcs (online version), with their radius equal to the detected σ. The two directions for FLOG kernel are specified manually, as indicated by the two blue lines (online version)

As we can see in Fig. 3(a) and (d), when no clutter exists around the fan corner, both kernels can correctly reflect the extents of the fan patterns. Specifically, the scales detected by the two kernels are exactly the same, roughly $1/\sqrt{2}$ of the fan radius. However, when other patterns coexist, LoG attempts to find some uniform scales for the whole point neighborhood, leading to undesired or inaccurate scales as shown in Fig. 3(b) and (e). In comparison, FLOG only concerns the given subregion. Signal changes elsewhere will never affect the scale selection for the target subregion. Therefore, identical scales are repeatedly detected despite the nearby clutter, as we compare Fig. 3(a) and (b) and Fig. 3(d) and (e). In Fig. 3(c) and (f), considerable errors of direction estimation and keypoint localization are introduced. As we can see, these errors have little influence on the extents detected by FLOG, except for a new scale arising in Fig. 3(f) because an additional signal change is included in the fan subregion. However, if there is no salient signal change within the region, the FLOG response may not present any extremum and could be more sensitive to errors and noise, which is true for LoG as well.

III. Scale- and Affine-Invariant Fan Feature

Here, we describe how to extract from images the Fan features that are invariant to scale and affine change. Basically it consists of four steps: 1) keypoint detection; 2) edge association; 3) scale selection; and 4) affine normalization. In the following subsections, the details of each step are described.

A. *Keypoint Detection*

As the Fan features are specially designed for the keypoints located on surface boundaries, a natural choice will be to extract the keypoints from image boundaries [27], [28]. However, for the sake of accurate localization and computational efficiency, we prefer to extract keypoints from Canny edges [29]. In order to guarantee the accurate localization of edges and keypoints, the gradients are computed at a single fine scale. However, many clutters will arise from detailed textures at the fine scale. Therefore, the texture suppression technique [30] is employed before doing nonmaximum suppression. In addition, after hysteresis thresholding, there are usually many short and weak edge fragments. An edge cleaning procedure is introduced to eliminate these fragments, as we believe that strong and long edges are more likely to represent object boundaries.

Keypoints should present good repeatability under various imaging conditions. Here, we propose a multiscale selection scheme to select salient keypoints from the edge points. Let E denote the set of edge points detected in the image. Let $H(\mathbf{e}, \sigma_{\mathrm{k}})$ denote the Harris measure [31] of an edge point $\mathbf{e} \in E$ at the scale $\sigma_{\mathrm{k}} = \sigma_0^{\mathrm{k}-1}$ $(k = 1, 2, \ldots K)$. The spatial neighbor of $\mathbf{e}$ is defined as $N(\mathbf{e}) = \{\boldsymbol{\varepsilon} \in E, \boldsymbol{\varepsilon} \neq \mathbf{e} \mid \|\boldsymbol{\varepsilon} - \mathbf{e}\|_2 \le D_1\}$. At each scale σ_{k}, we select a subset S_{k} of the salient edge points as the candidate keypoints by performing the nonmaximum suppression

$$S_k = \{\mathbf{e} \in E \mid H(\mathbf{e}, \sigma_k) \ge H(\boldsymbol{\varepsilon}, \sigma_k) \quad \forall \boldsymbol{\varepsilon} \in N(\mathbf{e})\}. \quad (10)$$

Note that the keypoints that represent the same local structure but are detected at different scales may shift a little from each other. As we believe that the more scales a local structure survives the more stable it is, we then track each candidate keypoint across scales, trying to find its affinities at different scales and

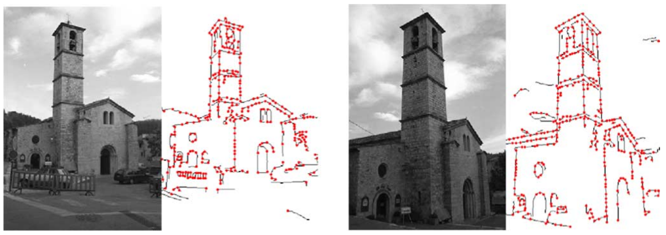


Fig. 4. Results of edge detection and keypoint extraction for two wide baseline images.

group them together as a single representative keypoint. Specifically, for a keypoint $\mathbf{p}_k \in S_k$ detected at scale σ_{k}, its affinity $\mathbf{p}_{k+1}$ at the next scale is defined as

$$\mathbf{p}_{k+1} = \arg\min_{\mathbf{x}\in S_{k+1}} \|\mathbf{x}-\mathbf{p}_k\|_2 \quad \text{subject to} \quad \|\mathbf{x}-\mathbf{p}_k\|_2 \le D_2. \tag{11}$$

If $\mathbf{p}_{k+1}$ exists, it will be removed from the set S_{k+1}, and we then try to find $\mathbf{p}_{k+2}$ for $\mathbf{p}_{k+1}$. Otherwise, the tracking is stopped and we obtain a group of keypoints $\{\mathbf{p}_m, \mathbf{p}_{m+1}\ldots, \mathbf{p}_{m+n}\}$. This group of keypoints will be combined into a single representative $\mathbf{p}_m$, i.e., the one detected at the finest scale, and its saliency is measured by $n+1$, i.e., the number of consecutive scales it survives. The tracking will be performed for each candidate keypoint $\mathbf{p} \in S_k$ $(k = 1, 2, \ldots K)$ until all of the keypoints have been removed from the candidate sets. Finally, we keep those representatives whose saliency is not smaller than T. In our experiments, the above parameters are empirically set to $\sigma_0 = 1.4$, $K = 5, T = 3$. The distance measure D_1 and D_2 are efficiently implemented by 7 × 7 and 3 × 3 windows, respectively. Fig. 4 shows an example of the edges and the keypoints detected in two wide baseline images. We can see the high repeatability of the keypoints. Note that a few keypoints may not have associated edges or characteristic scales (see Sections III-B and III-C) and hence are removed later.

B. Edge Association

For each keypoint, nearby edge fragments are associated to guide its neighborhood division. Inspired by [28], we first approximate each edge fragment by one or several straight line segments, as represented by the dotted lines in Fig. 5. Then, the correlation score between the keypoint $\mathbf{p}$ and a line segment l_k within the local window W is calculated by (12), where $d(\mathbf{p}, l_k)$ is the Euclidean distance of the keypoint $\mathbf{p}$ from the line segment l_k. The parameter ε is used to control the distance tolerance and is set to a small number such that the score drops fast as the distance increases. A Gaussian weighting $G(\mathbf{x}-\mathbf{p};\sigma)$ is introduced centered on $\mathbf{p}$ to emphasize the edge points near the keypoint.

$$\mathrm{Score}_k = \sum_{\mathbf{x}\in l_k, \mathbf{x}\in W} G(\mathbf{x}-\mathbf{p};\sigma)\cdot \exp\left(-d(\mathbf{p}, l_k)^2/\varepsilon^2\right). \tag{12}$$

Line segments with high scores are chosen to associate with the keypoint. In practice, as the truly related line segments usually have salient scores, a simple thresholding is sufficient. Finally, a line emitting from the keypoint is fitted to each associated line segment, indicating a division direction. Accordingly, multiple subregions are constructed around the keypoint. An example of edge association is shown in Fig. 5.

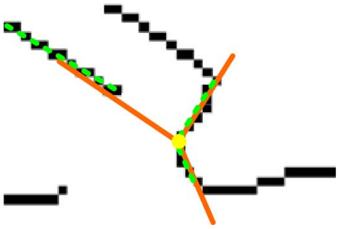


Fig. 5. Edge association. The yellow dot represents the keypoint. By edge association, the three line segments (green dotted lines) are associated to the keypoint. The estimated division directions are indicated by the three red solid lines through the yellow dot.

