Defocus Blur-Invariant Scale-Space Feature Extractions

Elhusain Saad, Member, IEEE, Keigo Hirakawa, Senior Member, IEEE

Abstract—We propose modifications to scale-space feature extraction techniques (Scale-Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF)) that make the feature detection and description invariant to defocus blur. Specifically, scale-space blob detection relies on the second derivative responses of images. Our analysis of circular defocus blur (which sufficiently approximates a real camera blur kernel) and its effect on scale-space blob detection suggests that fourth derivative—and not the usual second derivative—is optimal for detecting the blurred blobs while multi-scale descriptors of blurred blobs are effective at establishing correspondences between blurred images. The proposed defocus blur-invariant (DBI) scale-space feature techniques—which we refer to as DBI-SIFT and DBI-SURF—do not require image deblurring nor blur kernel estimation, meaning that their accuracy does not depend on the quality of image deblurring. We offer empirical evidence of blur invariance by establishing interest point correspondences between sharp or blurred reference images and blurred target images.

Index Terms—SIFT, SURF, Defocus Blur.

I. INTRODUCTION

IMAGE correspondence is a fundamental tool in a variety of computer vision tasks, including object recognition, image registration, tracking, and stereo matching. It is a difficult task, since pixel-by-pixel correspondences fail with variations in lighting, perspective, rotation, noise, camera parameters, blur, etc. Instead, modern correspondence methods make use of distinctive and robust invariant features that describe the local region by unique signature. High level image features describing the shape are more dependable for real scenarios [2].

Local features can be defined as an image pattern with properties that differs from the surroundings. The detector detects specific intensity patterns from an image (e.g. edges, corners, or blobs) and their spatial locations are recorded as interest points. Since these are intrinsic attributes of a scene, the features should be repeatable, stable, and local under geometric and photometric transformations. In order that interest point detection be useful, it must also be distinctive, have enough sensitivity to detect features in low illumination, and have accurate localization [2]–[5]. Each region around the interest point can be described by a descriptor, where a match based on descriptor similarities establishes accurate correspondences that are useful for a variety of recognition, tracking, and detection tasks.

A variety of interest points and their descriptors have been proposed in the past [2]–[9]. There have also been a number of rigorous studies to assess the relative invariance to scale, rotation, noise, etc [4]. Indeed, to a large extent, many techniques are found to be robust to these deformations and degradations. Scale-Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) are considered to be the state-of-art method to detect and describe a blob [3], [5]—a group of pixels whose intensities differ from the pixel intensities of the surrounding area. As we explain in Section II, blobs have shape similar to Laplacian of Gaussian (LOG), playing the role of a matched filter for blob detection in SIFT and SURF, while the scale of LOG can be varied to match the blob sizes. A bilevel refinement to LOG-type filters has been shown to improve the feature detection repeatability [10], [11]. Maximally Stable Extremal Regions (MSER) is an alternative to LOG-type approach to blob detection [7]. It assesses the changes to the contiguous grouping of pixels over varying intensity threshold values, under the assumption that blobs are insensitive to modest changes in thresholding. Among other works of scale-space feature detection is Binary Robust Invariant Scalable Keypoints (BRISK), aimed at assessing saliency scale-space [6] as described by FAST [12]. Other popular feature extraction techniques include detection and description of corners. Harris detector is a classical detection scheme based on second moment matrix of derivatives that is said to be invariant to illumination and rotation [13]. Harris-Laplace detector combines the classical corner detection scheme with LOG to make the detector invariant to scale and view angle [4], while the segment test assesses pixel intensities along a circular path around the corner [12], [14]–[16].

Among the intensity pattern descriptors, SIFT, Gradient Location-Orientation Histogram (GLOH), and DAISY takes cues from weighted histogram of gradient magnitudes and orientations in the vicinity of the detected blobs [3], [17], [18], while statistics of gradient intensities at various scales comprise descriptor vectors of SURF [5]. Geometric shape, line, and angle descriptors were considered in [19]. Recent focus on binary descriptors based on intensity derivative/bandpass/difference signs [6], [8], [16], [20]–[22], intensity orderings [9], [23], [24], and hypothesis testing [25] are aimed at reducing computational complexity and increasing robustness to more complex illumination changes. By orientation ordering and imposing symmetries in pooling the spatially local statistics, these descriptors are shown to be
Fig. 1. Example of matching the interest points detected over blurred reference image with the interest points detected over a blurred target image. (a) Failed detection and matching using SIFT with RANSAC. (b) Successful detection and matching using the proposed DBI-SIFT with RANSAC.

Fig. 2. Example showing stitching blurred images 1 and 2 using SIFT and the proposed DBI-SIFT.

largely invariant to deformations such as zoom, rotation, and view angle changes (to varying degrees) when defocus blur is negligible [26], [27].

Yet, despite the claims of the contrary in the prior investigations (which only considered modest levels of blur), however, the effects of blur on feature extraction are more harmful than other deformations such as view angle, scale, rotation [4], [5], [17], [28]–[30]. Specifically, as the severity of defocus increases, the features that we rely on to detect intensity patterns such as blobs becomes distorted; and interest point descriptor vectors loose discriminability, increasing the likelihood of false matches and misses. The experimental evidences presented in Section IV confirm this, as the repeatability scores (detection quality), precision (descriptor quality), and recall (descriptor quality) of even the newest interest point detection/description techniques decay with the increasing severity of defocus blur (perhaps with the exception of the method in [4]). Figure 1 also illustrates the influence of blur on the applications of image correspondence tasks—attempts to match blurred objects using existing techniques clearly fail. As evidenced by Figure 2, applications such as image stitching that rely on image correspondences are also sensitive to blur as well.

One obvious remedy to this problem is to perform blind image deblurring—a process of recovering the latent sharp image from the given blurry image—as a preprocessing step to image correspondence [31], [32]. However, this is an undesirable solution because deblurring is computationally expensive and the reconstructed images are far from perfect. Furthermore, most existing image deblurring algorithms do not handle multiple depth/spatially varying defocus blur, limiting the deblurring approach to constant depth scenes only. Alternatively, there exists a body of literature on the topic of blur invariants, aimed at recovering attributes of images that are insensitive to blur [33], [34]. The blur invariant moment statistics features extracted from the blurred image are blur kernel-specific, rather than the intensity pattern- or shape-specific features such as the scale-space features under the consideration of this paper. Although blur-invariant statistics have been used successfully in image registration, stereo matching, and template matching, work along this line does not naturally lead to detection/description of specific intensity patterns such as blob and corner in the presence of defocus blur. As such, the appropriateness of choosing the blur invariant moment statistics or the proposed DBI-SIFT/DBI-SURF depends on the needs of the application (which is beyond the scope of this paper). Another approach is to develop intensity pattern-based feature extraction techniques that are invariant to blur. Such attempts have been made in video tracking [17], [30], [35], region entropy-based feature extraction [29], [32], [35], [36], and corner detection [37]. Yet, some of these techniques may not be invariant to other photometric or geometric transforms. Overall, studies of blur-invariant feature extraction has been limited to date, and further investigation is warranted.

The main contribution of this paper is a modification we introduce to SIFT and SURF that makes the interest point detection and descriptors robust to defocus blur. The blob detection meets the strict criteria for “invariance” in the sense that the local maxima of modified Hessian determinants/LOG are expected to be in the same spatial location when the blur kernel shape is symmetrical. The modified blob descriptors are asymptotically invariant to defocus blur, in the sense that the corresponding modified descriptor vector computed from defocus blurred images and the conventional SIFT/SURF descriptors computed from sharp images converge exponentially as a function of the blob size. The contribution is significant as it gives rise to the possibility that objects in a blurry image can be recognized. To verify the claim, we assess the performance of the image correspondence techniques for varying degrees of blur (and combinations of blur and other...
deformations such as zoom, rotation, and viewing angle) using
real sensor data. Although a handful of newer interest point
detectors and descriptors [4], [6]–[8], [16], [20], [21] have
been proposed since the introduction of SIFT/SURF (many
focused on speed), we emphasize that SIFT and SURF are
widely used in a variety of computer vision applications
today because of their robustness and accuracy [11], [18], [28], [38]–
[45]. Thus the proposed improvements on SIFT/SURF have
a potential to widely influence computer vision applications
that already utilize SIFT/SURF. The proposed algorithms
show good matching performance when the defocus blur is
severe comparing with the SIFT/SURF, as illustrated in Fig.
1 and Fig. 2. Also, our experiments also demonstrate that
our proposed modifications to SIFT/SURF are more robust
to defocus blur than the more recently proposed interest point
detectors and descriptors [4], [6]–[10], [16], [20], [37].

