DEFOCUS BLUR-INVARIANT SCALE-SPACE FEATURE EXTRACTIONS

Dissertation

Submitted to

The School of Engineering of the

UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for

The Degree

Doctor of Philosophy in Electrical Engineering

By

Elhusain Salem Saad

UNIVERSITY OF DAYTON

Dayton, Ohio

December 2014

DEFOCUS BLUR-INVARIANT SCALE-SPACE FEATURE EXTRACTIONS

Name: Saad, Elhusain Salem

APPROVED BY:

Keigo Hirakawa, Ph.D.
Advisor Committee Chairman
School of Engineering, Department of
Electrical and Computer Engineering

John Loomis, Ph.D.
Committee Member
School of Engineering, Department of
Electrical and Computer Engineering

Vijayan K. Asari, Ph.D.
Committee Member
School of Engineering, Department of
Electrical and Computer Engineering

Muhammad Usman, Ph.D.
Committee Member
College of Arts and Sciences,
Department of Mathematics

John G. Weber, Ph.D.
Associate Dean
School of Engineering

Eddy M. Rojas, Ph.D., M.A., P.E.
Dean
School of Engineering

ABSTRACT

DEFOCUS BLUR-INVARIANT SCALE-SPACE FEATURE EXTRACTIONS

Name: Saad, Elhusain Salem
University of Dayton

Advisor: Dr. Keigo Hirakawa

In this dissertation, we propose a modification 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 since SIFT and SURF degraded with defocus blur more than other transformation. Specifically, scale-space blob detection relies on the second derivative responses to the image. Our analysis of defocus blur 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 orientations/descriptors will make an optimal for SIFT and SURF descriptors defocus blu- invariant. The proposed defocus blur-invariant (DBI) scale-space feature extraction —which we refer to as *DBI-SIFT and DBI-SURF*—do not require image deblurring nor blur kernel estimation, meaning that its accuracy does not depend on the quality of image deblurring. We offer empirical evidence of blur invariance by establishing feature interest point correspondences between sharp or blurred reference image and a blurred target image.

To my mother Amena, my wife Aisha and my children Salem and Muhammad who supported me

to complete this work

ACKNOWLEDGMENTS

First, I would like to thank my supervisor Professor Keigo Hirakawa for giving me the opportunity to work with him. I got a lot of skills from him, and I would like to keep working with him.

I would like to thank also my committee members, Professor John Loomis, Professor Asari, Vijayan, and Professor Muhammad Usman for their valuable review of this dissertation.

I would like to thank my Colleague Chen for his help during collecting the results.

Last but not the least, a special thanks to my family especially my mother for her pray for me and my wife for her constant support and encouragement throughout this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 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 uses of distinctive and robust invariant features that describe the local region by unique signature. A higher level image features describing the shape is more dependable for real scenarios [1].

Local features can be defined as an image pattern with properties that differs from the surrounding. The detector detects specific intensity patterns from an image (e.g. edges, corners, or a 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 transformation. In order that the interest points detection be useful, it must also be distinctive, have enough sensitivity to detect features in low illumination, and have accurate localization [1–4]. Each region around the interest point can be described by a descriptor, where a match based on descriptor similarities establish 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 [1–4]. There have also been a number of rigorous studies to assess the relative invariance to scale, rotation, noise, etc [3]. Indeed, to a large extent, many techniques are found to be robust

<div align="center">(a)   (b)</div>
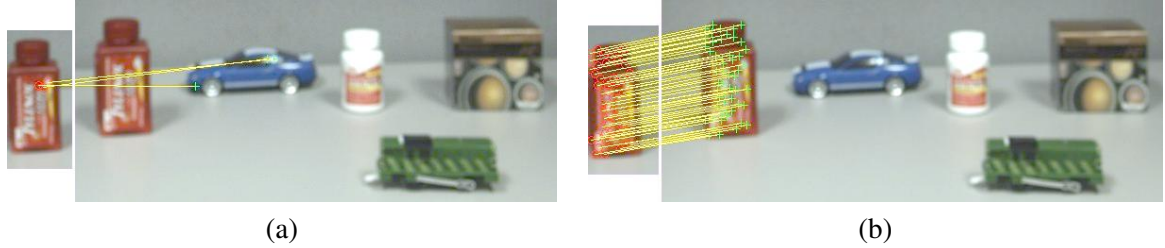
Figure 1.1: Example of matching the interest points detected over blurred reference image with the interest points detected over a blurred target image. (a) Failed detection and matching using SIFT with RANSAC. (b) Successful detection and matching using the proposed DBI-SIFT with RANSAC.

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—a group of pixels whose intensities differ from the pixel intensities of the surrounding area. The blob has shape similar to Laplacian of Gaussian (LOG). For this reason, LOG plays the role of a matched filter for blob detection in SIFT and SURF. The scale of LOG can be varied to match the blob sizes.

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 [5]. Harris-Laplace detector combines the classical corner detection scheme with LOG to make the detector invariant to scale and view angle [3]. Local Intensity Order Pattern (LIOP) is a modern descriptor based on difference of local intensities that is said to be invariant to geometric and photometric transformations [6].

As reported in [3, 4, 7, 8], the effects of blur on feature extraction are more harmful than other deformations (such as view angle, scale, rotation). Figure 1.1 supports this observation as well—attempts to match blurred objects using existing techniques clearly fail. As evidenced by Figure

(a) image 1
(b) image 2
(c) Using SIFT
(d) Using DBI-SIFT
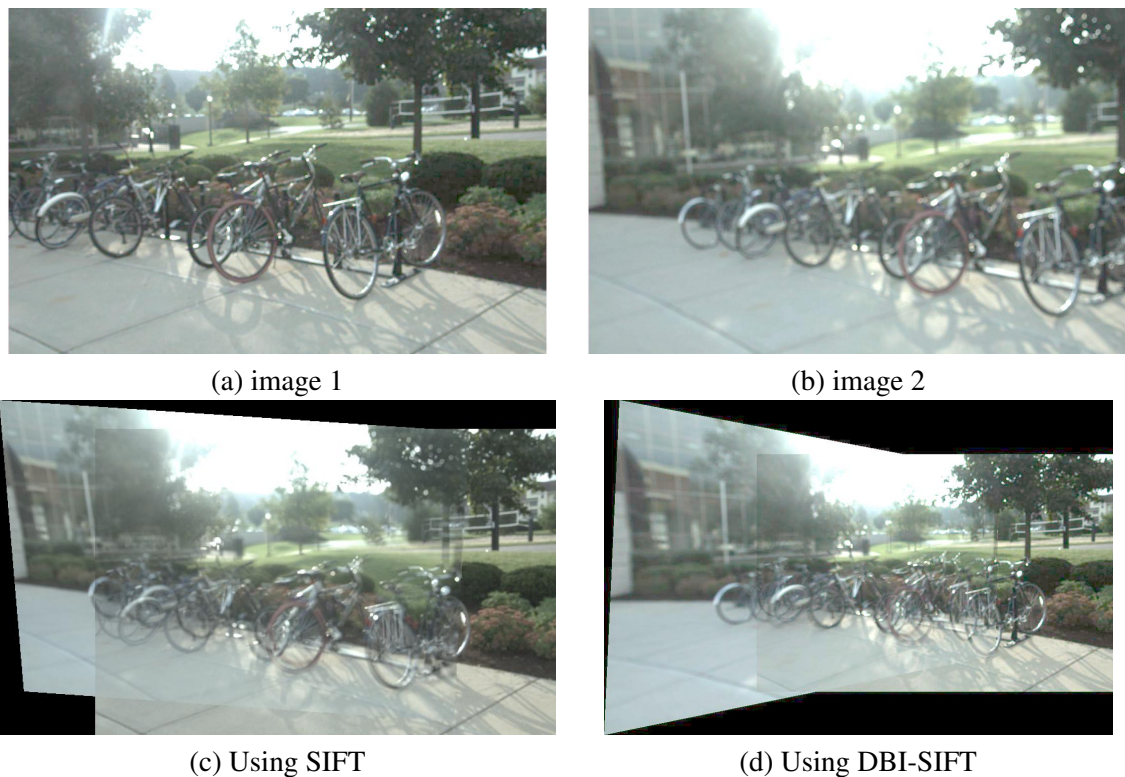
Figure 1.2: Example showing stitching blurred images 1 and 2 using SIFT and the proposed DBI-SIFT.