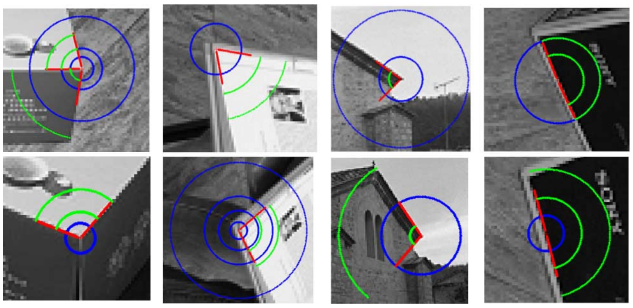


Fig. 6. Scale-invariant Fan features detected in real images taken from quite different viewpoints. The red lines are the division directions estimated by edge association. The scales selected by FLOG and LoG are indicated by the green arcs and the blue circles, respectively.

In our experiments of wide baseline matching, we choose to discard those subregions whose included angles are larger than 200°, because most of them capture either the background or multiple physical surfaces. Background subregions probably have no correspondences since the content of background could change a lot in wide baseline images. As for the regions comprised of multiple surfaces, we cannot use a single affine transform to model its geometric deformation. By removing these regions, there will be less clutter in final feature matching.

C. Scale Selection

The characteristic scales for each subregion are automatically selected by FLOG as described in Section II. As more than one scale can be detected for a subregion, multiple scale-invariant Fan features with different extents and from different subregions can be extracted for a single keypoint. For discrete implementation of the FLOG kernel, we face the problem of finite sampling approximation. In our experiments, the mask size of FLOG is set heuristically to $1+\mathrm{ceil}(3\sigma)$. To restore the zero mean property for the discrete FLOG mask, all the positive coefficients are uniformly scaled such that their sum equals to the absolute sum of all the negative coefficients. Of course, this procedure will slightly distort the mask shape. By experiments, we find that it usually leads to more distinctive extrema of FLOG response, but has little influence on the scales where the extrema are detected.

Fig. 6 gives some examples of the scale-invariant Fan features detected in wide baseline image pairs. The scales selected by LoG are also displayed for comparison. We can observe that the Fan features together with their FLOG scales can be detected consistently between widely separated views, whereas the scales selected by LoG are largely affected by nearby clutters. As suggested in (9) and Figs. 3 and 6, in general, the region extent detected by FLOG is shrunk compared to the location of salient signal change. To make the Fan features more distinctive, the detected subregions should be further enlarged to include the signal changes. However, large regions may lose the local properties such as the local planarity and the robustness to occlusion. In our experiments, the extents of all the subregions are enlarged by three times.

D. Affine Normalization

In addition to the scale change, fan subregions may suffer geometric deformation when observed from different viewpoints. Under the assumption that each subregion represents a locally planar surface, such a deformation can be modeled by an affine transform and hence can be addressed by affine normalization. This will make the scale invariant Fan features further possess affine invariance.

The second moment matrix [19], [20], [10] can be used to measure the affine deformation of an isotropic structure. This method works well for a circular support region, but is not suitable for a fan subregion. Indeed, other subregions attached to the keypoint should never be involved in estimating the affine shape of the concerned subregion, because they are supposed to represent different physical surfaces.

On the other hand, the covariance matrix [21], [22], [12] has also been successfully employed to diagnose the affine shape. For an image region R with arbitrary shape, its local image moments M_{ij} and the covariance matrix $\mathbf{C}$ can be computed by

$$M_{ij} = \iint_R x^i y^j I(x,y)\,dxdy \tag{13}$$

$$\mathbf{x}_c = [\bar{x}, \bar{y}]^{\mathbf{T}} = [M_{10}/M_{00}, \quad M_{01}/M_{00}]^{\mathbf{T}}$$
$$\mathbf{C} = \begin{bmatrix} M_{20}/M_{00} - \bar{x}^2 & M_{11}/M_{00} - \bar{x}\cdot\bar{y} \\ M_{11}/M_{00} - \bar{x}\cdot\bar{y} & M_{02}/M_{00} - \bar{y}^2 \end{bmatrix} \tag{14}$$

where $\mathbf{x}_c$ is the region centroid. Let λ_a and λ_b be the largest and smallest eigenvalues of the covariance matrix $\mathbf{C}$, respectively. Let v_a and v_b be the two corresponding eigenvectors. An important property of $\mathbf{C}$ is that its v_a and v_b indicate the semi-major and semi-minor axes of the ellipse (affine) shape of R, and λ_a and λ_b are proportional to their squared lengths. If $\lambda_a \neq \lambda_b$, which is usually the case in practice, we can use the affine transformation given in

$$\hat{\mathbf{x}} = \mathbf{A}(\mathbf{x}-\mathbf{x}_c) = s\begin{bmatrix} \lambda a^{-1/2} & 0 \\ 0 & \lambda b^{-1/2} \end{bmatrix}\begin{bmatrix} v_a^{\mathbf{T}} \\ v_b^{\mathbf{T}} \end{bmatrix}(\mathbf{x}-\mathbf{x}_c) \tag{15}$$

to project the anisotropic image pattern to an isotropic one. Here, $\mathbf{A}$ denotes the affine transform matrix, $\mathbf{x}$ and $\hat{\mathbf{x}}$ are the image coordinates before and after the affine transformation, respectively. s is a scaling factor. In this paper, it is decided to normalize $s\lambda_a^{-1/2}$ to 1, such that the image pattern is only expanded in the direction of v_b.

Note that the affine normalization is performed centered on the estimated region centroid $\mathbf{x}_c$ that is definitely not the position of the keypoint to which the fan subregion is attached. As shown in Fig. 7(a), directly computing the covariance matrix on the fan subregion will give us a diagnosis of the affine deformation centered on the region centroid, i.e., the red dots (online version) in Fig. 7(a). Affine normalization based on this shape estimation, as indicated by the green ellipses (online version) in Fig. 7(a), cannot accurately compensate the true deformation centered on the yellow-colored keypoint (online version). Fig. 7(b) shows the considerable differences of the normalized subregions detected in the two images $\mathbf{I}_1$ and $\mathbf{I}_2$ with significant viewpoint change. Here, the fan directions are represented by the red and blue lines, and the region extent determined by the FLOG scale is indicated by the green arcs. Fig. 7(c) shows the corresponding affine shapes in the original images, where we can also observe that the subregions are inconsistently detected in the two images.

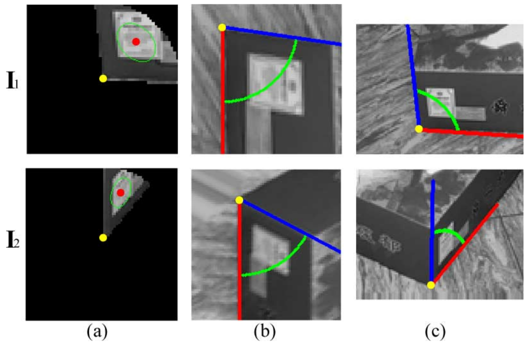


Fig. 7. Traditional affine normalization applied to the Fan subregions detected in two images $\mathbf{I}_1$ and $\mathbf{I}_2$. (a) Original image patches for affine shape diagnosis. (b) Normalized image patches. (c) Corresponding affine shapes in the original images.