The remainder of the paper is organized as follows. In
Section II, SIFT and SURF are analyzed with defocus blur.
The proposed defocus blur-invariant SIFT and SURF detectors
and descriptors are introduced in Section III. In Sections IV
and IV, we develop and provide empirical evidence to verify
the claims by matching between feature descriptors found in
a blurred image and a sharp image. We draw conclusions in
Section VI. We emphasize that all images shown in this
paper are taken from real cameras (i.e. no simulated blur).
We also emphasize that the proposed technique is designed
to counter defocus blur only—meaning it provides no benefit
to images corrupted by non-circular blur kernels (caused by
object motion, camera shake, etc.).

II. BLUR ANALYSIS OF SCALE-SPACE FEATURE
EXTRACTION

In this section, we analyze the impact of defocus blur on the
state-of-art scale-space feature detection methodologies. While
requisite mathematics for SIFT and SURF will be covered, we
refer the readers to [3], [5] for more detailed explanations.

A. SIFT interest point detection and descriptor

Let \( I : \mathbb{Z}^2 \rightarrow \mathbb{R} \) be the latent sharp image, where \( I(x) \) is
the pixel intensity at location \( x = (x, y) \in \mathbb{Z}^2 \). Define \( G(x, \sigma) \) as
a two-dimensional Gaussian function with scale \( \sigma \):

\[
G(x, \sigma) = \frac{1}{2\pi\sigma^2} \exp \left( -\frac{(x^2 + y^2)}{2\sigma^2} \right),
\]

and the corresponding Laplacian of Gaussian (LOG):

\[
\nabla^2 G(x, \sigma) = \left\{ \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right\} G(x, \sigma) = G_{xx}(x, \sigma) + G_{yy}(x, \sigma),
\]

where \( G_{xx}(x, \sigma) := \frac{\partial^2}{\partial x^2} G(x, \sigma) \), etc. The LOG response of
an image \( I \) is therefore

\[
I(x) \star \nabla^2 G(x, \sigma) = I_{xx}(x, \sigma) + I_{yy}(x, \sigma),
\]

where \( \star \) denotes the spatial convolution operator, and
\( I_{xx}(x, \sigma) := I(x) \star G_{xx}(x, \sigma) \) is the response to the Gaussian
derivative operator \( I_{yy} \) and \( I_{xy} \) defined similarly. Then LOG
has matched filter-like qualities for disk-shaped ideal blob,
attaining a local maximum response at the center of the blob
\( x_0 \) when Gaussian scale \( \sigma \) and blob radius \( r \) have the relation
\( \sigma = r/\sqrt{2} \) [2]:

\[
\max_{x, \sigma} I(x) \star \nabla^2 G(x, \sigma) = \left\{ I(\cdot) \star \nabla^2 G \left( \cdot, \frac{r}{\sqrt{2}} \right) \right\}(x_0).
\]

Using the principle of heat equation, LOG in (2) can be ap-
proximated efficiently by the Difference of Gaussians (DOG)
[3]:

\[
\nabla^2 G(x, \sigma) \approx \frac{G(x, \sqrt{2}\sigma) - G(x, \sigma)}{\sigma^2 \sqrt{2} - \sigma^2}.
\]

Substituting (5) to (3) results in a simplification:

\[
I(x) \star \nabla^2 G(x, \sigma) \approx \frac{I(x) \star G(x, \sqrt{2}\sigma) - I(x) \star G(x, \sigma)}{\sigma^2 \sqrt{2} - \sigma^2}.
\]

Hence, the local extrema of DOG is declared as SIFT interest points
identifying “blobs.”

Define \( G_x(x, \sigma) = \frac{\partial}{\partial x} G(x, \sigma) \) and \( G_y(x, \sigma) = \frac{\partial}{\partial y} G(x, \sigma) \).
The classical SIFT descriptor is based on the gradient magni-
itude \( M(x, \sigma) \) and orientation \( \theta(x, \sigma) \):

\[
M(x, \sigma) = \sqrt{I_x(x, \sigma)^2 + I_y(x, \sigma)^2},
\]

\[
\theta(x, \sigma) = \arctan \left( \frac{I_y(x, \sigma)}{I_x(x, \sigma)} \right),
\]

where \( I_x(x, \sigma) := I(x) \star G_x(x, \sigma) \), etc. SIFT interest point
detector is a vector of the form

\[
U(x, \sigma) = \left( \frac{U_1(x, \sigma)}{||U_1(x, \sigma)||}, \ldots, \frac{U_K(x, \sigma)}{||U_K(x, \sigma)||} \right)^T, \]

comprised of \( K \) sub-descriptors \( U_k(x, \sigma) \). Each sub-descriptor
is an orientation histogram weighted by \( M(x, \sigma) \) [3] computed
within a series of subregions \( \Lambda_k \subset \mathbb{Z}^2 \) as defined in [5]:

\[
U_k(x, \sigma) = \sum_{m \in \Lambda_k} M(x + m, \sigma) \left( \begin{array}{c} 1_{A_1}(\theta(x + m, \sigma)) \\ 1_{A_2}(\theta(x + m, \sigma)) \\ \vdots \\ 1_{A_N}(\theta(x + m, \sigma)) \end{array} \right),
\]

where \( A_1, \ldots, A_N \) are the \( N \) bins of orientation histogram, and

\[
1_A(\theta) = \begin{cases} 0 & \text{if } \theta \not\in A \\ 1 & \text{if } \theta \in A. \end{cases}
\]

is an indicator function. Reassigning orientation angle \( \theta \) relative
to the dominant angle makes this weighted histogram rotation
invariant. Normalization of this histogram in (8) makes it illumination
invariant as well [3].

B. SURF interest point detection and descriptor

Interpreting \( I_{xx}(x, \sigma) \) as a proxy for \( \frac{\partial^2}{\partial x^2} I(x) \), etc., the
Taylor series expansion of \( I(x + \Delta x) \) is:

\[
I(x + \Delta x, \sigma) \approx I(x, \sigma) + \Delta x^T \begin{pmatrix} I_x(x, \sigma) \\ I_y(x, \sigma) \end{pmatrix} + \frac{1}{2} \Delta x^T \begin{pmatrix} I_{xx}(x, \sigma) & I_{xy}(x, \sigma) \\ I_{yx}(x, \sigma) & I_{yy}(x, \sigma) \end{pmatrix} \Delta x + \ldots.
\]
where the matrix $\mathcal{H}(x, \sigma) \in \mathbb{R}^{2 \times 2}$ is known as Hessian. Suppose $x_0$ is a local extrema. Then $\nabla I(x_0) = 0$ because $I_x(x_0) = I_y(x_0) = 0$. Then the difference between the extrema $I(x_0)$ and a surrounding pixel $I(x_0 + \Delta x)$ can be approximated as:

$$I(x_0 + \Delta x, \sigma) - I(x_0, \sigma) \approx \frac{I_{xx}(x_0, \sigma)}{2} \Delta x^2 + \frac{I_{xy}(x_0, \sigma)}{2} \Delta x \Delta y + \frac{I_{yy}(x_0, \sigma)}{2} \Delta y^2,$$

where $\Delta x = (\Delta x, \Delta y) \in \mathbb{Z}^2$. The roots of this equation are the local extrema locations, and it can be found by the discriminant:

$$I_{xx}(x_0, \sigma)I_{yy}(x_0, \sigma) - I_{xy}(x_0, \sigma)^2,$$

which is also the determinant of the Hessian matrix $\det(\mathcal{H}(x_0, \sigma))$. Hence if $\det(\mathcal{H}(x_0, \sigma)) > 0$, intensity pattern is assumed to be the extrema of $I(x)$ at $x = x_0$. If $\det(\mathcal{H}(x_0, \sigma)) < 0$, it denotes a saddle point while $\det(\mathcal{H}(x_0, \sigma)) = 0$ is inconclusive. The determinant and the trace of $\mathcal{H}(x, \sigma)$ computed with different scale of LOG are used to identify the extrema of $I(x)$ and recorded as the SURF interest point. Bay used a Haar-type approximation of the Gaussian derivative to speed up the implementation [5].