1.2, 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 [9, 10]. However, this is an undesirable solution because deblurring is computationally expensive and the reconstructed images are far from perfect. The other approach is to develop feature extraction techniques that are invariant to blur. Such attempt has been made in video tracking [7, 11, 12] and region entropy-based feature extraction [11, 13–15]. To date, the study of blur-invariant feature extraction has been limited, and future investigation is warranted.

CHAPTER II

STATE-OF-ART INTEREST POINT DETECTION AND DESCRIPTORS

In this chapter, we are going to discuss MSER, SIFT, SURF and LIOP descriptor.

2.1   Region Detector

Maximally Stable Extremal Regions (MSR) is a simple, reliable, and common used feature extraction. It takes a set of all possible thresholds values to convert gray scale image to a binary image as in (2.1). By connecting the pixels that either have bright extremal regions or dark extremal regions intensities to get stable regions that apear and grow with increasing the threshold by starting from a tiny seed until becomes stable. The number of thresholding depends on the stability of the evaluation of the connected components [16].

$$E_t = \begin{cases} 1 & if \ \ I(x) \geqslant t \\ 0 & otherwise \end{cases} \tag{2.1}$$

MSER region is computed by the by local minima of the rate of change of connected components area.

$$var_i = \frac{Area_{i+\Delta} - Area_i}{Area_i} \tag{2.2}$$

where i represents the maximal intensity for extremal region while $i+\Delta$ represents the ith expanded region and $Area_i = size(E_{t_i})$. When the variation is smaller than the parent and ancestor, then

4

this region is considered to be maximally stable extremal region. MSER is invariant under monotonic and homeomorphic transformation, and it is considered as stable because the extremal regions unchanged over the threshold range. It is considered as sensitive to blur as reported in [3].

2.2 SIFT interest point detection and descriptor

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

$$G(\boldsymbol{x}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x^2 + y^2)}{2\sigma^2}} \tag{2.3}$$

and the corresponding Laplacian of Gaussian (LOG):

$$\nabla^2 G(\boldsymbol{x}, \sigma) := \left\{ \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right\} G(\boldsymbol{x}, \sigma)$$
$$= G_{xx}(\boldsymbol{x}, \sigma) + G_{yy}(\boldsymbol{x}, \sigma), \tag{2.4}$$

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

$$I(\boldsymbol{x}) \star \nabla^2 G(\boldsymbol{x}, \sigma) = I_{xx}(\boldsymbol{x}, \sigma) + I_{yy}(\boldsymbol{x}, \sigma), \tag{2.5}$$

where $\star$ denotes the spatial convolution operator, and $I_{xx}(\boldsymbol{x}, \sigma) := I(\boldsymbol{x}) \star G_{xx}(\boldsymbol{x}, \sigma)$ 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 $\boldsymbol{x}_0$ when Gaussian scale σ and blob radius r have the relation $\sigma = r/\sqrt{2}$ [1]:

$$\max_{\boldsymbol{x}, \sigma} I(\boldsymbol{x}) \star \nabla^2 G(\boldsymbol{x}, \sigma) = \left\{ I(\cdot) \star \nabla^2 G\left(\cdot, \frac{r}{\sqrt{2}}\right) \right\} (\boldsymbol{x}_0). \tag{2.6}$$

Using the principle of heat equation, LOG in (2.4) can be approximated efficiently by the Difference of Gaussians (DOG) [2]:

$$\nabla^2 G(\boldsymbol{x}, \sigma) \approx \frac{G(\boldsymbol{x}, \sqrt{2}\sigma) - G(\boldsymbol{x}, \sigma)}{\sigma^2\sqrt{2} - \sigma^2}. \tag{2.7}$$

Substituting (2.7) to (2.5) results in a simplification:

$$I(\boldsymbol{x}) \star \nabla^2 G(\boldsymbol{x}, \sigma) \approx \frac{I(\boldsymbol{x}) \star G(\boldsymbol{x}, \sqrt{2}\sigma) - I(\boldsymbol{x}) \star G(\boldsymbol{x}, \sigma)}{\sigma^2\sqrt{2} - \sigma^2}. \tag{2.8}$$

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

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

$$\begin{aligned} M(\boldsymbol{x}, \sigma) &= \sqrt{I_x(\boldsymbol{x}, \sigma)^2 + I_y(\boldsymbol{x}, \sigma)^2} \\ \theta(\boldsymbol{x}, \sigma) &= \tan^{-1}\left(\frac{I_y(\boldsymbol{x}, \sigma)}{I_x(\boldsymbol{x}, \sigma)}\right), \end{aligned} \tag{2.9}$$

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

$$\boldsymbol{U}(\boldsymbol{x}, \sigma) = \left(\frac{\boldsymbol{U}_1(\boldsymbol{x}, \sigma)^T}{\|\boldsymbol{U}_1(\boldsymbol{x}, \sigma)\|}, \cdots, \frac{\boldsymbol{U}_K(\boldsymbol{x}, \sigma)^T}{\|\boldsymbol{U}_K(\boldsymbol{x}, \sigma)\|}\right)^T, \tag{2.10}$$

comprised of K sub-descriptors $\boldsymbol{U}_k(\boldsymbol{x}, \sigma)$. Each sub-descriptor is an orientation histogram weighted by $M(\boldsymbol{x}, \sigma)$ [2] computed within a series of subregions $\Lambda_k \subset \mathbb{Z}^2$ as defined in [4]:

$$\boldsymbol{U}_k(\boldsymbol{x}, \sigma) = \sum_{\boldsymbol{m} \in \Lambda_k} M(\boldsymbol{x} + \boldsymbol{m}, \sigma) \begin{pmatrix} 1_{A_1}(\theta(\boldsymbol{x} + \boldsymbol{m}, \sigma)) \\ 1_{A_2}(\theta(\boldsymbol{x} + \boldsymbol{m}, \sigma)) \\ \vdots \\ 1_{A_N}(\theta(\boldsymbol{x} + \boldsymbol{m}, \sigma)) \end{pmatrix}. \tag{2.11}$$

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

$$1_A(\theta) = \begin{cases} 0 & \text{if } \theta \notin A \\ 1 & \text{if } \theta \in A. \end{cases} \tag{2.12}$$

is an indicator function. Reassigning orientation angle θ relative to the dominant angle makes this weighted histogram rotation invariant. Normalization of this histogram in (2.10) makes it illumination invariant as well [2].

2.3   SURF interest point detection and descriptor

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

$$
\begin{aligned}
I(\boldsymbol{x} + \Delta\boldsymbol{x}, \sigma) \approx & I(\boldsymbol{x}, \sigma) + \Delta\boldsymbol{x}^T \begin{pmatrix} I_x(\boldsymbol{x}, \sigma) \\ I_y(\boldsymbol{x}, \sigma) \end{pmatrix} + \\
& \frac{1}{2}\Delta\boldsymbol{x}^T \underbrace{\begin{pmatrix} I_{xx}(\boldsymbol{x}, \sigma) & I_{xy}(\boldsymbol{x}, \sigma) \\ I_{xy}(\boldsymbol{x}, \sigma) & I_{yy}(\boldsymbol{x}, \sigma) \end{pmatrix}}_{\mathcal{H}(\boldsymbol{x}, \sigma)} \Delta\boldsymbol{x} + ...,
\end{aligned}
\tag{2.13}
$$

where the matrix $\mathcal{H}(\boldsymbol{x}, \sigma) \in \mathbb{R}^{2\times2}$ is known as *Hessian*. Suppose $\boldsymbol{x_0}$ is a local extrema. Then $\nabla I(\boldsymbol{x_0}) = 0$ because $I_x(\boldsymbol{x_0}) = I_y(\boldsymbol{x_0}) = 0$. Then the difference between the extrema $I(\boldsymbol{x_0})$ and a surrounding pixel $I(\boldsymbol{x_0} + \Delta\boldsymbol{x})$ can be approximated as:

$$
\begin{aligned}
& I(\boldsymbol{x_0} + \Delta\boldsymbol{x}, \sigma) - I(\boldsymbol{x_0}, \sigma) \approx \\
& \frac{I_{xx}(\boldsymbol{x_0}, \sigma)}{2}\Delta x^2 + I_{xy}(\boldsymbol{x_0}, \sigma)\Delta x \Delta y + \frac{I_{yy}(\boldsymbol{x_0}, \sigma)}{2}\Delta y^2,
\end{aligned}
\tag{2.14}
$$