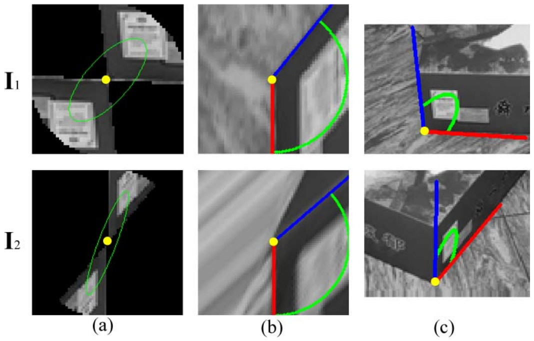


Fig. 8. Improved affine normalization applied to the fan subregions detected in two images $\mathbf{I}_1$ and $\mathbf{I}_2$. (a) Mirror-predicted image patches for affine shape diagnosis. (b) Normalized image patches. (c) Corresponding affine shapes in the original images.

Here we introduce a simple and efficient method to address this problem. Suppose that a subregion represents an incomplete planar surface attached to a keypoint, in comparison with a circular feature whose support region is a complete surface around the keypoint. To estimate the affine shape of a subregion, we propose to predict the image pattern of the missing part of the planar surface by mirroring the known subregion, as shown in Fig. 8(a). For a mirror-predicted image patch, its region centroid is guaranteed to locate on the keypoint, and hence the affine shape diagnosed by the covariance matrix, as indicated by the green ellipses in Fig. 8(a), can give us a better estimation of the local geometric deformation around the keypoint. The improvement in affine normalization can be clearly observed as we compare Figs. 7(b) and (c) with Figs. 8(b) and (c), where the appearances of the two subregions normalized by using the mirror prediction are much more similar than using the traditional way.

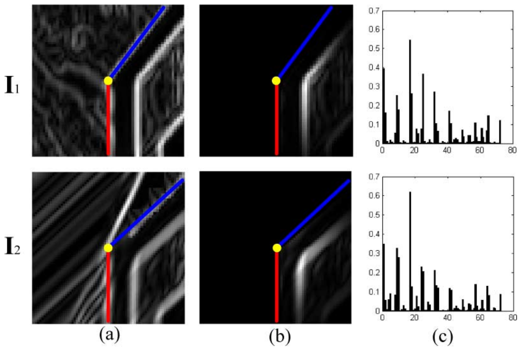


Fig. 9. Computation of Fan-SIFT descriptor. (a) Gradients computed in the normalized image patch. (b) Relevant gradients after global Gaussian weighting and boundary suppression. Gradients outside the subregions are also eliminated. (c) Fan-SIFT descriptors.

An iterative estimation method similar to [10] can be employed to further improve the scale and affine normalization of Fan features. To save computations, however, we adopt a single scale selection plus affine normalization, and we found that it works well in the experiments.

To conclude the whole section, the proposed method can efficiently extract consistent subregions from two images despite significant viewpoint changes, and normalize the regions to present very similar appearance, which is essential for feature description and matching.

IV. Fan-SIFT Descriptor

The well-known SIFT descriptor [11] is an invariant and stable representation of region appearance by a weighted histogram of gradient locations and orientations. It performs best in the context of matching and recognition [3]. The Fan SIFT descriptor proposed in this section is an extension of the SIFT descriptor for describing the fan subregions. The technical details are described below.

First, the intensity gradients are computed in the normalized image patch generated by the method described in Section III, as shown in Fig. 9(a). The smoothing scale σ_g for computing the gradients is chosen as $\sigma_g = k\sigma_s\sqrt{\theta/2\pi}$, where σ_s is the FLOG scale, θ is the fan angle of the normalized subregion, and the control parameter k is set to 1/3 in our experiments, so as to preserve the fine texture details for high discrimination. Following [11], the gradients are weighted by a global Gaussian function centered on the keypoint to provide the robustness to occlusion to some extent.

As can be observed in Fig. 9(a), there are always strong gradients around the subregion boundaries. These gradients actually depict the region shape rather than its inner texture. To suppress these boundary gradients, we introduce a boundary mask defined in

$$\mathrm{Mask}(\mathbf{x}) = \begin{cases} 0 & , \quad \mathrm{dis}(\mathbf{x}) \le \varepsilon_1 \\ 1 & , \quad \mathrm{dis}(\mathbf{x}) \ge \varepsilon_2 \\ \dfrac{\exp(\mathrm{dis}(\mathbf{x}) - \varepsilon_1) - 1}{\exp(\varepsilon_2 - \varepsilon_1) - 1}, & \quad \text{otherwise} \end{cases} \tag{16}$$

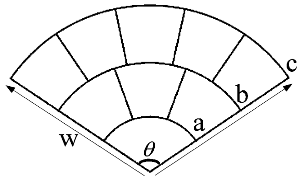


Fig. 10. Design of fan grids for Fan-SIFT descriptor.

where $\mathbf{x}$ is the sample position and $\mathrm{dis}(\mathbf{x})$ is the minimal distances from $\mathbf{x}$ to the two boundaries. The threshold ε_2 is set to $1+3\sigma_g$ by taking into account the diffusion of Gaussian smooth for computing gradients. The threshold ε_1 is simply set to $\varepsilon_2/2$. Unlike the one in [26], our suppression is only performed on samples very close to the region boundaries, such that the inner texture can be well preserved for the purpose of discrimination. Fig. 9(b) shows the relevant gradients after the boundary suppression. Gradients outside the subregions are eliminated as well.

Next, fan grids are introduced to distribute the gradients into nine discrete locations, as shown in Fig. 10. The width of the fan region is three times the FLOG scale, i.e., $w = 3\sigma_s$. The radius of the three fan rings are set to $a = w/3$, $b = 2w/3$ and $c = w$, and the fan angles for each ring are equally divided, such that all of the fan grids have the same areas $(S_f = \theta\sigma_s^2/2)$, i.e., all of the discrete locations have the same number of gradient samples. To achieve a rotationally invariant description, a coordinate system is aligned to the direction from the fan vertex to the region centroid, which is unique once the keypoint and the fan region are determined. In this way, we avoid estimating the dominant gradient orientation as the SIFT does. Gradient orientations are then computed in this coordinate frame and are quantized into eight bins.

Finally, a histogram of gradient locations and orientations is built in a way similar to [11]. Based on the histogram, a vector with $9 \times 8 = 72$ dimensions is composed as the Fan-SIFT descriptor. The descriptor is further normalized into a unit vector to compensate for affine changes in illumination. Fig. 9 illustrates the computation of Fan-SIFT descriptors for the two Fan features detected in different images. As we can see in Fig. 9(c), the two Fan-SIFT descriptors can be reliably matched.

V. Matching Based on Fan Features

The similarity of two Fan features is measured by the Euclidean distance between their descriptors. The *nearest neighbor distance ratio* [11], [3] is employed to match the descriptors. Specifically, two descriptors $D_{\mathbf{A}}$ and $D_{\mathbf{B}}$ are matched if $||D_{\mathbf{A}} - D_{\mathbf{B}}||/||D_{\mathbf{A}} - D_{\mathbf{C}}|| < t$, where the descriptors $D_{\mathbf{B}}$ and $D_{\mathbf{C}}$ are the first and second nearest neighbor to $D_{\mathbf{A}}$. The threshold t is set to 0.8 in our experiments.

To obtain the tentative correspondences of keypoints based on the matching of Fan features, we introduce the following strategy. Two keypoints $\mathbf{p}_1$ and $\mathbf{p}_2$ are matched as long as one of the Fan features attached to $\mathbf{p}_1$ can be matched to one of those attached to $\mathbf{p}_2$. This is based on our assumption that each fan subregion represents a local physical surface attached to the keypoint. As a result, any one of them can be used as the signature of the keypoint.