SURF defines the interest point descriptor for interest point at $x$ as vector

$$V(x, \sigma) = \left( \frac{V_1(x, \sigma)}{\|V_1(x, \sigma)\|}, \ldots, \frac{V_K(x, \sigma)}{\|V_K(x, \sigma)\|} \right)^T,$$

where $k$th sub-descriptor is

$$V_k(x, \sigma) = \sum_{m \in \Lambda_k} \begin{pmatrix} I_x(x + m, \sigma) \\ I_y(x + m, \sigma) \\ I_x(x + m, \sigma) \\ I_y(x + m, \sigma) \end{pmatrix}.$$

Again, reassigning subregions $\Lambda_k \subset \mathbb{Z}^2$ and the pixel coordinates relative to the dominant angle makes this feature vector rotation invariant. Normalization in (14) makes it illumination invariant as well [3].

C. Blur types

Image blur is caused by a pixel recording light from multiple sources. Assuming Lambertian surfaces, blur is typically represented by the implied blur kernel that acts on the unobserved sharp in-focus image [31], [32], [35], [46]. There are three common types of blur: camera shake, object motion and defocus. Camera motion during the exposure results in near-global motion blur where the same point in the scene is observed by multiple moving pixels [31], [32]. Object motion causes each pixel to observe multiple points in the scene, resulting in spatially varying motion blur proportional in length to the object speed [46]. Defocus blur is caused by a wide aperture that prevents light rays originating from the same point in the scene from converging at the focal plane. The defocus blur kernel varies with the object distance/depth, and is therefore spatially varying. While defocus blur can be useful for three dimensional scene reconstruction from a single camera [46], [47], it also interferes with feature extraction [36], [46]–[48].

The support of the defocus blur kernel $B(x)$ takes the shape of the aperture opening, which is a circular disk in most typical cameras. Though $B(x)$ may not be known exactly, the following approximation is well accepted [46], [49], [50]:

$$B(x) \approx \begin{cases} \frac{1}{\pi q^2} & \|x\| \leq q, \\ 0 & \text{otherwise} \end{cases}$$

where $q$ is the radius of the defocus blur disk (a.k.a. circle of confusion). Then the blurred image $\tilde{I}(x)$ is modeled as:

$$\tilde{I}(x) = B(x) \ast I(x) + Z(x),$$

where $I(x)$ is the sharp image and $Z(x)$ is a realization of noise.

D. Scale-Space Interest Point Detectors Are Not Defocus Blur Invariant

Let us now examine the effects of defocus blur kernel on LOG and Hessian matrix. Suppose we compute the derivatives of $\tilde{I}$ instead of $I$:

$$\tilde{\partial_x^2 \tilde{I}}(x) = \left\{ B(\cdot) \ast \left( \tilde{\partial_x^2 I}(\cdot) \right) \right\}(x) + \tilde{\partial_x^2}Z(x).$$

Or by the way of Gaussian second derivative,

$$\tilde{I}_{xx}(x, \sigma) := \tilde{I}(x) \ast G_{xx}(x, \sigma) = B(x) \ast I(x) \ast G_{xx}(x, \sigma) + Z(x) \ast G_{xx}(x, \sigma) = B(x) \ast I_{xx}(x, \sigma) + Z_{xx}(x, \sigma),$$

where $Z_{xx}(x, \sigma) := Z(x) \ast G_{xx}(x, \sigma)$ (similar analysis applies to $\tilde{I}_{xy}$ and $\tilde{I}_{yy}$). We may conclude that the second derivative responses $\{I_{xx}, \tilde{I}_{xx}, I_{yy}\}$ of $I$ are blurred (and slightly noisier) versions of the ideal second derivatives $\{I_{xx}, I_{xy}, I_{yy}\}$.

Substituting (19) to (3) yields an interesting analysis of LOG response of the blurred image $\tilde{I}$:

$$\tilde{I}(x) \ast \nabla^2 G(x, \sigma) = \tilde{I}(x) \ast G_{xx}(x, \sigma) + \tilde{I}(x) \ast G_{yy}(x, \sigma) = B(x) \ast (I_{xx}(x, \sigma) + I_{yy}(x, \sigma)) + Z_{xx}(x, \sigma) + Z_{yy}(x, \sigma) = B(x) \ast (I(x) \ast \nabla^2 G(x, \sigma)) + Z(x) \ast \nabla^2 G(x, \sigma).$$

Thus we conclude that LOG response of a blurred image $\tilde{I}$ is a blurred LOG response of a sharp image $I$.

Suppose that $x_0$ is a local maximum of $I \ast \nabla^2 G$ (i.e. SIFT interest point of a sharp image). Given a blur radius of $q$ and with no noise, $x_0$ is also a local maximum of $B \ast (I \ast \nabla^2 G)$ (i.e. SIFT interest point of a blurred image) when the LOG response $I \ast \nabla^2 G$ is monotonically increasing toward $x_0$ over a circular neighborhood of radius $2q$ around $x_0$. That is,

$$\{I(\cdot) \ast \nabla^2 G(\cdot, \sigma)\}(x_0 + m) \geq \{I(\cdot) \ast \nabla^2 G(\cdot, \sigma)\}(x_0) + n$$

for all $m, n \in \mathbb{Z}^2$ satisfying $\|m\|_2 < \|n\|_2 \leq 2q$ (justification in Appendix). Since the condition in (21) is harder to meet when $q$ is large, the likelihood that $I \ast \nabla^2 G$ and $\tilde{I} \ast \nabla^2 G$ share the interest point $x_0$ decreases rapidly as blur radius $q$ increases. We conclude that SIFT interest point detection is not defocus blur invariant.

Similar analysis applies to the determinant of Hessian matrix. Supposing that $x_0$ is a local maximum of $\det(\mathcal{H}(x)),$
one important question to ask is whether $x_0$ would also emerge as the local maximum of the determinant $\det(\mathcal{H}(x, \sigma)) = I_{xx}(x, \sigma)I_{yy}(x, \sigma) - I_{xy}(x, \sigma)^2$. Unfortunately, when the pixel $x_0$ (locally) maximizes $\det(\mathcal{H}(x, \sigma))$, we expect to have satisfied the condition

$$\det(\mathcal{H}(x_0 + m, \sigma)) \geq \det(\mathcal{H}(x_0 + n, \sigma)) \tag{22}$$

for all $m, n \in \mathbb{Z}^2$ satisfying $\|m\|_2 < \|n\|_2 < 2q$ (justification in Appendix). That is to say that $\det(\mathcal{H}(x, \sigma))$ is monotonically increasing towards $x = x_0$ over a circular neighborhood of radius $2q$. As a result, we expect even with no noise that the number of interest points surviving the blur decays rapidly as the size of the blur kernel grows. Of course, the requirement of (21) and (22) become even more difficult with noise. We conclude that SURF interest point detection is also not invariant to blur.

E. Scale-Space Descriptors Are Asymptotically Defocus Blur Invariant

Denote by $\widetilde{U}(x, \sigma)$ and $\widetilde{V}(x, \sigma)$ the SIFT and SURF interest pointer descriptor corresponding to the blurred image $\widetilde{I}$, respectively. Let $\sigma$ denote the scale of a sharp blob—we cannot know $\sigma$ exactly from a blurry image with blurred blobs, obviously. But for now let us assume $\sigma$ is available. A question central to the context of blur is whether $U(x, \sigma)$ and $\widetilde{U}(x, \sigma)$ (or $V(x, \sigma)$ and $\widetilde{V}(x, \sigma)$) are similar enough to establish a match. For this, we rely on the Fourier analysis.

Drawing on the analysis of (19), we have the relation

$$\widetilde{I}_x(x, \sigma) = B(x) \ast I_x(x, \sigma) + Z_x(x, \sigma), \tag{23}$$

that is, first derivatives $\{I_x, I_y\}$ of $I$ are blurred (and slightly noisier) versions of the ideal first derivatives $\{I_x, I_y\}$. Let $i, i_x, i_y, b, g, z_x : \{\mathbb{R}/2\pi\}^2 \to \mathbb{C}$ be the discrete space Fourier transform of $I, I_x, I_y, B, G, Z_x$, respectively, where $i(\omega)$ is the Fourier coefficient at frequency $\omega = (\omega_x, \omega_y) \in \{\mathbb{R}/2\pi\}^2$, etc. Then by convolution theorem, we have the relation:

$$\widehat{i}_x(\omega, \sigma) = b(\omega) \cdot i_x(\omega, \sigma) + z_x(\omega) = b(\omega) \cdot \{j \cdot \omega_x \cdot g(\omega, \sigma) \cdot i(\omega)\} + z_x(\omega). \tag{24}$$

Here, $j \cdot \omega_x \cdot g(\omega, \sigma)$ is the Fourier transform of Gaussian derivative filter $G_x(x, \sigma)$ stemming from the relation [51]:

$$\frac{\partial^n}{\partial x^n} G(x, \sigma) \to (\omega_x)^n g(\omega, \sigma) = (\omega_x)^n e^{-\frac{\omega^2}{2}}. \tag{25}$$