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

$$
I_{xx}(\boldsymbol{x_0}, \sigma)I_{yy}(\boldsymbol{x_0}, \sigma) - I_{xy}(\boldsymbol{x_0}, \sigma)^2,
\tag{2.15}
$$

which is also the determinant of the Hessian matrix $\det(\mathcal{H}(\boldsymbol{x_0}, \sigma))$. Hence if $\det(\mathcal{H}(\boldsymbol{x_0}, \sigma)) > 0$, intensity pattern is assumed to be the extrema of $I(\boldsymbol{x})$ at $\boldsymbol{x} = \boldsymbol{x_0}$. If $\det(\mathcal{H}(\boldsymbol{x_0}, \sigma)) < 0$, it denotes a saddle point while $\det(\mathcal{H}(\boldsymbol{x_0}, \sigma)) = 0$ is inconclusive. The determinant and the trace of $\mathcal{H}(\boldsymbol{x}, \sigma)$ computed with different scale of LOG are used to identify the extrema of $I(\boldsymbol{x})$ and recorded as the SURF interest point. Bay used a Haar-type approximation of the Gaussian derivative to speed up the implementation [4].

SURF defines the interest point descriptor for interest point at $\boldsymbol{x}$ as vector

$$
\boldsymbol{V}(\boldsymbol{x}, \sigma) = \left(\frac{\boldsymbol{V_1}(\boldsymbol{x}, \sigma)^T}{\|\boldsymbol{V_1}(\boldsymbol{x}, \sigma)\|}, \cdots, \frac{\boldsymbol{V_K}(\boldsymbol{x}, \sigma)^T}{\|\boldsymbol{V_K}(\boldsymbol{x}, \sigma)\|} \right)^T,
\tag{2.16}
$$

where kth sub-descriptor is

$$V_k(\boldsymbol{x}, \sigma) = \sum_{\boldsymbol{m} \in \Lambda_k} \begin{pmatrix} I_x(\boldsymbol{x} + \boldsymbol{m}, \sigma) \\ I_y(\boldsymbol{x} + \boldsymbol{m}, \sigma) \\ |I_x(\boldsymbol{x} + \boldsymbol{m}, \sigma)| \\ |I_y(\boldsymbol{x} + \boldsymbol{m}, \sigma)| \end{pmatrix}. \tag{2.17}$$

Again, reassigning orientation angle θ relative to the dominant angle makes this feature vector rotation invariant. Normalization in (2.16) makes it illumination invariant as well [2].

a Local Intensity Order Pattern (LIOP) is a new (2011) robust descriptor proposed in [6]. The LIOP descriptor lacks interest point detection, and therefore relies on SIFT or SURF detector as preprocessing. LIOP does not need to get orientation angle estimation. The intensity values of the local regions are sorted in ascending order, then it is quantized into B ordinal bins according to their orders. Mathematically, let $P(x)$ an $N-$ dimensional vector of neighboring points $\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \boldsymbol{x_4}$ around the point $\boldsymbol{x}$ of local patch which is a point around the interest point $\boldsymbol{x_o}$, where x is the location of the interest point. Let γ be the mapping function that maps the vector P to permutation $\pi \in \Pi^N$ which non-descending order as follows [6]:

$$P(x) = (I(x_1), I(x_2), I(x_3), I(x_4)) \tag{2.18}$$

$$\gamma(P) = \pi, \quad P \in p^N, \quad \pi \in \Pi^N \tag{2.19}$$

Where $\pi = (i_1, i_2, ..., i_N)$ and $p_{i_1} \leqslant p_{i_2} \leqslant ...., \leqslant p_{i_N}$. Because we have $N!$ permutations in Π^N, so the mapping function γ divides the set p^N into $N!$ partitions. This partitions will be encoded using index table which contains all possible permutations, and let ϕ is feature mapping function that maps permutation to $N!$ dimensional vector which all entries zeros except the ith element which corresponding to the index table $Ind(\pi)$ [6] :

$$V_{N!}^{Ind(\gamma(P(x)))} = (0, 0, 0, ..., \underbrace{1,}_{Ind(\gamma(P(x)))} 0, .., 0) \tag{2.20}$$

8

$$\phi(\pi) = V_{N!}^{Ind(\pi)} \;\;, \pi \in \Pi^N \tag{2.21}$$

Therefore, the LIOP descriptor of a point x is:

$$LIOP(x) = \phi(\gamma(P(x))) = V_{N!}^{Ind(\gamma(P(x)))} = (0, 0, 0, ..., \underbrace{1,}_{Ind(\gamma(P(x)))} \;\; 0, .., 0) \tag{2.22}$$

The N neighboring points are distributed on a circle which has radius R and its center at x, so the first selected point will be the one that located on radial direction which connect the center of the local pattern x_o to this point x, then $N-1$ points are sampled in anticlockwise direction. This way of choosing the sampling makes the descriptor rotation invariant . The ordering of the intensity making the descriptor invariant to illumination. The final descriptor will concatenate all of these LIOP's in one vector as follows [6]:

$$LIOP_{descriptor} = (des_1, \; des_2, \; , ..., \; des_B) \tag{2.23}$$

$$des_i = \sum_{x \in bin_i} w(x) \; Liop(x) \tag{2.24}$$

$$w(x) = \sum_{i,j} sgn(|I(x_i) - I(x_j)| - T_{lp}) + 1 \tag{2.25}$$

Where $w(x)$ is a window to make the descriptor more robust to noise and T_{lp} is a preset threshold.

CHAPTER III

BLUR TYPES AND ITS IMPACT ON FEATURE EXTRACTIONS

3.1 Blur types

Image blur is caused by a pixel recording lights from multiple sources. Assuming Lambertian surfaces, blur is typically represented by the implied blur kernel that acts on the unobserved sharp in-focus image [9–11, 17]. 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 on the scene is observed by multiple moving pixels [9, 10]. Object motion causes each pixel to observe multiple points on the scene, resulting in spatially varying motion blur proportional in length to the object speed [17]. 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 the defocus blur can be useful for three dimensional scene reconstruction from a single camera [17, 18], it also interferes with feature extraction [13, 17–19].

The support of the defocus blur kernel $B(\boldsymbol{x})$ takes the shape of the aperture opening, which is a circular disk in most typical cameras. Though $B(\mathbf{x})$ may not be known exactly, consider the

following approximation [17]:

$$B(\boldsymbol{x}) \approx \begin{cases} \frac{1}{\pi q^2} & ||\boldsymbol{x}||_2 \le q, \\ 0 & \text{oterwise} \end{cases} \tag{3.1}$$

where q is the radius of the defocus blur disk. Then the blurred image $\widetilde{I}(\boldsymbol{x})$ is modeled as:

$$\widetilde{I}(\boldsymbol{x}) = B(\boldsymbol{x}) \star I(\boldsymbol{x}) + N(\boldsymbol{x}), \tag{3.2}$$

where $I(\boldsymbol{x})$ is the sharp image and $N(\boldsymbol{x})$ is a realization of noise.