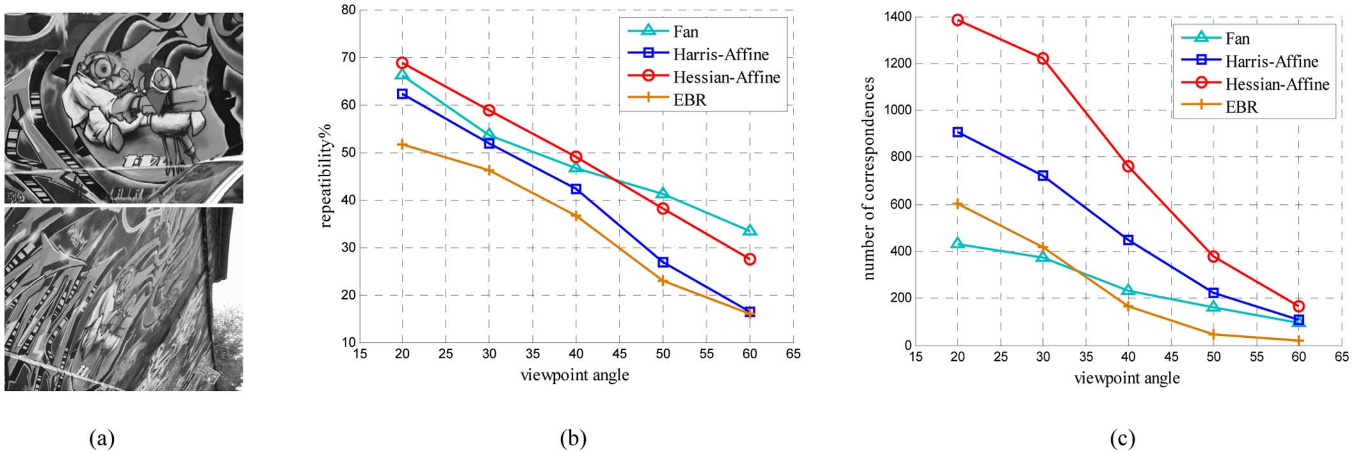


Fig. 11. Test of viewpoint invariance for Graffiti sequence , [2]. (a) Exemplar images. (b) Repeatability score. (c) Number of correspondences.

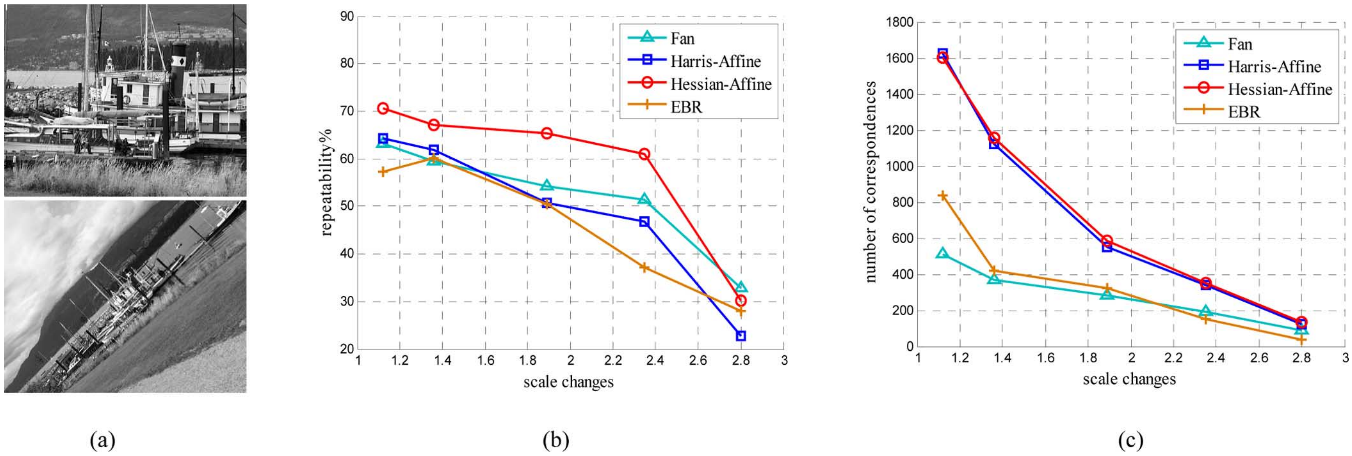


Fig. 12. Test of scale (+ rotation) invariance for Boat sequence , [2]. (a) Exemplar images. (b) Repeatability score. (c) Number of correspondences.

To automatically reject false matches, we apply a global and a semi-local geometric filter. The global one is the epipolar test using RANSAC. Matches that violate the estimated epipolar geometry will be discarded. The semi-local filter [37] is based on the consensus of nearby local affine transforms. This filter can also be applied to the Fan feature because a correspondence of fan subregions can provide sufficient information to infer the local affine transform between two images.

VI. Experimental Results

To evaluate our method, we compare the proposed Fan feature with *Harris Affine*, *Hessian Affine* [10], [2] and EBR [13], all of which have been efficiently implemented[1] and are publicly available. Different features basically capture different image structures. The Harris and Hessian features detect corner-like and blob-like structures within object surfaces [3]. Both EBR features and Fan features are extracted from edges. EBR arises from well-formed edge junctions, while Fan feature aims at both edge junctions and the salient points along edges. Through the following experiments, we show that not only does the Fan feature possess nice invariance property that is comparable to the state-of-the-art features [2], but also it can successfully match image structures near surface discontinuities, and hence contributes to the variety of the bag of features.

A. *Repeatability Under Viewpoint and Scale Change*

Here, we follow the standard test [2] to evaluate the repeatability and accuracy of Fan features under viewpoint and scale changes. The results for Graffiti and Boat sequences are shown in Figs. 11 and 12, respectively. The Image pairs in these sequences can be related by a single homography. Thus, we can determine the feature correspondence by measuring the overlap of their elliptical regions which are mapped onto the same image by the known homography (the support region of a Fan feature is deemed to be a complete ellipse in this test). Following [2], the region size is normalized to a radius of 30 pixels prior to computing the overlap measure. Two features are considered as a correspondence if the overlap error is smaller than 40%. The repeatability score is computed as the ratio between the number of correspondences and the smaller of the number of features extracted in a pair of images.

Fig. 11(b) shows that for the Graffiti scene, Fan feature has better repeatability than Harris Affine and EBR under viewpoint changes. Though Hessian Affine performs the best for small and median viewpoint angles, Fan feature exceeds it in the case of large viewpoint changes. The test results of scale invariance in

[1] [Online]. Available: http://www.robots.ox.ac.uk/~vgg/research/affine/

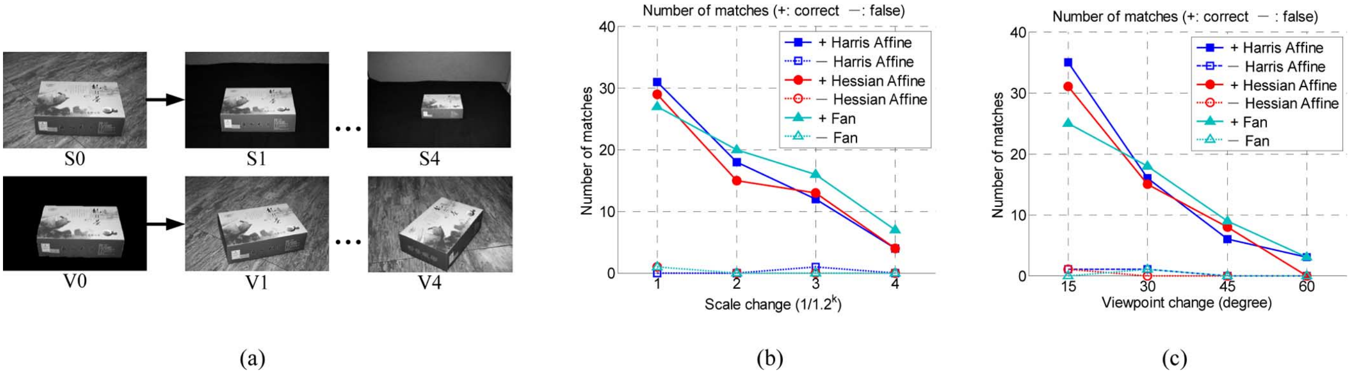


Fig. 13. Test of scale, viewpoint, and background invariance for Box sequence. (a) Test images with different scales and viewpoint angles, (b) Scale and background invariance. (c) Viewpoint and background invariance.