Thus, differentiated Gaussian results in a bandpass filter where the pass band width and center frequency scale inversely with the Gaussian scale $\sigma$. One can show that the frequency center $\omega_o$ is:

$$\omega_o = \pm \frac{\sqrt{n}}{\sigma}. \tag{26}$$

Together with the fact that frequency response of blur kernel $b(\omega)$ is relatively smooth, the following approximations to (24) and (23) hold when $\sigma$ sufficiently large:

$$\widehat{i}_x(\omega, \sigma) \approx b(\frac{\omega}{\sigma}) \cdot \{j \cdot \omega_x \cdot g(\omega, \sigma) \cdot i(\omega)\} + z_x(\omega), \tag{27}$$

where $b(\frac{\omega}{\sigma})$ is the magnitude at center frequency of the pass band $\omega_o$. The approximate relation in (27) asymptotically approaches equality as $\sigma$ becomes large. (See justification provided in Appendix.) With the interpretation of (27) that $I_x(x, \sigma)$ is attenuated by the frequency response $b(\frac{\omega}{\sigma})$ (i.e. a constant), we arrive at the conclusion that the SURF interest point descriptor $V(x, \sigma)$ is...
asymptotically invariant to blur when noise is sufficiently low:
\[ \tilde{V}(x, \sigma) \approx \left( \frac{b(\frac{1}{\sigma})V_1(x, \sigma)}{\|b(\frac{1}{\sigma})V_1(x, \sigma)\|}, \ldots, \frac{b(\frac{1}{\sigma})V_K(x, \sigma)}{\|b(\frac{1}{\sigma})V_K(x, \sigma)\|} \right)^T \]
\[ = V(x, \sigma). \tag{28} \]

Note, however, that division by \( b(\frac{1}{\sigma}) \) can also boost noise. It is straightforward to extend the above conclusions to the Haar wavelet approximation of Gaussian derivative filters also.

A similar analysis leads to the notion of defocus blur-invariance for SIFT descriptor. Substituting (27) into (7), we can show that the orientation is unaffected by defocus blur:
\[ \tilde{\theta}(x, \sigma) \approx \tan^{-1}\left( \frac{b(\frac{1}{\sigma})I_y}{b(\frac{1}{\sigma})I_x} \right) = \tan^{-1}\left( \frac{I_y}{I_x} \right) = \theta(x, \sigma). \tag{29} \]

Similarly for the magnitude, we have the relation:
\[ \tilde{M}(x, \sigma) \approx \sqrt{\left( \frac{b(\frac{1}{\sigma})I_x(x, \sigma) \right)^2 + \left( \frac{b(\frac{1}{\sigma})I_y(x, \sigma) \right)^2}} = b(\frac{1}{\sigma})M(x). \tag{30} \]

Subsequent normalization of the descriptor vector \( \tilde{U}(x, \sigma) \) divides out \( b(\frac{1}{\sigma}) \) from \( \tilde{M}(x) \):
\[ \tilde{U}(x, \sigma) \approx \left( \frac{b(\frac{1}{\sigma})U_1(x, \sigma)}{\|b(\frac{1}{\sigma})U_1(x, \sigma)\|}, \ldots, \frac{b(\frac{1}{\sigma})U_K(x, \sigma)}{\|b(\frac{1}{\sigma})U_K(x, \sigma)\|} \right)^T \]
\[ = U(x, \sigma). \tag{31} \]

We conclude that SURF and SIFT descriptors are asymptotically defocus blur-invariant, but only if the scale \( \sigma \) of the unobserved sharp blob is known, and noise level is not reasonably high (no free lunch here). Though this is encouraging, question remains as to how to handle an appropriate choice of \( \sigma \) that enables the approximation \( \tilde{V}(x, \sigma) \approx V(x, \sigma) \) and \( \tilde{U}(x, \sigma) \approx U(x, \sigma) \). We address this issue in Section III-C.

### III. Proposed: Defocus Blur-Invariant Scale-Space Feature Extraction

#### A. Defocus Blur-Invariant Scale-Space Derivatives

Based on the blur analysis of the previous section, we propose novel technique to detect interest points using scale-space detectors from defocus blurred images. Together with the blur-invariant interest point descriptor matching we propose in Section III.B, blur-invariant scale-space feature extraction gives rise to the possibility that objects in a blurry image can be detected and recognized.

Recall (4)—the disk-shaped ideal blob attains a local maximum response \( I_{xx}(x, \sigma) \) at the center of blob \( x = x_0 \) with Gaussian scale \( \sigma = r/ \sqrt{2} \) where \( r \) is the blob radius. This matched filter-like response to Gaussian second derivative \( I_{xx}(x, \sigma) \), is shown in Figure 3(b). However, as illustrated by Figure 3(e), \( I_{xx} \) fails to yield a local maximum at the center of the blob when defocus blur \( B \) is introduced—consistent with our findings in Section II-D.

Consider now the fact that the point spread function \( B(x) \) of a defocus blur also has a disk shape owing to the circular aperture in a typical camera lens system (i.e., the so-called

circle of confusion)—very similar to the ideal blob shape, in fact. It follows, then, that the Gaussian second derivative filter \( G_{xx}(x, \nu) \) applied to blur kernel \( B(x) \) with blur radius \( q \) also attains a sharp maximum at the blur center \( x = 0 = (0, 0)^T \) and scale \( \nu = q/ \sqrt{2} \):
\[ \max_{x, \nu} \left\{ \left( \frac{\partial^2}{\partial x^2} G(\cdot, \nu) \right) \ast B(\cdot) \right\}(x) = B_{xx}(0, q/ \sqrt{2}). \tag{32} \]

Leveraging (32), consider Gaussian fourth derivative filter applied to the blurred ideal blob \( \tilde{I} \):
\[ \tilde{I}_{xxxx}(x, \zeta) = \left\{ \left( \frac{\partial^2}{\partial x^2} G(\cdot, \zeta) \right) \ast \tilde{I}(\cdot) \right\}(x) \]
\[ = \left\{ \left( \frac{\partial^2}{\partial x^2} G(\cdot, v) \right) \ast B(\cdot) \right\} \ast \left\{ \left( \frac{\partial^2}{\partial x^2} G(\cdot, \sigma) \right) \ast I(\cdot) \right\}(x) \]
\[ + \left( \frac{\partial^2}{\partial \nu^2} G(\cdot, v) \right) \ast \left\{ \left( \frac{\partial^2}{\partial x^2} G(\cdot, \sigma) \right) \ast Z(\cdot) \right\}(x) \]
\[ = B_{xx}(x, v) \ast I_{xx}(x, \sigma) + G_{xx}(x, v) \ast Z_{xx}(x, \sigma). \tag{33} \]

Though far from perfect, the peak of \( B_{xx} \) approximate an impulse function better than \( B \), the end result of which is the softening of the conditions (21) and (22) that causes SIFT and SURF interest point detection to be sensitive to defocus blur. Specifically, \( I_{xxxx}(x_0, \sigma) \) has local maxima at \( x_0 \) if (21) is satisfied for \( \|n\|_2 < \|n\|_2 < 2p \), where \( p \) is the width of \( B_{xx} \) peak. Since \( p < q \), condition in (21) is more easily met with Gaussian fourth derivative than with the usual second derivative. As for noise, \( Z_{xx}(x, \sigma) \) is the usual amount of noise seen in a second derivative detector. The additional second derivative filter \( G_{xx}(x, v) \) is the “penalty” paid by the scheme in (33). While the transformation of \( B \mapsto B_{xx} \) or \( \tilde{I} \mapsto \tilde{I}_{xx} \) may not suffice as image deblurring technique, it is adequate for preserving local extremal structure.

#### B. Defocus Blur-Invariant Interest Point Detection

Leveraging the analysis from the previous section, we propose modifications to SIFT and SURF scale-space detectors that are invariant to defocus blur.