3.2 Analysis of Feature Extraction Performance in Existing of Defocuse Blur

MSER, SIFT and SURF interest point detection are shown to be invariant to scale, rotation,...etc. However, in presence of blur, the detection of the interest point becomes unreliable or random . Therefore, in the next section we analyze the effects of the defocus blur on the performance of these selected algorithms. random

3.2.1 MSER

MSER depends on sharp boundaries and stable region inside the boundary. However as shown in the Figure 3.1 and equations (3.3) and (3.4), this conditions deteriorate in presence of defocus blur because the boundaries will no longer be sharp and the margin region will be influnced by the blur kernel. Recalling equations (2.1), (2.2) and (3.2)

$$\tilde{E}_{t(x)} = \begin{cases} 1 & if \tilde{I}(x) \geqslant t \\ 0 & otherwise \end{cases} \tag{3.3}$$

$$\tilde{var}_t = \frac{\tilde{Area}_{t-\Delta t} - \tilde{Area}_t}{\tilde{Area}_t}, \tag{3.4}$$

where $\tilde{I}$ is the blurred version of latent sharp image I. The region area increas and consequently the variation in equation (3.4) will not be small, so region will not be decleared as stable region.
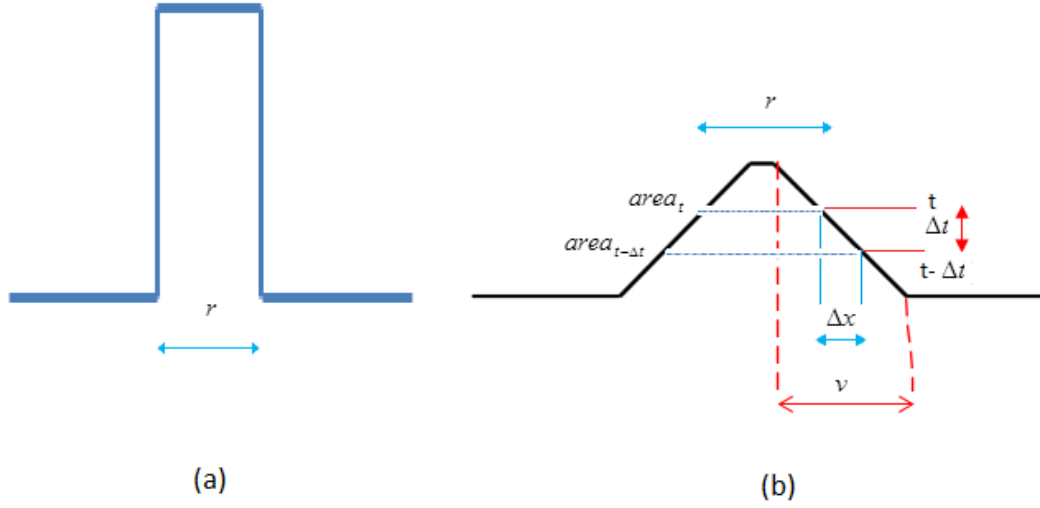
Figure 3.1: Example shows how the MSER get effected by defocus blur. (a)- sharp blob. (b)- blurry blob.

This will occur when ever the blur kernel radius has an effect on the region or blobs that their radius not big comparable with blur radius, but for those regions "blobs" that have big radius still MSER is able to detect them if the contrast did not affected. However, as the blur kernel increase, it does affect the area inside the gradual boundary because it is not stable area. Based on the Figure 3.1 and equations (3.3) and (3.4), we can say that the MSER get effected by defocus blur as follows:

$$\tilde{var}_t = \frac{\tilde{Area}_{t-\Delta t} - \tilde{Area}_t}{\tilde{Area}_t} = \frac{-\Delta t \frac{\partial}{\partial t} Area_t}{Area_t} \tag{3.5}$$

$$\tilde{var}_t = \frac{-\Delta t \frac{\partial}{\partial t} Area_t}{Area_t} = \frac{-\Delta t \int_t^\infty \int_{-\infty}^\infty \delta(\tau - I(\boldsymbol{x})) d_x d_\tau}{Area_t} = \frac{-\Delta t \int_{-\infty}^\infty \delta(t - I(\boldsymbol{x})) d_x}{Area_t}, \tag{3.6}$$

where $\Delta x = \frac{\Delta t}{\alpha}$. Let $I(\boldsymbol{x}) = \alpha x + \beta$, and $\tau = t - \alpha x - \beta$. Therefore $\partial x = \frac{\partial \tau}{\alpha}$

$$\tilde{var}_t = \frac{-\Delta t \int_{-\infty}^\infty \delta(\tau) d_\tau}{\alpha \, Area_t} = \frac{constant}{\alpha} \int_{-\infty}^\infty \delta(\tau) d_\tau \tag{3.7}$$

Basedd on (3.7) we can say that:

(a) ideal blob I (b) 2nd derivative of ideal blob   (c) $det(\mathcal{H}(\mathbf{x}, \sigma))$ of ideal blob   (d) blurred blob

(e) 2nd derivative of blurred blob   (f) $det(\mathcal{H}(\mathbf{x}, \sigma))$ of blurred blob   (g) 4th derivative of blurred blob   (h) $det(\widehat{\mathcal{H}}(\mathbf{x}, \zeta))$ of blurred blob
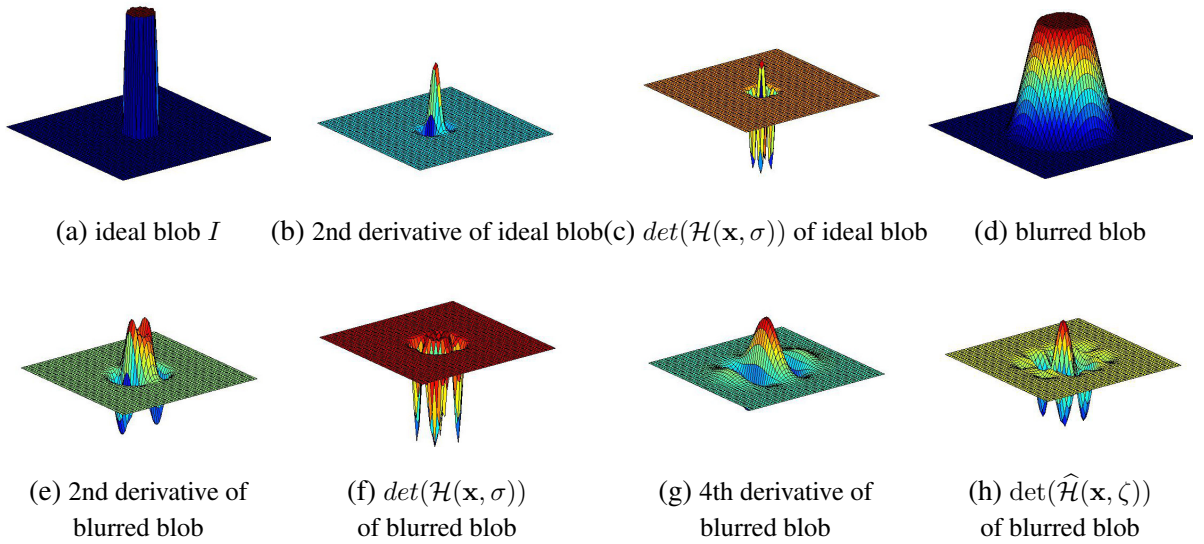
Figure 3.2: Example illustrating why interest point detection fails with defocus blur. (a-c) SURF is designed to detect ideal/sharp blob by the local maxima of Hessian determinant. (d-f) Second derivative $\widetilde{I}_{xx}$ of a blurred blob does not produce a single peak. As a result, the Hessian determinant fails to yield a local maximum at the center of the blurred blob. (g) Fourth derivative $\widetilde{I}_{xxxx}$ can reverse the effects of blur to yield a single peak. (h) Modified Hessian matrix based on the Fourth derivative response yields a single peak at the center of the blurred blob.

1. : If α is large (high contrast or small blur), then $\tilde{var}_t$ is small and consequently stable.

2. : If α is small (low contrast or large blur), then $\tilde{var}_t$ is big and consequently unstable.