Fig. 12(b) are slightly different. The repeatability of Fan feature falls below that of Harris Affine for small scale change, yet still better than EBR. And the gap between Hessian Affine and Fan feature becomes larger for small and median scale changes. On the other hand, the invariance of the feature under the studied transformation is reflected in the slope of the curves, i.e., how much does a given curve degrade with increasing transformations. In this sense, we can see from Figs. 11(b) and 12(b) that the Fan feature has better invariance to scale and viewpoint changes than the other features.

Finally, both Figs. 11(c) and 12(c) indicate that the correspondences of Fan features are fewer than the other features, especially for small image changes. This is because the Fan features are essentially extracted in a smaller number due to the strict selection that aims to ensure the good repeatability. Yet we note that it still contributes a lot to the quantity of matches in case of strong scale and viewpoint changes. In addition, like all edge-based features, Fan feature performs worse for purely textured scenes such as the Wall and Bark sequences in [2]. For this kind of scenes, Fan feature is therefore not recommended.

B. Scale, Viewpoint and Background Invariance

Images used in the standard test [2] are mostly of planar scenes with no background change or clutter. Experiments presented in this subsection will further take into account the background variation around the surface discontinuity, in addition to the changes of scale and viewpoint. Fig. 13(a) shows the test images (all with resolution 300×200 pixels). The first group (S0 $\sim$ S4) is used to test the scale invariance. Attempt is made to match S0-S1, S0-S2, S0-S3 and S0-S4. The scale changes of the four image pairs are $1/1.2^k$ $(k = 1, 2, 3, 4)$ successively. The second group (V0 $\sim$ V4) is used to test the viewpoint invariance. We try to match the image pairs of V0-V1, V0-V2, V0-V3 and V0-V4, where the viewpoint changes are approximately 15, 30, 45 and 60 degree. All the image pairs have quite different backgrounds, so as to test the background invariance in the meantime. Note that the 3-D box in the test images provides sufficient surface textures to raise Harris and Hessian features. Meanwhile, it is well structured for extracting edge-based features like EBR and Fan feature.

As the image pairs can no longer be related by simple homographies, we now focus on the performance of actual feature matching based on the feature descriptors. Through our experiments, the standard SIFT descriptor is used to describe the Harris Affine, Hessian Affine, and EBR, while the Fan feature uses the Fan-SIFT descriptor instead. The similarity measure based on Euclidean distance and the strategy of *nearest neighbor distance ratio* $(t = 0.8)$ are adopted to initially match these features. We then apply the semi-local filter [37] to reject false matches.

Test results are presented in Fig. 13(b) and (c), where the solid marks indicate the number of correct matches and the hollow ones indicate the false matches. As EBR generates few matches for small scale and viewpoint changes and totally fails in case of large changes, it is not plotted in Fig. 13(b) and (c). From the results, we can see that when there is only slight change in scale or viewpoint, Harris and Hessian Affine produce more correct matches than Fan feature. This again indicates that Fan feature may contribute less to the match quantity in case of small image changes. However, when these changes become more significant, Fan feature tends to preserve more correct matches than Harris and Hessian Affine, which also suggests that the Fan feature has better invariance under scale, viewpoint and background changes.

In Fig. 14, some typical matching results are shown for visual comparison, where the keypoints are represented by the red dots and their associated support regions are represented by the green ellipses. False matches are indicated by red ellipses instead. Note that the support region of Fan feature is only a fan part of the ellipse which can be distinguished by the clear image edges. As we can see in Fig. 14(a)–(c), the keypoints extracted by Harris and Hessian Affine are quite different from those by Fan feature, while EBR find few matches as shown in Fig. 14(d). Hessian Affine generally extracts image blobs. Harris Affine detects image corners, but has little chance to match the corners on or near the object boundaries, because their support regions probably contain different backgrounds. Note that, for a keypoint near the object boundary, Harris and Hessian Affine may adapt its support region to a small or highly deformed one to avoid crossing the surface boundary. However, such features are usually less distinctive or unstable under scale or viewpoint change. In comparison, Fan features are specially designed to save the keypoints on or near surface boundaries. As shown in Fig. 14(c), most keypoints matched by Fan features are lo-

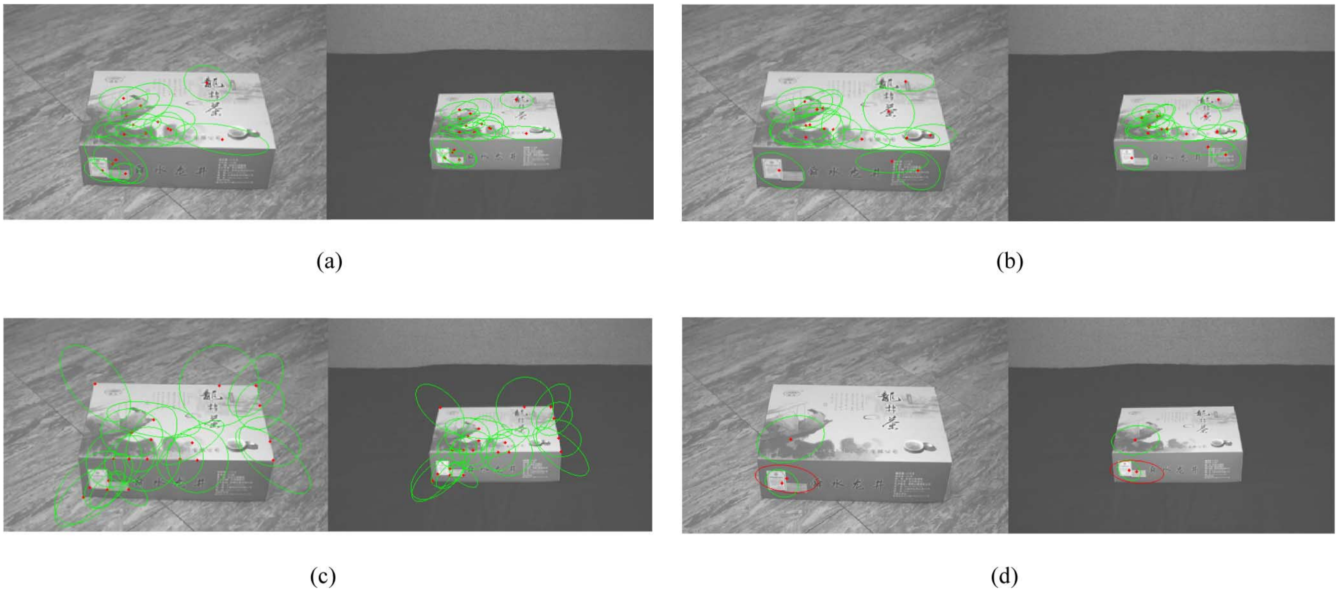


Fig. 14. Selected matching results from the test of scale & background invariance. (a) $\mathbf{S}0$-$\mathbf{S}2$ by Hessian Affine. (b) $\mathbf{S}0$-$\mathbf{S}2$ by Harris Affine. (c) $\mathbf{S}0$-$\mathbf{S}2$ by Fan feature. (d) $\mathbf{S}0$-$\mathbf{S}2$ by EBR.

cated on the box boundaries, including the 3-D box corners. Since the keypoints are extracted from edges, some of them may arise from salient surface textures. In this case, even when the keypoints are close to the surface boundaries, they can still be matched by Fan features provided that one of the subregions is distinctive enough and does not cross the surface boundary. Similar observations can be found as well in the image matching results presented in Section VI-C. In conclusion, the Fan feature is complementary to the classical circular features such as SIFT [11], Harris Affine, and Hessian Affine [12].