1) DBI-SIFT blob detector: We propose to replace (3) with
\[ \tilde{I}(x) \ast \nabla^2 G(x, \nu) \ast \nabla^2 G(x, \sigma) = \tilde{I}(x) \ast \nabla^4 G(x, \zeta), \tag{34} \]
where we invoke the DOG approximation:
\[ \nabla^2 G(x, \nu) \approx G(x, \sqrt{2} \nu) - G(x, \nu) \]
\[ \nabla^2 G(x, \sigma) \approx G(x, \sqrt{2} \sigma) - G(x, \sigma). \tag{35} \]
Combining, we have,

$$\nabla^4 G(x, \zeta) \approx \{G(x, \sqrt{2}v) - G(x, \nu)\} \ast \{G(x, \sqrt{2}\sigma) - G(x, \sigma)\}, \quad \text{where} \quad \zeta = \sqrt{\nu^2 + \sigma^2}$$

as before. Hence DBI-SIFT interest point is the local extrema of:

$$\tilde{I}(x) \ast \nabla^4 G(x, \zeta) \approx \{B(\cdot) \ast \nabla^2 G(\cdot, \nu)\} \ast \{I(\cdot) \ast \nabla^2 G(\cdot, \sigma)\}(x) + \{Z(\cdot) \ast \nabla^2 G(\cdot, \sigma) \ast \nabla^2 G(\cdot, \nu)\}(x). \quad (36, 37)$$

Fig. 5. Example images from our data set, containing scenes with four variations in zoom, rotation, and viewing angles, and 7 variations in blur. The sharpest (blur=1) and blurriest (blur=7) images are taken with f/3.5 and f/22, respectively.
As before, \( Z(x) \times \nabla^2 G(x, \sigma) \) is the usual amount of noise seen in (5). The additional \( \nabla G(x, v) \) is the “penalty” paid by scheme in (37). The location \( x_0 \) is declared as DBI-SIFT interest point—that is, meets a strict sense of defocus blur-invariance—if
\[
\| \tilde{I}(x_o) \times \nabla^4 G(x_o, \zeta) \|
\geq \| \tilde{I}(x_o + \Delta x) \times \nabla^4 G(x_o + \Delta x, \zeta + \Delta \zeta) \|
\tag{38}
\]
over some spatial-scale neighborhood covered by \((\Delta x, \Delta \zeta)\).

2) DBI-SURF blob detector: We propose to replace (12) with the following determinant of the modified Hessian matrix.

\[
\tilde{H}(x, \zeta) = \begin{pmatrix}
I_{xxx}(x, \zeta) & I_{xyy}(x, \zeta) \\
I_{xxx}(x, \zeta) & I_{yyy}(x, \zeta)
\end{pmatrix}
= \begin{pmatrix}
B_{xx}(x, v) * I_{xx}(x, \sigma) & B_{xy}(x, v) * I_{xy}(x, \sigma) \\
B_{yx}(x, v) * I_{xy}(x, \sigma) & B_{yy}(x, v) * I_{yy}(x, \sigma)
\end{pmatrix}
+ \begin{pmatrix}
G_{xx}(x, v) * Z_{xx}(x, \sigma) & G_{xy}(x, v) * Z_{xy}(x, \sigma) \\
G_{yx}(x, v) * Z_{xy}(x, \sigma) & G_{yy}(x, v) * Z_{yy}(x, \sigma)
\end{pmatrix},
\tag{39}
\]

where \( \zeta = \sqrt{\sigma^2 + \eta^2} \). The location \( x_0 \) is declared as DBI-SURF interest point if \( \det(\tilde{H}(x_0, \zeta)) \geq \det(\tilde{H}(x_0 + \Delta x, \zeta + \Delta \zeta)) \) over some spatial-scale neighborhood covered by \((\Delta x, \Delta \zeta)\). The design of modified Hessian matrix in (39) is motivated as a way to satisfy (58) in Appendix (but with \( p \) instead of \( q \)), for which
\[
\det(\tilde{H}(x_0 + m, \sigma)) \geq \det(\tilde{H}(x_0 + n, \sigma))
\tag{40}
\]
for all \( m, n \in \mathbb{Z}^2 \) satisfying \( \| m \|_2 < \| n \|_2 < 2p \) is a necessary condition. We expect the local maximum of \( \tilde{H}(x) \) to coincide with the local maximum of \( \det(\tilde{H}(x)) \)—meeting a strict sense of defocus blur-invariance when (40) is satisfied, which is easier to fulfill than (22) because \( p \ll q \). Although the fourth derivative filter gives small magnitude of the response, the local maxima point response has a sharp enough peak to be detected reliably over some spatial-scale neighborhood covered by \((\Delta x, \Delta \zeta)\).

C. Defocus Blur-Invariant Descriptor Matching Criteria

Let us begin by reviewing a classical descriptor matching criteria. Suppose \( \{x_1, \ldots, x_M\} \) and \( \{y_1, \ldots, y_N\} \) are pixel coordinates of interest points in reference and target images with detected scales \( \{\eta_1, \ldots, \eta_M\} \) and \( \{\sigma_1, \ldots, \sigma_N\} \), respectively. Let \( D_r(x_m, \eta_m) \) and \( D_t(y_n, \sigma_n) \) be the descriptors of \( x_m \) and \( y_n \), respectively—they may refer to SIFT \((U(x, \sigma))\) or SURF \((V(x, \sigma))\) descriptors but it makes no difference to the presentation below. In a criteria known as nearest-neighborhood (NN) search, a target interest point \( y_n \) is matched to the reference interest point \( x_m \) by [17]:
\[
\tilde{m} = \arg \min_n \| D_r(x_m, \eta_m) - D_t(y_n, \sigma_n) \|_2.
\tag{41}
\]
To safeguard against false matches, two additional conditions must be met before a match is declared valid. The first condition is the fidelity of the match based on the descriptor error:
\[
\| D_r(x_m, \eta_m) - D_t(y_n, \sigma_n) \|_2 < \tau
\tag{42}
\]
for some threshold \( \tau \). The second condition is known as nearest neighborhood distance ratio (NNR) [3], [17]:
\[
\frac{\| D_r(x_m, \eta_m) - D_t(y_n, \sigma_n) \|_2}{\| D_r(x_m, \eta_m) - D_t(y_{\tilde{n}}, \sigma_{\tilde{n}}) \|_2} < t, \forall \tilde{n} \neq \tilde{m}.
\tag{43}
\]

The intuition behind the NNR is that descriptor error \( \| D_r(x_m, \eta_m) - D_t(y_n, \sigma_n) \|_2 \) must be sufficiently smaller than the error of a non-match \( \| D_r(x_m, \eta_m) - D_t(y_{\tilde{n}}, \sigma_{\tilde{n}}) \|_2 \). Hence, varying \( t \in [0, 1] \) gives us a sense of distinctiveness or discriminativity of the descriptors. In [3], the optimum NNR threshold value \( t \) was experimentally determined to be 0.80 for sharp images.

Consider now the scenario where the target image is blurred. Define \( \tilde{D}_r(\cdot, \cdot) \) as the SIFT/SURF descriptor computed from a blurred target image instead of the sharp one. Recalling that SIFT and SURF descriptors are invariant to defocus blur if scale \( \sigma \) is known, there exists \( \sigma \) such that the following approximation holds:
\[
\tilde{D}_t(y_n, \sigma) \approx D_t(y_n, \sigma_n),
\tag{44}
\]
that is, a descriptor \( \tilde{D}_t \) computed with a scale other than the detected scale \( \zeta \). In practice, the detected blur size \( \zeta = \sqrt{\sigma^2 + \eta^2} \) is larger than the \( \sigma \) we desire. To handle image correspondences based on SIFT/SURF descriptors with unknown scale \( 0 < \sigma < \zeta \), we adopt a multi-scale approach similar to that of [52] that interprets descriptor matching criteria as a function of \( \tilde{\sigma} \). We modify NN search as a minimization of the form
\[
\tilde{m} = \arg \min_n \left\{ \min_{\sigma} \| D_r(x_m, \eta_m) - D_t(y_n, \sigma) \|_2 \right\}.
\tag{45}
\]
That is, we draw on the distinctiveness of the descriptors to claim that \( D_t(y_n, \tilde{\sigma}) \) with the minimum descriptor error suffices as a proxy for \( D_t(y_n, \sigma_{\tilde{n}}) \). A one-dimensional search over \( 0 < \tilde{\sigma} < \zeta \) can be executed in a reasonable time.

When both the reference and the target images are blurred, there also exists a scale \( \tilde{\eta} \) such that
\[
\tilde{D}_r(x_m, \tilde{\eta}) \approx D_r(x_m, \eta_{\tilde{m}}),
\tag{46}
\]
where the descriptor $\hat{D}_r$ is computed from a blurry reference image. Drawing on the distinctiveness of the descriptors again, the following is proxy for (41):

$$\hat{m} = \arg \min_n \left\{ \min_{\tilde{t}, \tilde{\eta}} \| \hat{D}_r(x_n, \tilde{\eta}) - \hat{D}_t(y_n, \tilde{\eta}) \|_2 \right\}. \quad (47)$$

Although the blurred reference/blurred target matching is computationally expensive owing to the increased search dimension, the matches made in this manner are accurate, as our evaluation in Section IV verifies. The proposed defocus blur-invariant descriptor matching criteria of (45) and (47) are used in conjunction with thresholds in (42) and (43).