3.2.2 Scale-Space Interest Point Detectors *Are Not* Defocus Blur Invariant

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

$$\frac{\partial^2}{\partial x^2} \widetilde{I}(\boldsymbol{x}) = \left\{ B(\cdot) \star \left(\frac{\partial^2}{\partial x^2} I(\cdot) \right) \right\} (\boldsymbol{x}) + \frac{\partial^2}{\partial x^2} N(\boldsymbol{x}). \tag{3.8}$$

Or by the way of Gaussian second derivative,

$$\widetilde{I}_{xx}(\boldsymbol{x}, \sigma) := \widetilde{I}(\boldsymbol{x}) \star G_{xx}(\boldsymbol{x}, \sigma)$$

$$= B(\boldsymbol{x}) \star I(\boldsymbol{x}) \star G_{xx}(\boldsymbol{x}, \sigma) + N(\boldsymbol{x}) \star G_{xx}(\boldsymbol{x}, \sigma) \tag{3.9}$$

$$= B(\boldsymbol{x}) \star I_{xx}(\boldsymbol{x}, \sigma) + N_{xx}(\boldsymbol{x}, \sigma),$$

where $N_{xx}(\boldsymbol{x}, \sigma) := N(\boldsymbol{x}) \star G_{xx}(\boldsymbol{x}, \sigma)$ (similar analysis applies to $\widetilde{I}_{xy}$ and $\widetilde{I}_{yy}$). We may conclude

that the second derivative responses $\{\widetilde{I}_{xx}, \widetilde{I}_{xy}, \widetilde{I}_{yy}\}$ of $\widetilde{I}$ are blurred (and slightly noisier) versions

of the ideal second derivatives $\{I_{xx}, I_{xy}, I_{yy}\}$.

Substituting (3.9) to (2.5) yields an interesting analysis of LOG response of the blurred image

$\widetilde{I}$:

$$\widetilde{I}(\boldsymbol{x}) \star \nabla^2 G(\boldsymbol{x}, \sigma) = \widetilde{I}(\boldsymbol{x}) \star G_{xx}(\boldsymbol{x}, \sigma) + \widetilde{I}(\boldsymbol{x}) \star G_{yy}(\boldsymbol{x}, \sigma)$$

$$= B(\boldsymbol{x}) \star I(\boldsymbol{x}) \star G_{xx}(\boldsymbol{x}, \sigma) + B(\boldsymbol{x}) \star I(\boldsymbol{x}) \star G_{yy}(\boldsymbol{x}, \sigma)$$

$$= B(\boldsymbol{x}) \star (I_{xx}(\boldsymbol{x}) + I_{yy}(\boldsymbol{x})) \tag{3.10}$$

$$= B(\boldsymbol{x}) \star (I(\boldsymbol{x}) \star \nabla^2 G(\boldsymbol{x}, \sigma)).$$

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

image I.

Supposing that $\boldsymbol{x_0}$ is the local maximum of $I \star \nabla^2 G$ (i.e. SIFT interest point), we are interested

to know whether $\boldsymbol{x_0}$ is also the local maximum of $B \star (I \star \nabla^2 G)$. For this to occur, we want

$$\{B(\cdot) \star I(\cdot) \star \nabla^2 G(\cdot, \sigma)\}(\boldsymbol{x}_0) \geq$$
$$\tag{3.11}$$
$$\{B(\cdot) \star I(\cdot) \star \nabla^2 G(\cdot, \sigma + \Delta\sigma)\}(\boldsymbol{x}_0 + \Delta\boldsymbol{x})$$

for all scale-space neighborhood $(\Delta\boldsymbol{x}, \Delta\sigma)$ centered around $(\boldsymbol{x}, \sigma)$. Consider the specific case of

$\Delta\sigma = 0$. By algebra, we obtain

$$0 \leq \{\delta(\boldsymbol{x}_0) - \delta(\boldsymbol{x}_0 - \Delta\boldsymbol{x})\} \star \{B(\boldsymbol{x}_0) \star I(\boldsymbol{x}_0) \star \nabla^2 G(\boldsymbol{x}_0, \sigma)\}$$
$$\tag{3.12}$$
$$= \{B(\boldsymbol{x}_0) - B(\boldsymbol{x}_0 - \Delta\boldsymbol{x})\} \star \{I(\boldsymbol{x}_0) \star \nabla^2 G(\boldsymbol{x}_0, \sigma)\}$$

One may interpret this as $B(\boldsymbol{x}_0) - B(\boldsymbol{x}_0 - \Delta\boldsymbol{x})$ filtering LOG response $I \star \nabla^2 G$. The following is the condition for the quantity on the right hand side of (3.12) would be positive valued:

$$\{I(\cdot) \star \nabla^2 G(\cdot, \sigma)\}(\boldsymbol{x}_0 + \boldsymbol{m}) \geq \{I(\cdot) \star \nabla^2 G(\cdot, \sigma)\}(\boldsymbol{x}_0 + \boldsymbol{n}) \tag{3.13}$$

for all $\boldsymbol{m}, \boldsymbol{n} \in \mathbb{Z}^2$ satisfying $\|\boldsymbol{m}\|_2 < \|\boldsymbol{n}\|_2 \leq 2q$, where q is the blur radius. That is, the LOG response $I \star \nabla^2 G$ must be monotonically increasing toward $\boldsymbol{x}_0$ over a circular neighborhood of radius $2q$. This is a far stronger constraint that must be satisfied—the likelihood that $I \star \nabla^2 G$ and $\widetilde{I} \star \nabla^2 G$ share the interest point $\boldsymbol{x}_0$ 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 $\boldsymbol{x}_0$ is a local maximum of $\det(\mathcal{H}(\boldsymbol{x}))$, one important question to ask is whether $\boldsymbol{x}_0$ would also emerge as the local maximum of the determinant $\det(\widetilde{\mathcal{H}}(\boldsymbol{x}, \sigma)) = \widetilde{I}_{xx}(\boldsymbol{x}, \sigma)\widetilde{I}_{yy}(\boldsymbol{x}, \sigma) - \widetilde{I}_{xy}(\boldsymbol{x}, \sigma)^2$. Unfortunately, one can show that the condition for $\boldsymbol{x}_0$ to (locally) maximize $\det(\widetilde{\mathcal{H}}(\boldsymbol{x}, \sigma))$ is that

$$\det(\mathcal{H}(\boldsymbol{x}_0 + \boldsymbol{m}, \sigma)) \geq \det(\mathcal{H}(\boldsymbol{x}_0 + \boldsymbol{n}, \sigma)) \tag{3.14}$$

for all $\boldsymbol{m}, \boldsymbol{n} \in \mathbb{Z}^2$ satisfying $\|\boldsymbol{m}\|_2 < \|\boldsymbol{n}\|_2 < 2q$. That is to say that $\det(\mathcal{H}(\boldsymbol{x}, \sigma))$ is monotonically increasing towards $\boldsymbol{x} = \boldsymbol{x}_0$ over a circular neighborhood of radius $2q$. As a result, we expect that the number of interest points surviving the blur decays rapidly as the size of the blur kernel grows. Of course, the requirement of (3.13) and (3.14) become even more difficult with noise. We conclude that SURF interest point detection is also not invariant to blur.

3.2.3 Scale-Space Descriptors *Are Almost* Defocus Blur Invariant

Denote by $\widetilde{U}(\boldsymbol{x}, \sigma)$ and $\widetilde{V}(\boldsymbol{x}, \sigma)$ the SIFT and SURF interest pointer descriptor corresponding to the blurred image $\widetilde{I}$, respectively. Let σ denote the scale of a *sharp* blob—we cannot know σ
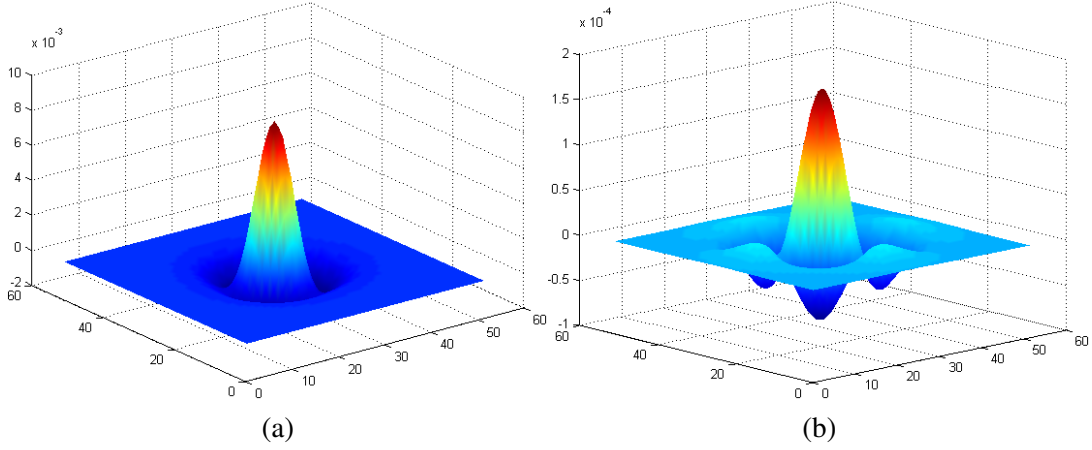
Figure 3.3: (a) Laplacian of Gaussian. (b) Fourth derivative of Gaussian.