C. Image Matching

More matching examples are presented to demonstrate the utility and effectiveness of Fan feature for matching structured scenes with significant scale, viewpoint (pose), and background changes and clutters. The matching results are visually shown in Fig. 15, where the image resolution, the number of correct and false matches and the major difficulties in matching the images are annotated as well. The results of Fan feature are compared with EBR and Harris Affine, and Hessian Affine, which is an efficient combination of both Harris Affine and Hessian Affine . EBR fails for the image pairs in Fig. 15(b) and (c), and hence is not displayed there. As we can see, the Fan feature consistently outperforms the other features in terms of the number of correct matches. It is necessary to point out that besides the strong image changes, the test objects do not possess many distinctive textures, which result in only a small number of correspondences. Since the textures at small scales are not distinctive enough, the support regions need to be enlarged to increase the discriminating power. Large Harris and Hessian features, however, are very likely to cover surface discontinuities and as a consequence cannot be matched in case of changing viewpoints or backgrounds. In comparison, there is less risk of crossing surface discontinuity by matching large subregions. In this sense, Fan feature is superior in matching the weakly textured surfaces under changing viewpoints or backgrounds.

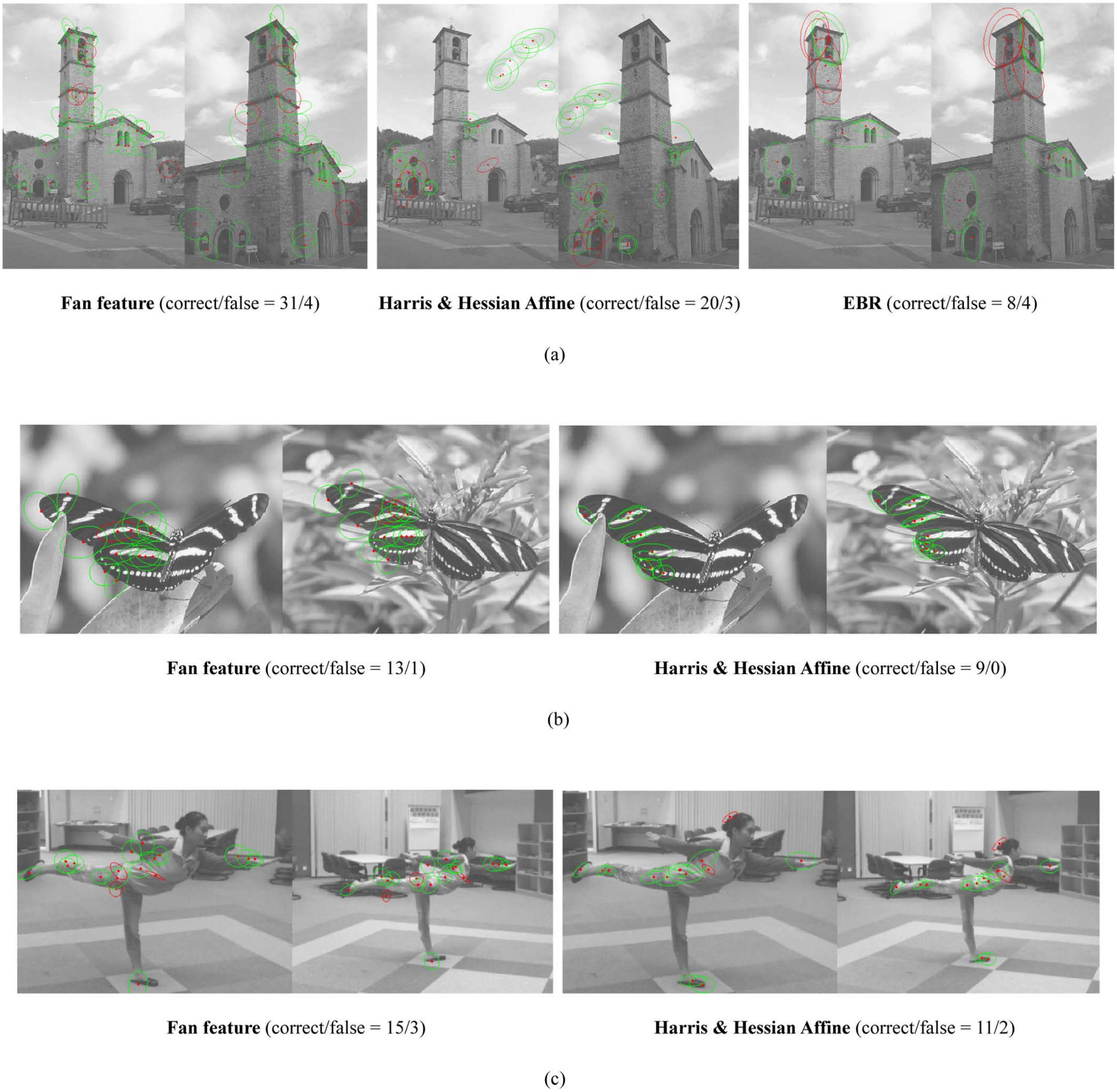
TABLE I
PROPORTION OF REGION OVERLAP

Image pairs	Overlap / Fan	Overlap / Harris & Hessian
Church	0.096	0.122
Butterfly	0.365	0.637
Yoga	0.432	0.419

Another important observation is that the matched regions found by the Fan feature is complementary to those found by other features, which is true even for EBR in Fig. 15(a). Keypoints on surface boundaries are successfully matched by Fan features despite the changing viewpoints and the background clutter. In comparison, most matched Harris and Hessian features are far away from the surface discontinuities. Their support regions are restricted within the object surfaces, unless the covered backgrounds have little change between the images. To further quantify the complementary relationship between the Fan feature and the Harris & Hessian Affine feature, we measure the proportion of overlap of their regions with each other (elliptical regions for Harris and Hessian, fan subregions for Fan feature; for correct matches only). The results are summarized in Table I. We can see that it is a better choice to incorporate both of them to give a more complete representation of image content.