We claim that the distinctiveness of the interest point description deteriorated by the blur is restored by the proposed defocus blur-invariant descriptor matching criteria of (45) and (47). This is clearly evidenced by Figure 6. Specifically, the ability for the descriptor matching criteria to differentiate one population (positive match) from another (false match) depends on whether the probability density functions (pdf) of positive and false matches admitted by the NNR thresholding in (43) have significant overlap. As clearly shown by Figure 6, the proposed multi-scale descriptor matching criteria enjoys very little risk of false rejection when thresholding by a $t$ value (to eliminate false matches), while ordinary NN descriptor matching criteria in (41) for SIFT/SURF clearly suffers from false rejection/matching. Based on this analysis, the thresholding value of $t = 0.80$ as empirically determined in [3] remains to be an appropriate choice for blurred scale-space feature matching.

IV. EXPERIMENTAL PERFORMANCE EVALUATION

A. Data Set

The proposed algorithms are tested on 168 real images taken by Nikon D9 with Nikkor 18-105mm f/3.5-5.6G ED VR lens of scenes containing different structures, texture, and depth. Each image was captured in a “raw sensor” mode, which was treated with demosaicking method of [53], color correction, and white balance [54] before the feature extraction. Gamma correction was applied after feature extraction owing to the fact that the blur model of (17) will not hold after nonlinear tone mapping. Example images are shown in Figure 5. (Upon acceptance of this paper, we will make this dataset available on our research website.)

Similar to [4], [17], each scene is photographed 4 times with zoom, rotation, and/or view angles of varying degrees. For each zoom/rotation/angle setting, we captured 7 images while varying the aperture sizes so that we obtain sharp and blurry images—we refer to the grouping of these 7 images as “aperture stack.” We also claim this procedure results in variation in lighting and noise, since large f-number images are very dark. In fact, the light efficiency of f/3.5 (minimum Nikkor setting) is 39.51 = $(\frac{22}{35})^2$ times more than the f/22 (maximum Nikkor setting)—for display, images in Figure 5 are shown with the lighting variation normalized, however. We also point out that the Nikkor lens used in our experiments employs 7-fold aperture (yielding heptagonal blur kernel shape). Even though the proposed DBI-SIFT/DBI-SURF methods did not compensate for the minor differences between the circular and 7-fold apertures, our experimental results below yielded convincing results. See Figure 26 in the supplementary submission for experiments verifying that the proposed DBI-SIFT/DBI-SURF techniques remain effective for N-fold apertures (where a $N$ ranges from 3 to 10). There were no pixels that are saturated other than the sky region. The scenes captured with viewing angle difference are of a planar scene, since no “global” homography transformation exists for scenes of discontinuous depth. However, image with zoom/rotation changes were of scenes of varying depth. As result, each blurred image contains multiple defocus blur sizes that depend on the (unknown) scene depth.

A homography matrix $H$ describes the coordinate transforms between different zoom/rotation/view angle images. In our dataset, homography matrices were computed by following a procedure described in [4], [17] using the f/22 images, which are the sharpest images taken within an aperture stack. For a fixed zoom/rotation/angle setting, blurred images within the same aperture stack are assumed to have a homography identical to the f/22 image.

B. Interest Point Detection Repeatability

Pixel coordinates $x$ and $y$ of the detected interest points in reference and target images, respectively, are said to be repeated or corresponding if for some threshold $\varepsilon > 0$,

$$\|x - Hy\|_2 < \varepsilon,$$  

where $H$ is the homography matrix between the two images. If the detected interest point features are indeed invariant to zoom/rotation/angle changes, then there should be a high probability of repeated interest points. If on the other hand the blur negatively influences interest point detection, the number of repeated interest points should decrease as a result. Indeed, Figures 18 and 19 in the supplementary material suggests that the detected SIFT/SURF interest points are randomly scattered when the blur is severe. Contrast this to the DBI-SIFT/DBI-SURF interest points that are stable across blur.
JOURNAL OF [TEX CLASS FILES, VOL. 13, NO. 9, SEPTEMBER 2014 10

Fig. 8. Repeatability of interest point detection as a function of blur severity (the reference image is sharp). (a-f) We used images in Figures 5(a,i,u) as sharp reference images. (a-c) The target images are the blurred versions (e.g. Figures 5(b,j,v)) of the reference images. (d-f) The target images are the zoomed/rotated/angled versions (Figures 5(c,k,w)) of the reference images (Figure 5(a,i,u)) with varying degrees of blur (e.g. Figures 5(d,l,x)). For DBI-SIFT and DBI-SURF only, the reference interest points were detected using ordinary SIFT and SURF methods (i.e. local extrema of DOG or Hessian determinant response), respectively (recall goal of DBI-SIFT and DBI-SURF is to recover these points from blurred images). We conclude that DBI-SIFT, DBI-SURF, and Harris-Laplace detectors are largely invariant to defocus blur. More complete results are provided in supplementary submission.

Figure 8 provides a quantitative analysis of the influence blur has on the interest point detection. Repeatability is defined as the percentage of detected target interest points that correspond with the interest points of the reference image [4]:

\[
\text{repeatability} = \frac{\# \text{correspondences}}{\# \text{detection}}. \tag{49}
\]

Figure 8 shows the repeatability result when the reference image is sharp and target image varies in sharpness/blurriness. If the target images are also sharp, the proposed DBI-SIFT/DBI-SURF as well as conventional SIFT [3], conventional SURF [5], BRISK [6], Harris-Laplace [4], STAR [11], ORB [16], FAST [12], and MSER [7] enjoy relative invariance to zoom, rotation, and view angle, as evidenced by the high repeatability score for “blur=1” in each plot of Figure 8. However, SIFT, SURF, BRISK, STAR, ORB, FAST, and MSER were sensitive to blur, as indicated by repeatability score degrading with increased aperture size. By comparison, DBI-SIFT and DBI-SURF demonstrated considerable robustness against blur. In fact, the repeatability score of DBI-SIFT/DBI-SURF improved with a small amount of blur, since they were not designed to work with sharp images. The Harris-Laplace [4] and CPC [37] corner detectors also showed robustness to blur. Though the original intention of the Laplace operator in [4] was for detecting corners through different scales, our analysis in Section III leads us to suspect that higher order derivative helped increased its robustness against defocus blur. On the other hand, CPC was designed to be invariant to blur—indeed Figures 8(a-c) confirm this. However, CPC is sensitive to zoom, rotation, and view angle, as evidenced by poor performance in Figures 8(d-f).

C. Recall and Precision

Interest point descriptors are useful only if it can be used to match a large percentage of the corresponding interest points and if the false match rates are low. A match between reference interest point \( x_m \) and target interest point \( y_{\tilde{m}} \) established by (42) and (43) is “correct” if they are also repeated (i.e. satisfying (48)). The recall describes the ability for descriptors to correctly match the corresponding interest points by the ratio:

\[
\text{recall} = \frac{\# \text{correct match}}{\# \text{correspondences}}. \tag{50}
\]

Since recall does not penalize against false matches made by the matching algorithm, a separate quantity to assess precision is needed. In [17], precision score is a percentage of interest points satisfying (42) and (43) that are correct (i.e. satisfy (48)):

\[
\text{precision} = \frac{\# \text{correct match}}{\# \text{total match}}. \tag{51}
\]

If the value of \( t \in [0, 1] \) in (43) increased, the number of correctly and incorrectly matched interest points increase.
Fig. 9. Precision of descriptor matching as a function of target image blur severity (the reference image is sharp). (a-f) Reference descriptors were computed on the interest points detected in Figures 5(a,i,u) by SIFT. (a-c) The locations of the interest points were the same as the reference descriptors, but the actual target descriptor vectors were computed from the blurred images (e.g. Figures 5(b,j,v)). (d-f) The locations of the interest points were determined by SIFT detector applied to Figures 5(c,k,w) (i.e. sharp target images), but the actual descriptor vectors are computed from the blurred images (e.g. Figures 5(d,l,x)). For DBI-SIFT and DBI-SURF only, the proposed matching criteria in (45) combined with NNR in (43) were used; the remainder of plots were generated by a combination of NN matching criteria in (41) combined with NNR in (43). We conclude that DBI-SIFT and DBI-SURF are largely invariant to defocus blur. (g-l) Recalling that BRISK, FREAK, ORB, and BRIEF descriptors are designed to describe corners, the dotted lines indicate the experiments in (a-f) repeated using Harris-Laplace corner detector to determine the location of reference and target interest points (i.e. instead of blob detectors).
As such, the recall is monotonically increasing function of $t \in [0, 1]$ (though it may never reach 100%); whether precision scores increases with $t \in [0, 1]$ depends on whether descriptors are discriminative. As such, precision-recall curves as a parametric function of $t \in [0, 1]$—such as the one shown in Figure 7—are popular ways to characterize the discriminative trends of the descriptors.