exactly from a blurry image with blurred blobs, obviously. But for now let us assume σ is available. A question central to the context of blur is whether $U(\boldsymbol{x}, \sigma)$ and $\widetilde{U}(\boldsymbol{x}, \sigma)$ (or $V(\boldsymbol{x}, \sigma)$ and $\widetilde{V}(\boldsymbol{x}, \sigma)$) are similar enough to establish a match. For this, we rely on the Fourier analysis.

Drawing on the analysis of (3.9), we have the relation

$$\widetilde{I}_x(\boldsymbol{x}, \sigma) = B(\boldsymbol{x}) \star I_x(\boldsymbol{x}, \sigma) + N_x(\boldsymbol{x}, \sigma), \tag{3.15}$$

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

$$\begin{aligned}
\widetilde{i}_x(\boldsymbol{\omega}, \sigma) &= b(\boldsymbol{\omega}) \cdot i_x(\boldsymbol{\omega}, \sigma) + n_x(\boldsymbol{\omega}) \\
&= b(\boldsymbol{\omega}) \cdot \{j \cdot \omega_x \cdot g(\boldsymbol{\omega}, \sigma) \cdot i(\boldsymbol{\omega})\} + n_x(\boldsymbol{\omega}).
\end{aligned} \tag{3.16}$$

Here, $j \cdot \omega_x \cdot g(\boldsymbol{\omega}, \sigma)$ is the Fourier transform of Gaussian derivative filter $G_x(\boldsymbol{x}, \sigma)$ stemming from the relation [20]:

$$\frac{\partial^n}{\partial x^n} G(\boldsymbol{x}, \sigma) \rightarrow (-j\omega_x)^n g(\boldsymbol{\omega}, \sigma) = (-j\omega_x)^n\, e^{\frac{-\sigma^2 \omega^2}{2}}. \tag{3.17}$$

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

$$\omega_o = \pm \frac{\sqrt{n}}{\sigma}. \tag{3.18}$$

Together with the fact that frequency response of blur kernel $b(\boldsymbol{\omega})$ is relatively smooth, the following approximations to (3.16) and (3.15) hold when σ sufficiently large:

$$\widetilde{i}_x(\boldsymbol{\omega}, \sigma) \approx b(\tfrac{1}{\sigma}) \cdot \{ j \cdot \omega_x \cdot g(\boldsymbol{\omega}, \sigma) \cdot i(\boldsymbol{\omega}) \} + n_x(\boldsymbol{\omega})$$

$$\widetilde{I}_x(\boldsymbol{x}, \sigma) \approx b(\tfrac{1}{\sigma}) \cdot I_x(\boldsymbol{x}, \sigma) + N_x(\boldsymbol{x}, \sigma), \tag{3.19}$$

where $b(\tfrac{1}{\sigma})$ is the magnitude at center frequency of the pass band ω_o. This is clearly shown also by Figure 3.4, where an approximation in (3.19) improves as σ increases. With the interpretation of (3.19) that $I_x(\boldsymbol{x}, \sigma)$ is attenuated by the frequency response $b(\tfrac{1}{\sigma})$ (i.e. a constant), we arrive at the conclusion that the SURF interest point descriptor $\boldsymbol{V}(\boldsymbol{x}, \sigma)$ is invariant to blur when noise is sufficiently low which can be seen in Figure 3.4:

$$\widetilde{\boldsymbol{V}}(\boldsymbol{x}, \sigma) \approx \left(\frac{b(\tfrac{1}{\sigma}) \boldsymbol{V}_1(\boldsymbol{x}, \sigma)^T}{\| b(\tfrac{1}{\sigma}) \boldsymbol{V}_1(\boldsymbol{x}, \sigma) \|}, \cdots, \frac{b(\tfrac{1}{\sigma}) \boldsymbol{V}_K(\boldsymbol{x}, \sigma)^T}{\| b(\tfrac{1}{\sigma}) \boldsymbol{V}_K(\boldsymbol{x}, \sigma) \|} \right)^T$$

$$= \boldsymbol{V}(\boldsymbol{x}, \sigma). \tag{3.20}$$

Note, however, that division by $b(\tfrac{1}{\sigma})$ 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 (3.19) into (2.9), we can show that the orientation is unaffected by defocus blur:

$$\widetilde{\theta}(\boldsymbol{x}, \sigma) \approx \tan^{-1}\left(\frac{b(\tfrac{1}{\sigma}) I_y}{b(\tfrac{1}{\sigma}) I_x} \right) = \tan^{-1}\left(\frac{I_y}{I_x} \right) = \theta(\boldsymbol{x}, \sigma). \tag{3.21}$$

(a) $\omega^2 g(\omega)$ (σ small)

(b) $\omega^2 g(\omega)b(\omega)$ (σ small)

(c) $\omega^2 g(\omega)$ (σ large)

(d) $\omega^2 g(\omega)b(\omega)$ (σ large)
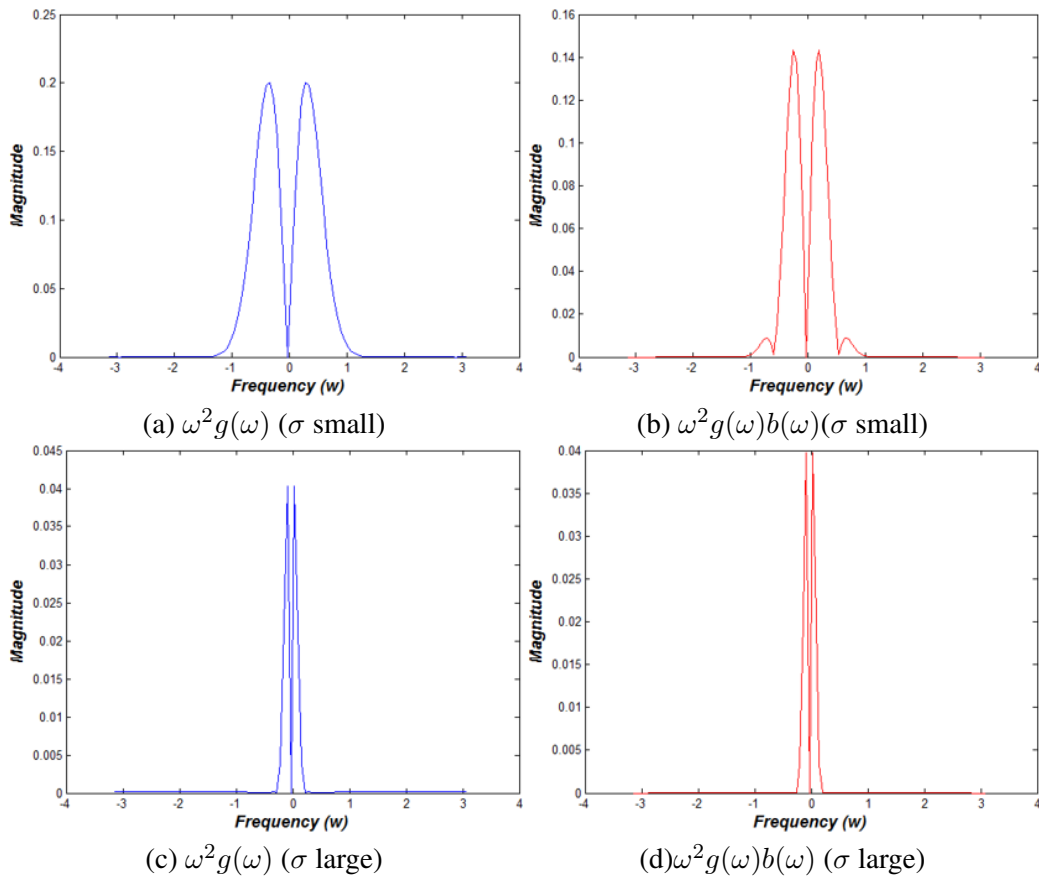
Figure 3.4: The impact of blur on the frequency response of the Gaussian derivative operators. Gaussian derivative operators are bandpass filters, where the width of the passband decreases rapidly with the scale. If the passband is narrow enough, the derivative operator's frequency response with blur is simply a scaled version with the derivative operator's frequency response without blur. See text.