D. Application to Object Rendering From Sparse Views

IBR has been extensively researched in recent years, especially for IBR using densely sampled images [32]. Here, we demonstrate that, by putting the Fan features and Harris and Hessian features together, object rendering with good quality can be achieved using only two wide baseline images as inputs. The object silhouette is only required in one view, while the other view can have annoying background clutter. The main



Fan feature (correct/false = 31/4) **Harris & Hessian Affine** (correct/false = 20/3) **EBR** (correct/false = 8/4)

(a)

Fan feature (correct/false = 13/1) **Harris & Hessian Affine** (correct/false = 9/0)

(b)

Fan feature (correct/false = 15/3) **Harris & Hessian Affine** (correct/false = 11/2)

(c)

Fig. 15. Image matching results of Fan feature and Harris & Hessian Affine. (a) Church (300 × 450): viewpoint change+scale change+lack of texture. (b) Butterfly (400 × 300): pose change+background clutter+homogeneous texture. (c) Yoga (600 × 450): significant viewpoint change (about 75°)+scale change+background clutter+lack of texture.

idea is that both Fan features and Harris and Hessian features can provide not only a few sparse matched keypoints but also the matched ellipse regions that can help infer the local surfaces. First, we perform initial matching by Fan feature and Harris nd Hessian features separately and then put them together into the affine filter and the epipolar test so that they can support each other and consequently output more correct matches, as shown in Fig. 16. Note that, since we know the exact object boundary for the left image, we only extract the Fan features along the object contour and detect the Harris and Hessian features within the contour. Next, we use a method similar to [35] to refine these matched affine features and use them as seeds to propagate more densely and uniformly sampled matching points. Finally, we perform triangulation to generate a high-quality 3-D mesh representation, based on which we combine the textures of the two input images to render the object in a novel view.

The rendering result of an intermediate view is presented in Fig. 17(c). This view is about 25° apart to both the left and right views. For comparison, we also provide the ground truth image captured by our camera in Fig. 17(b) and give the rendering result initialized by Harris and Hessian features alone in Fig. 17(a). As we can observe in the rendering result Fig. 17(a), the girl’s two legs have obvious and annoying deformations compared with the ground truth. This is because Harris and Hessian features cannot provide correct matches that cover these regions to guide accurate feature propagation. By incorporating the Fan features for initialization, such deformations are successfully reduced as shown in Fig. 17(c), and we can hardly ob-

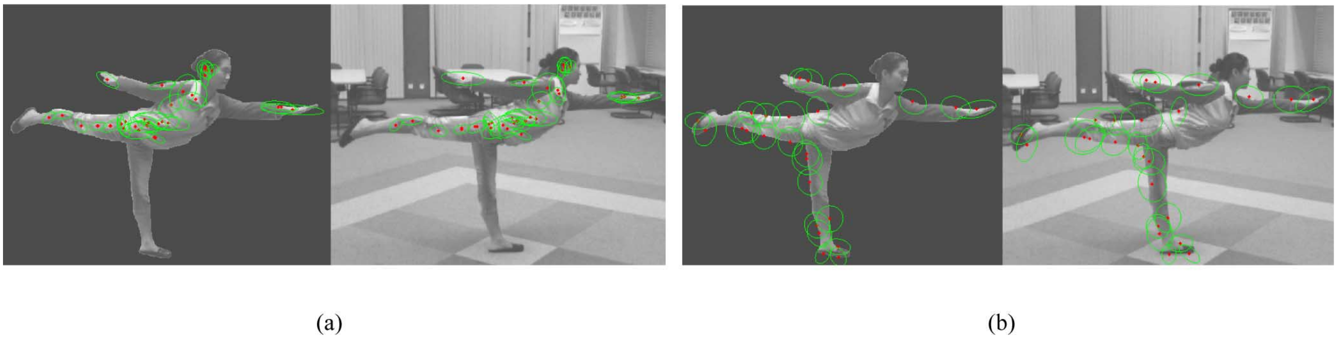


Fig. 16. Correspondences between the masked left image and the right image: (a) by Harris and Hessian Affine extracted within the object contour and (b) by Fan features extracted along the object contour.

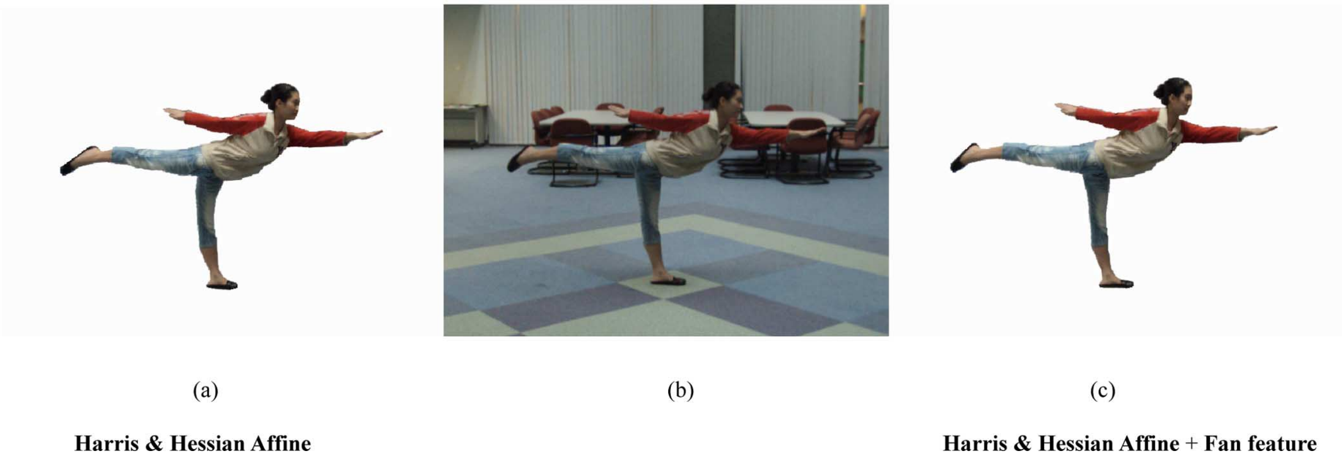


Fig. 17. Rendering the object in the intermediate view from two widely separated views (about 50^circ). (a) Rendering results initialized by Harris and Hessian Affine. (b) Ground truth. (c) Rendering results initialized by Harris and Hessian Affine+Fan feature.

serve the difference between the rendered result and the ground truth, except for the girl's front face for which it is inherently difficult to find any matches between the two wide baseline input images.

Moreover, it is important to point out that the quantity of matched keypoints may be overemphasized in many applications such as recognition and reconstruction. Actually, a small number of matched features with large and representative support regions may be more useful than a lot of small and duplicated features. Here, for example, the initial match quantity does not play a big role. What is crucial is that the initially matched features should cover sufficient regions of the object and that they can survive the significant viewpoint changes.

E. Computational Complexity

The extraction and description of Fan features involve a number of steps. The edge detection is performed at a single scale, and is slightly slower than the standard Canny edge detector due to additional texture suppression and edge cleaning. The Harris measure is computed at five scales, but only for the edge points, so is the nonmaximum suppression (NMS). Keypoint tracking is only for the edge points that survive NMS. Basically, the keypoint selection is very fast and the number of detected keypoints is typically much smaller than Harris and Hessian Affine [10]. For each keypoint, we then perform edge association (sub regions larger than 200 degree are discarded), scale selection (12 scales are explored), affine normalization and Fan-SIFT description. There is no iteration of scale and shape adaptation as is used in Harris and Hessian Affine [10].

TABLE II
COMPUTATION TIMES OF FEATURE EXTRACTION AND DESCRIPTION FOR THE LEFTMOST IMAGE IN FIG. 15(A)

Feature + Descriptor	Run time (sec)	Number of features
Fan feature + Fan-SIFT	1.95	354
Harris Affine + SIFT	5.48	1374
Hessian Affine + SIFT	3.62	926

Table II gives the computation time measured on a Core Duo T2400 1.83 GHz Windows laptop, for the leftmost 300×450 image shown in Fig. 15(a). It also gives the number of features extracted from this image. Though the run time may change depending on the image content, the table can give us a reasonable indication of typical time consumption. Note that the Fan feature is implemented without any optimization. In addition, because the Fan features are extracted in smaller quantity and the Fan-SIFT descriptor has lower dimensions, matching Fan features is much faster than matching Harris and Hessian Affine features.