To understand the impact of defocus blur on the descriptors, we computed the precision-recall curve for varying severities of blur. In Figure 9, we summarize the overall precision-recall trend by plotting the precision at a fixed recall rate (a vertical line in Figure 7) as a function of blur. The NN matching criteria of (41) together with the NNR thresholding in (43) shown in Figure 9 reveals that interest point descriptors of SIFT [3], SURF [5], LIOP [9], BRISK [6], FREAK [8], ORB [16], and BRIEF [20], [21] are largely invariant to deformations such as zoom, rotation, and view angle changes (to varying degrees) when defocus blur is negligible. It is also clear from these plots that blur is detrimental to the precision of descriptor based on NN-search criteria, however. The downward trend as blur index goes up (from blur=1 to blur=7) indicates that defocus blur negatively impacts matching. One unusual trend in Figure 9 is a local minimum at blur=2—this is likely caused by the fact that defocus blur affects recall and precision differently.

Since blur decreases the number of correct matches, recall in (50) decreases steadily as blur increase. On the other hand, blur influences precision in a non-trivial way: when the blur is not severe, the number of false matches in (51) is relatively high. On the other hand, as blur becomes severe, fewer matches are declared. Thus precision depends on the delicate balance between the rates at which correct and total numbers of matches decay.

By contrast, the multi-scale DBI-SIFT/DBI-SURF matching criteria proposed in (45) together with the NNR thresholding in (43) successfully matches the SIFT/SURF descriptors of the sharp reference image $D_r(x, \eta)$ with the SIFT/SURF descriptors of the blurry target image $\tilde{D}_t(y, \tilde{\sigma})$. The rate at which the precision of the multi-scale matching criteria deteriorates is far slower than the conventional matching criteria of (41). We may conclude that DBI-SIFT/DBI-SURF retain discriminativity more effectively under the influence of severe defocus blur.

We also repeated this experiment for matching descriptors computed from a blurred reference image $\tilde{D}_r(x, \tilde{\eta})$ to descriptors computed from a blurred target image $\tilde{D}_t(y, \tilde{\sigma})$. SIFT and SURF descriptors were correctly matched between the blurred images using the proposed criteria in (47) in conjunction with NNR thresholding in (43), even when the same descriptors failed to yield meaningful matches with the conventional NN matching criteria in (41). The precision-recall of DBI-SIFT/DBI-SURF also outperformed matching using descriptors of LIOP, BRISK, FREAK, ORB, and BRIEF in the presence of blur.

The remainder of the experimental results in the supplemental material submission follow the same format as Figures 8-10, where the detected interest points and repeatability/precision-recall are shown for reference and target images of varying degrees. Our study is conducted over
The proposed DBI-SIFT/DBI-SURF blob detections in (38) and (39) make use of higher order derivatives. To understand the impact of noise on the fourth order derivatives, we performed an additional experiment. An indoor low-light scene was photographed by defocused Nikon D90 at ISO200 multiple times—once with small aperture (f/29) and shutter speed of 25 seconds to obtain a sharp, low-noise reference image; and 23 more times with large aperture (f/4.2) with varying shutter speed (from 4 seconds to 1/40 seconds) to obtain blurry and noisy images. Recalling that signal-to-noise in power scales linearly with the light intensity, 1/40 second integration time is unreasonably short and noisy for such a scene. The repeatability performances as a function of integration time is reported in Figure 11. It demonstrates that DBI-SIFT blob detection is surprisingly resilient to noise, even with excessively short integration time. DBI-SURF blob detection was more sensitive to noise—the repeatability of conventional SURF was higher than the proposed DBI-SURF when the integration time was less than 1/4 seconds (though the repeatability of the latter is reasonable). By contrast, Harris-Laplace corner detection was sensitive to noise, decaying rapidly as shutter speed increased. We conclude that DBI-SIFT/DBI-SURF blob detections have acceptable performance for reasonable indoor low-light operating conditions.

Table I shows the computation time needed for matching. A Lenovo laptop (Intel core i7, 2.4 GHz processor, 8GB RAM) running Matlab R2013b was been used to record execution for matching between the descriptors extracted from reference sharp image and blurry target images. Though our Matlab implementations are unoptimized for speed, codes for executing SIFT and SURF are near-identical to DBI-SIFT and DBI-SURF. SURF/DBI-SURF are implemented using Gaussian derivative filters instead of their Haar-type approximations, since the near-defocus blur-invariance of SIFT/SURF descriptors in Section II-E depends on the approximation in (27) hold better.

It is clear from Table I, the time needed for matching for proposed descriptors is greater than the time needed for comparable SIFT and SURF implementations. However, the number of correct matches for the proposed descriptors is significantly more than the number of correct matches of ordinary SIFT or SURF matching. We plan to investigate techniques for improving the execution speed in the future. Strategies that will be considered will include Haar-type approximations of Gaussian fourth derivative [5] and faster searches for matches inspired by KD-trees [56], FLANN [57], and IVFADC [58]. Since our multi-scale descriptor matching criteria is similar to that of [52], we also expect speed improvements by leveraging the optimization strategies taken in [52].
TABLE I
EXECUTION AVERAGE TIME FOR MATCHING IMAGE SHARP IMAGE AND TRANSFORMED BLURRED TARGET IMAGES (21 IMAGES).

<table>
<thead>
<tr>
<th></th>
<th>average</th>
<th>SIFT</th>
<th>DBI-SIFT</th>
<th>SURF</th>
<th>DBI-SURF</th>
</tr>
</thead>
<tbody>
<tr>
<td># points in ref. image</td>
<td>4700</td>
<td>4700</td>
<td>4700</td>
<td>4700</td>
<td></td>
</tr>
<tr>
<td># points in target image</td>
<td>4355</td>
<td>4355</td>
<td>4355</td>
<td>4355</td>
<td></td>
</tr>
<tr>
<td># of correct match</td>
<td>1247</td>
<td>1928</td>
<td>998</td>
<td>1558</td>
<td></td>
</tr>
<tr>
<td>execution time</td>
<td>23s</td>
<td>80s</td>
<td>17s</td>
<td>63s</td>
<td></td>
</tr>
</tbody>
</table>

VI. CONCLUSION

We proposed DBI-SIFT and DBI-SURF detectors that use Gaussian fourth derivative filters on blurred blobs, restoring the SIFT and SURF responses that Gaussian second derivative filters had on a sharp blob. The proposed DBI-SIFT and DBI-SURF multi-scale descriptor matching criteria overcomes the influence of defocus blur by searching over the unobserved blob scale that makes the blurred interest point descriptor similar to the sharp descriptor. Experimental results from real images confirmed that repeatability and precision of the proposed feature extraction techniques were far more robust to defocus blur than state-of-art interest point detectors and descriptors—even while retaining invariance to zoom/rotation/angle/lighting. We also verified the invariance to blur under the scenario when both the reference and the target images are blurred. Drawback to our method is the increased complexity of matching with DBI-SIFT/DBI-SURF descriptors. Although the use of fourth derivative in the DBI-SIFT/DBI-SURF blob detection increased sensitivity to noise in theory, we experimentally verified that repeatability of the detection is acceptable for practical low-light camera setting. Our future research direction may include extending scale-space feature extraction techniques using other known blur-invariant operators [33], [34].

APPENDIX

Before we justify (21) and (22), we emphasize that we are not interested in solving these statements in the most general case. Instead, we are interested in characterizing the behavior of (21) and (22) near the blob centers only—as our intended goal is to describe whether blob detections would be invariant to defocus blurs. As we are also interested in the corner cases (when blob detections from sharp and defocus blurred images coincide due to unreasonable/obscure combinations of image statistics), it is more useful to describe the defocus blur invariance in terms of “sufficient conditions” stemming from likely scenarios rather than “necessary conditions” (that must not exclude corner cases). In other words, they carry the weight of “necessary conditions” for all pragmatic scenarios.