Similarly for the magnitude, we have the relation:

$$\widetilde{M}(\boldsymbol{x}, \sigma) \approx \sqrt{(b(\tfrac{1}{\sigma})I_x(\boldsymbol{x}, \sigma))^2 + (b(\tfrac{1}{\sigma})I_y(\boldsymbol{x}, \sigma))^2}$$

$$= b(\tfrac{1}{\sigma})M(\boldsymbol{x}).$$

(3.22)

Subsequent normalization of the descriptor vector $\widetilde{U}(\boldsymbol{x}, \sigma)$ divides out $b(\tfrac{1}{\sigma})$ from $\widetilde{M}(\boldsymbol{x})$:

$$\widetilde{U}(\boldsymbol{x}, \sigma) \approx \left(\frac{b(\tfrac{1}{\sigma})\boldsymbol{U}_1(\boldsymbol{x}, \sigma)^T}{\|b(\tfrac{1}{\sigma})\boldsymbol{U}_1(\boldsymbol{x}, \sigma)\|}, \cdots, \frac{b(\tfrac{1}{\sigma})\boldsymbol{U}_K(\boldsymbol{x}, \sigma)^T}{\|b(\tfrac{1}{\sigma})\boldsymbol{U}_K(\boldsymbol{x}, \sigma)\|} \right)^T$$

$$= \boldsymbol{U}(\boldsymbol{x}, \sigma).$$

(3.23)

We conclude that SURF and SIFT descriptors are defocus blur-invariant, but *only* if the scale σ of the unobserved *sharp* blob is known, and noise level is not unreasonably high (no free lunch here). Though this is encouraging, question remains as to how to handle an appropriate choice of σ that enables the approximation $\widetilde{V}(\boldsymbol{x}, \sigma) \approx V(\boldsymbol{x}, \sigma)$ and $\widetilde{U}(\boldsymbol{x}, \sigma) \approx U(\boldsymbol{x}, \sigma)$. We address this issue in Chapter 4.

3.2.4   LIOP Descriptor

With noise, Let $\boldsymbol{m}$ represents the coordinates of pixels inside the circle around the $\tilde{x}_1$, and $\boldsymbol{n}$ represents the coordinates of pixels inside the circle around the $\tilde{x}_2$. For the sharp image, the ordering condition was:

$$I(x_1) > I(x_2)$$

(3.24)

For this condition to be valide with defocus blur.

$$\tilde{I}(x_1) > \tilde{I}(x_2)$$

(3.25)

$$\tilde{I}(x_1) - \tilde{I}(x_2) > 0$$

(3.26)

$$B(x) \star (I(x_1) - I(x_2)) > 0$$

(3.27)

19

$$B(x) \star [1\ 0\ 0, ... - 1] \star I(x) > 0 \tag{3.28}$$

So from (3.28), $\tilde{I}(x_1)$ comes from $\tilde{I}(x_1) = I(x_1) \star B(x)$, so $\tilde{I}(x_1)$ comes from pixel intensities around x_1 which are $\boldsymbol{m}$. Therefore (3.25) to be valid with defocus blur kernel radius r, it needs to satisfy the following condition:

$$\tilde{I}(x_1 + \boldsymbol{m}) > \tilde{I}(x_2 + \boldsymbol{n}), \tag{3.29}$$

where $|\boldsymbol{m}| < r$, and $|\boldsymbol{n}| < r$

CHAPTER IV

PROPOSED METHODS

4.1 Proposed: Defocus Blur-Invariant Scale-Space Feature Extraction

4.1.1 Defocus Blur-Invariant Scale-Space Derivatives

Based on the blur analysis of the previous Chapter, 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 4.1.3, blur-invariant scale-space feature extraction gives rise to the possibility that objects in a blurry image can be detected and recognized.

Recall (2.6)—the disk-shaped *ideal* blob attains a local maximum response $I_{xx}(\boldsymbol{x}, \sigma)$ at the center of blob $\boldsymbol{x} = \boldsymbol{x}_0$ with Gaussian scale $\sigma = r/\sqrt{2}$ where r is the blob radius. This *matched filter*-like response to Gaussian second derivative $I_{xx}(\boldsymbol{x}, \sigma)$, is shown in Figure 3.2(b). However, as illustrated by Figure 3.2(e), $\widetilde{I}_{xx}$ fails to yield a local maximum at the center of the blob when defocus blur B is introduced—consistent with our findings in Chapter 3.

Consider now the fact that the point spread function $B(\boldsymbol{x})$ 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 $G_{xx}(\boldsymbol{x}, \upsilon)$ applied to blur kernel $B(\boldsymbol{x})$ with blur radius q also attains a sharp
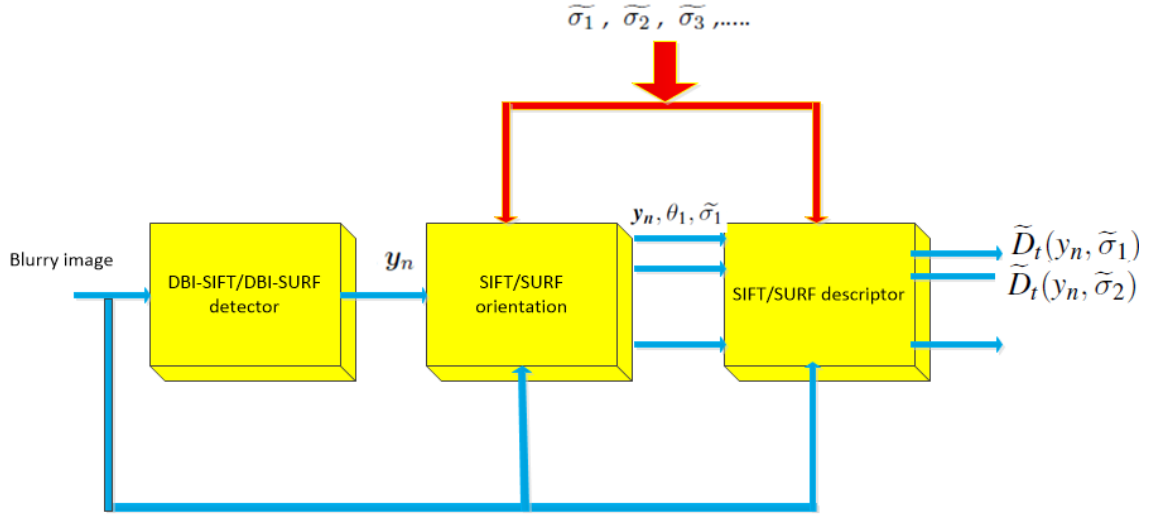
Figure 4.1: Block diagram of the proposed defocus blur-invariant blob feature extraction scheme.

maximum at the blur center $\boldsymbol{x} = \boldsymbol{0} = (0,0)^T$ and scale $\upsilon = q/\sqrt{2}$:

$$\max_{\boldsymbol{x},\upsilon} \underbrace{\left\{ \left(\frac{\partial^2}{\partial x^2} G(\cdot, \upsilon) \right) \star B(\cdot) \right\} (\boldsymbol{x})}_{B_{xx}(\boldsymbol{x},\upsilon)} = B_{xx}(\boldsymbol{0}, q/\sqrt{2}). \tag{4.1}$$