VII. Conclusion

In this paper, scale- and affine-invariant Fan features are proposed to match the keypoints located on or near surface boundaries. Multiple Fan features are attached to a single keypoint to provide robustness to image content change around the depth discontinuity (including the background change). For each Fan feature, its characteristic scale is selected based on the proposed FLOG kernel. Its affine shape is diagnosed from the mirror-predicted surface patch. In this way, the Fan features can be consistently extracted from two images despite scale change and geometric deformation. Fan-SIFT descriptor is then introduced to depict the feature's texture content. The Fan features are not extracted in a large quantity because the keypoints are carefully selected to guarantee the saliency and repeatability. Experimental results show that the Fan features have good repeatability for structured scenes and have superior invariance to strong scale, viewpoint and background changes. Moreover, the Fan feature is complementary to traditional circular features, especially for describing the surfaces that are weakly textured or close to the object boundaries. The combination of the Fan feature and the Harris and Hessian Affine features shows a promising result of object rendering using wide-baseline images. Adding Fan features into the bag of features may also benefit other applications like object recognition and image retrieval.

Acknowledgment

The authors would like to thank the associate editor and the anonymous reviewers for their insightful comments and suggestions on paper organization and experiment design, which helped to improve the quality of this paper.

References

[1] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *Int. J. Comput. Vis.*, vol. 73, no. 2, pp. 213–238, Jun. 2007.

[2] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 1, no. 65, pp. 43–72, 2005.

[3] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[4] J. Crowley, "A representation for visual information," Ph.D. dissertation, Inst. Robot., Carnegie Mellon Univ., Pittsburgh, PA, 1981.

[5] J. Crowley and A. Parker, "A representation for shape based on peaks and ridges in the difference of low pass transform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 2, pp. 156–170, Mar. 1984.

[6] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vis.*, vol. 30, no. 2, pp. 79–116, 1998.

[7] D. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vision*, 1999, pp. 1150–1157.

[8] K. Mikolajczyk, "Detection of local features invariant to affine transformations," Ph.D. dissertation, Institut National Polytechnique de Grenoble, Grenoble, France, 2002.

[9] K. Mikolajczyk, A. Zisserman, and C. Schmid, "Shape recognition with edge-based features," in *Proc. British Mach. Vis. Conf.*, 2003.

[10] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *Int. J. Comput. Vis.*, vol. 1, no. 60, pp. 63–86, 2004.

[11] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 2, no. 60, pp. 91–110, 2004.

[12] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. British Mach. Vis. Conf.*, 2002, pp. 384–393.

[13] T. Tuytelaars and L. Van Gool, "Matching widely separated views based on affine invariant regions," *Int. J. Comput. Vis.*, vol. 1, no. 59, pp. 61–85, 2004.

[14] T. Kadir and M. Brady, "Scale, saliency and image description," *Int. J. Comput. Vis.*, vol. 45, no. 2, pp. 83–105, 2001.

[15] F. Mindru, T. Moons, and L. Van Gool, "Recognizing color patterns irrespective of viewpoint and illumination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1999, vol. 1, pp. 368–373.

[16] J. J. Koenderink and A. J. van Doorn, "Representation of local geometry in the visual system," *Biol. Cybern.*, vol. 55, pp. 367–375, 1987.

[17] A. Vedaldi and S. Soatto, "Features for recognition: Viewpoint invariance for non-planar scenes," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, vol. 2, pp. 1474–1481.

[18] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1265–1278, Aug. 2005.

[19] T. Lindeberg and J. Garding, "Shape-adapted smoothing in estimation of 3-D shape cues from affine deformations of local 2-D brightness structure," *Image Vis. Comput.*, vol. 15, no. 6, pp. 415–434, 1997.

[20] A. Baumberg, "Reliable feature matching across widely separated views," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2000, pp. 774–781.

[21] D. Shen and H. Horace, "Generalized affine invariant image normalization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 5, pp. 431–440, May 1997.

[22] S. Obdrzalek and J. Matas, "Object recognition using local affine frames on distinguished regions," in *Proc. British Mach. Vis. Conf.*, 2002, pp. 113–122.

[23] P. Forssen and D. Lowe, "Shape descriptors for maximally stable extremal regions," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2007, pp. 1–8.

[24] O. Carmichael and M. Hebert, "Shape-based recognition of wiry objects," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2003, pp. 401–408.

[25] F. Jurie and C. Schmid, "Scale-invariant shape features for recognition of object categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2004, vol. 2, pp. II–90–II-96.

[26] A. Stein and M. Hebert, "Incorporating background invariance into feature-based object recognition," in *Proc. Workshop Applicat. Comput. Vis.*, 2005, vol. 1, pp. 37–44.

[27] D. Martin, C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 530–549, May 2004.

[28] M. Maire, P. Arbelaez, C. Fowlkes, and J. Malik, "Using contours to detect and localize junctions in natural images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[29] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[30] C. Grigorescu, N. Petkov, and M. A. Westenberg, "Contour and boundary detection improved by surround suppression of texture edges," *Image Vision Comput*, vol. 22, no. 8, pp. 609–622, Aug. 2004.

[31] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, 1988, pp. 147–151.

[32] L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, Szeliski, and R. Szeliski, "High-quality video view interpolation using a layered representation," in *Proc. Conf. SIGGRAPH*, 2004, pp. 600–608.

[33] J. Meltzer and S. Soatto, "Edge descriptors for robust wide-baseline correspondence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[34] S. Lazebnik, C. Schmid, and J. Ponce, "Semi-local affine parts for object recognition," in *Proc. British Mach. Vis. Conf.*, 2004, vol. 2, pp. 959–968.

[35] V. Ferrari, T. Tuytelaars, and L. Van Gool, "Simultaneous object recognition and segmentation by image exploration," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 40–54.

[36] A. Vedaldi and S. Soatto, "Viewpoint induced deformation statistics and the design of viewpoint invariant features: Singularities and occlusions," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 360–373.

[37] C. Cui and K. Ngan, "A novel geometric filter for affine invariant features," in *Proc. IEEE Int. Conf. Image Process.*, 2010, pp. 865–868.

Chunhui Cui received the B.E. and M.E. degrees from Huazhong University of Science and Technology, Wuhan, China, in 2004 and 2006, respectively, and the Ph.D. degree in electrical engineering from the Chinese University of Hong Kong, Hong Kong, China, in 2010.

His research interests include image and video processing, image and video coding, and computer vision.

King Ngi Ngan (M'79–SM'91–F'00) received the Ph.D. degree in electrical engineering from the Loughborough University, Loughborough, U.K.

He is currently a Chair Professor with the Department of Electronic Engineering, Chinese University of Hong Kong. He was previously a Full Professor with the Nanyang Technological University, Singapore, and the University of Western Australia, Australia. He holds honorary and visiting professorships with numerous universities in China, Australia, and Southeast Asia. He is an associate editor of the *Journal on Visual Communications and Image Representation*, as well as an area editor of *EURASIP Journal of Signal Processing: Image Communication* and an associate editor for the *Journal of Applied Signal Processing*. He has published extensively including three authored books, five edited volumes, over 300 refereed technical papers, and edited nine special issues in journals. In addition, he holds 10 patents in the areas of image/video coding and communications.

Prof. Ngan is a Fellow of the IET and IEAust (Australia) and was an IEEE Distinguished Lecturer during 2006–2007. He served as an associate editor of the IEEE *Transactions on Circuits and Systems for Video Technology*. He chaired a number of prestigious international conferences on video signal processing and communications, and served on the advisory and technical committees of numerous professional organizations. He was a general co-chair of the IEEE International Conference on Image Processing (ICIP), Hong Kong, September 2010.