Suppose \( x_0 \) is a local maximum of \( \nabla^2 I(x, \sigma) := I(x) \ast \nabla^2 G(x, \sigma) \) (i.e. SIFT interest point of a sharp image). The pixel \( x_0 \) is also a local maximum of \( B \ast \nabla^2 I \) (i.e. SIFT interest point of a blurred image) if

\[
\left\{ B(\cdot) \ast \nabla^2 I(\cdot, \sigma) \right\}(x_0) \geq \left\{ B(\cdot) \ast \nabla^2 I(\cdot, \sigma + \Delta \sigma) \right\}(x_0 + \Delta x)
\]

(52)

for all scale-space neighborhood \((\Delta x, \Delta \sigma)\) centered around \((x, \sigma)\). Consider the specific case of \(\Delta \sigma = 0\). By algebra, we obtain

\[
0 \leq \{\delta(x_0) - \delta(x_0 - \Delta x)\} \ast \{B(x_0) \ast \nabla^2 I(x_0, \sigma)\} = \{B(x_0) - B(x_0 - \Delta x)\} \ast \nabla^2 I(x_0, \sigma)
\]

(53)

\[
= \int_{\|x - x_0\|_2 < q} \int_{\|x - x_0 - \Delta x\|_2 < q} \frac{\nabla^2 I(x, \sigma)}{\pi q^2} \, dx.
\]

Clearly, the region \( \{x : \|x - x_0\|_2 < q, \|x - x_0 - \Delta x\|_2 < q\} \) overlapped by the two integrals cancels each other. Thus, the quantity in (53) is positive if every member of the set \( \{\nabla^2 I(x, \sigma) : \|x - x_0\|_2 < q, \|x - x_0 - \Delta x\|_2 > q\} \) is larger than the members of \( \{\nabla^2 I(x, \sigma) : \|x - x_0\|_2 > q, \|x - x_0 - \Delta x\|_2 < q\} \) for all \( \|\Delta x\| < q \) a condition met by (21).

To justify the statement (22), we make an additional assumption that \( I_{xy}(x) \) is insignificantly small near \( x \approx x_0 \). This is justifiable since the blob center does not resemble saddle point (and \( I_{xy}(x) \) is exactly zero if the blob shape is circular or oval). Suppose \( x_0 \) is a local maximum of \( \text{det}(\mathcal{H}(x)) \approx I_{xx}(x)I_{yy}(x) \) (i.e. SURF interest point of a sharp image). The pixel \( x_0 \) is also a local maximum of \( \text{det}(\mathcal{H}(x)) \approx I_{xx}(x)I_{yy}(x) \) (i.e. SURF interest point of a blurred image) if

\[
\left\{ B \ast I_{xx}(\cdot, \sigma) \right\}(x) \left\{ B \ast I_{yy}(\cdot, \sigma) \right\}(x) = \left\{ B \ast I_{xx}(\cdot, \sigma + \Delta x) \right\}(x) \left\{ B \ast I_{yy}(\cdot, \sigma) \right\}(x) \geq 0
\]

(54)

\[
\left\{ B \ast I_{xx}(\cdot, \sigma) \right\}(x) \left\{ B \ast I_{yy}(\cdot, \sigma) \right\}(x) \geq 0
\]

(55)

By algebra in a manner similar to (53), we have

\[
0 \leq \int_{\|x - x_0\|_2 < q} \int_{\|y - x_0\|_2 < q} I_{xx}(x, \sigma)I_{yy}(y, \sigma) \, dy \, dx
\]

(56)

\[
- \int_{\|x - x_0 - \Delta x\|_2 < q} \int_{\|y - x_0\|_2 < q} I_{xx}(x, \sigma)I_{yy}(y, \sigma) \, dy \, dx.
\]

(57)

Canceling out the overlapping regions of the two integrals,

\[
0 \leq \int_{\|x - x_0\|_2 < q} \int_{\|y - x_0\|_2 < q} I_{xx}(x, \sigma)I_{yy}(y, \sigma) \, dy \, dx
\]

(58)

\[
- \int_{\|x - x_0 - \Delta x\|_2 < q} \int_{\|y - x_0\|_2 < q} I_{xx}(x, \sigma)I_{yy}(y, \sigma) \, dy \, dx
\]

(59)

\[
- \int_{\|x - x_0\|_2 < q} \int_{\|y - x_0\|_2 < q} I_{xx}(x, \sigma)I_{yy}(y, \sigma) \, dy \, dx
\]

(60)

\[
- \int_{\|x - x_0 - \Delta x\|_2 < q} \int_{\|y - x_0\|_2 < q} I_{xx}(x, \sigma)I_{yy}(y, \sigma) \, dy \, dx.
\]
Naturally (i.e. excluding corner cases), inequality above decomposes into three parts:

\[
0 \leq \int_{\|x-x_0\| < q} I_{xx}(x, \sigma) \, dx \times \left( \int_{\|y-y_0\| < q} \left( I_{yy}(y, \sigma) \, dy - \int_{\|y-y_0\| > q} I_{yy}(y, \sigma) \, dy \right) \right)
\]

\[
0 \leq \int_{\|x-x_0\| < q} I_{xx}(x, \sigma) \, dx \times \left( \int_{\|y-y_0\| < q} \left( I_{yy}(y, \sigma) \, dy - \int_{\|y-y_0\| > q} I_{yy}(y, \sigma) \, dy \right) \right)
\]

\[0 \leq \int_{\|x-x_0\| < q} I_{xx}(x, \sigma) \, dx \times \left( \int_{\|y-y_0\| < q} \left( I_{yy}(y, \sigma) \, dy - \int_{\|y-y_0\| > q} I_{yy}(y, \sigma) \, dy \right) \right)
\]

(58)

Recall that near the blob centers, we expect \(\text{sgn}(I_{xx}) = \text{sgn}(I_{yy})\). Thus inequalities in (58) are fulfilled if every member of the set \(\{ I_{xx}(x, \sigma) : \|x-x_0\| < q, \|x-x_0-\Delta x\|_2 < q \}\) has a larger magnitude than the members of \(\{ I_{yy}(y, \sigma) : \|y-y_0\| < q, \|y-y_0-\Delta y\|_2 < q \}\) for all \(\|\Delta x\| < q\); and every member of \(\{ I_{yy}(x, \sigma) : \|x-x_0\| < q, \|x-x_0-\Delta x\|_2 < q \}\) has a larger magnitude than the members of \(\{ I_{xx}(y, \sigma) : \|y-y_0\| < q, \|y-y_0-\Delta y\|_2 < q \}\), which for (22) is a necessary condition.

To justify (27), take a Taylor expansion of \(b(\omega)\) near frequency center in (26):

\[b(\omega) = b\left(\frac{1}{\sigma}\right) + (\omega - \frac{1}{\sigma})b'\left(\frac{1}{\sigma}\right) + (\omega - \frac{1}{\sigma})^2b''\left(\frac{1}{\sigma}\right) + \ldots \]  

(59)

Multiplying this with \(\omega g(\omega, \sigma)\), we have

\[b(\omega)\omega g(\omega, \sigma) = b\left(\frac{1}{\sigma}\right)\omega g(\omega, \sigma) + (\omega - \frac{1}{\sigma})b'\left(\frac{1}{\sigma}\right)\omega g(\omega, \sigma) + \omega - \frac{1}{\sigma})^2b''(\frac{1}{\sigma})\omega g(\omega, \sigma) + \ldots \]  

(60)

The zeroth order Taylor expansion obviously leads to first order approximation error in (27). However, this is acceptable because the derivative terms contribute very little. Specifically, if \(\omega \approx \frac{1}{\sigma}\) then \((\omega - \frac{1}{\sigma})^n\) vanishes; but \(\omega g(\omega, \sigma)\) decays rapidly as \((\omega - \frac{1}{\sigma})}\) becomes large. Thus the approximation error in (27) is significant only if derivative term \(b'\left(\frac{1}{\sigma}\right)\) is of order \((\omega - \frac{1}{\sigma})\omega g(\omega, \sigma)\) (this is a large term). This also confirms that the convergence rate of asymptotic invariance in (27) is exponential (recalling (25), the error decays at the rate \(e^{-\frac{\omega^2}{2\sigma^2}}\) with respect to \(\sigma\)).

ACKNOWLEDGMENT

We would like to thank the authors of [3]–[5], [9], [33], [34], [59], [60] for providing the codes for comparisons. This work is funded in part by Libyan Ministry of High Education.

REFERENCES