Leveraging (4.1) , consider Gaussian fourth derivative filter applied to the blurred ideal blob $\widetilde{I}$:

$$\begin{aligned}
\widetilde{I}_{xxxx}(\boldsymbol{x}, \zeta) &= \left\{ \left(\frac{\partial^4}{\partial x^4} G(\cdot, \zeta) \right) \star \widetilde{I}(\cdot) \right\} (\boldsymbol{x}) \\
&= \left\{ \left\{ \left(\frac{\partial^2}{\partial x^2} G(\cdot, \upsilon) \right) \star B(\cdot) \right\} \star \left\{ \left(\frac{\partial^2}{\partial x^2} G(\cdot, \sigma) \right) \star I(\cdot) \right\} \right\} (\boldsymbol{x}) \\
&\quad + \left(\frac{\partial^2}{\partial x^2} G(\cdot, \upsilon) \right) \star \left\{ \left(\frac{\partial^2}{\partial x^2} G(\cdot, \sigma) \right) \star N(\cdot) \right\} (\boldsymbol{x}) \\
&= B_{xx}(\boldsymbol{x}, \upsilon) \star I_{xx}(\boldsymbol{x}, \sigma) + G_{xx}(\boldsymbol{x}, \upsilon) \star N_{xx}(\boldsymbol{x}, \sigma). \tag{4.2}
\end{aligned}$$

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 (3.13) and (3.14) that causes SIFT and SURF interest point detection to be sensitive to defocus blur. Specifically, $\widetilde{I}_{xxxx}(\boldsymbol{x_0}, \sigma)$ has local maxima at $\boldsymbol{x_0}$ if (3.13) is satisfied for $\|\boldsymbol{m}\|_2 \ll \|\boldsymbol{n}\|_2 < 2p$, where p is the width of B_{xx} peak. Since $p \ll q$,

22

condition in (3.13) is more easily met with Gaussian fourth derivative than with the usual second derivative. As for noise, $N_{xx}(\boldsymbol{x}, \sigma)$ is the *usual* amount of noise seen in a second derivative detector. The additional second derivative filter $G_{xx}(\boldsymbol{x}, \upsilon)$ is the "penalty" paid by the scheme in (4.2). While the transformation of $B \mapsto B_{xx}$ or $\widetilde{I} \mapsto \widetilde{I}_{xx}$ may not suffice as image deblurring technique, it is adequate for preserving local extremal structure.

4.1.2 Defocus Blur-Invariant Interest Point Detection

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

1. **DBI-SIFT blob detector**.

We propose to replace (2.7) with

$$\widetilde{I}(\boldsymbol{x}) \star \nabla^2 G(\boldsymbol{x}, \upsilon) \star \nabla^2 G(\boldsymbol{x}, \sigma) = \widetilde{I}(\boldsymbol{x}) \star \nabla^4 G(\boldsymbol{x}, \zeta), \tag{4.3}$$

where we invoke the DOG approximation:

$$\nabla^2 G(\boldsymbol{x}, \upsilon) \approx G(\boldsymbol{x}, \sqrt{2}\upsilon) - G(\boldsymbol{x}, \upsilon)$$

$$\nabla^2 G(\boldsymbol{x}, \sigma) \approx G(\boldsymbol{x}, \sqrt{2}\sigma) - G(\boldsymbol{x}, \sigma). \tag{4.4}$$

Combining, we have,

$$\nabla^4 G(\boldsymbol{x}, \zeta) \approx \{G(\boldsymbol{x}, \sqrt{2}\upsilon) - G(\boldsymbol{x}, \upsilon)\} \star$$

$$\{G(\boldsymbol{x}, \sqrt{2}\sigma) - G(\boldsymbol{x}, \sigma)\}, \tag{4.5}$$

where $\zeta = \sqrt{\upsilon^2 + \sigma^2}$ as before. Hence DBI-SIFT interest point is the local extrema of:

$$\widetilde{I}(\boldsymbol{x}) \star \nabla^4 G(\boldsymbol{x}, \zeta) \approx$$

$$\{B(\cdot) \star \nabla^2 G(\cdot, \upsilon)\} \star \{I(\cdot) \star \nabla^2 G(\cdot, \sigma)\}(x) + \tag{4.6}$$

$$\{N(\cdot) \star \nabla^2 G(\cdot, \sigma)\} \star \nabla^2 (G(\cdot, \upsilon)).$$

23

As before, $N(\boldsymbol{x}) \star \nabla^2 G(\boldsymbol{x}, \sigma)$ is the usual amount of noise seen in (2.7). The additional $\nabla G(\boldsymbol{x}, \upsilon)$ is the "penalty" paid by scheme in (4.6). The location $\boldsymbol{x}_0$ is declared as DBI-SIFT interest point if

$$
\begin{aligned}
&\|\widetilde{I}(\boldsymbol{x}_o) \star \nabla^4 G(\boldsymbol{x}_o, \zeta)\| \\
&\geq \|\widetilde{I}(\boldsymbol{x}_o + \Delta\boldsymbol{x}) \star \nabla^4 G(\boldsymbol{x}_o + \Delta\boldsymbol{x}, \zeta + \Delta\zeta)\|
\end{aligned}
\tag{4.7}
$$

over some spatial-scale neighborhood covered by $(\Delta\boldsymbol{x}, \Delta\zeta)$.

2. **DBI-SURF blob detector.**

We propose to replace (2.14) with the following modified determinant of the Hessian matrix.

$$
\begin{aligned}
\widehat{\mathcal{H}}(\boldsymbol{x}, \zeta) &= \begin{pmatrix} \widetilde{I}_{xxxx}(\boldsymbol{x}) & \widetilde{I}_{xxyy}(\boldsymbol{x}) \\ \widetilde{I}_{xxyy}(\boldsymbol{x}) & \widetilde{I}_{yyyy}(\boldsymbol{x}) \end{pmatrix} \\
&= \begin{pmatrix} \widetilde{I}(\boldsymbol{x}) \star G_{xxxx}(\boldsymbol{x}, \zeta) & \widetilde{I}(\boldsymbol{x}) \star G_{xxyy}(\boldsymbol{x}, \zeta) \\ \widetilde{I}(\boldsymbol{x}) \star G_{xxyy}(\boldsymbol{x}, \zeta) & \widetilde{I}(\boldsymbol{x}) \star G_{yyyy}(\boldsymbol{x}, \zeta) \end{pmatrix}.
\end{aligned}
\tag{4.8}
$$

The location $\boldsymbol{x}_0$ is declared as DBI-SURF interest point if $\det(\widehat{\mathcal{H}}(\boldsymbol{x}_0, \zeta)) \geq \det(\widehat{\mathcal{H}}(\boldsymbol{x}_0 + \Delta\boldsymbol{x}, \zeta + \Delta\zeta))$ Although the fourth derivative filter gives small magnitude of the response, the local maxima point response has enough sharp peak to be detected reliably over same spatial-scale neighborhood covered by $(\Delta\boldsymbol{x}, \Delta\zeta)$.

4.1.3 Defocus Blur-Invariant Descriptor Matching Criteria

Let us begin by reviewing a classical descriptor matching criteria. Suppose $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ and $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$ are pixel coordinates of interest points in reference and target images with detected

24

scales $\{\eta_1, \ldots, \eta_M\}$ and $\{\sigma_1, \ldots, \sigma_N\}$, respectively. Let $\boldsymbol{D}_r(\boldsymbol{x}_m, \eta)$ and $\boldsymbol{D}_t(\boldsymbol{y}_n, \sigma_n)$ be the descriptors of $\boldsymbol{x}_m$ and $\boldsymbol{y}_n$, respectively—they may refer to SIFT ($\boldsymbol{U}(\boldsymbol{x}, \sigma)$) or SURF ($\boldsymbol{V}(\boldsymbol{x}, \sigma)$) descriptors but it makes no difference to the presentation below. In a criteria known as nearest neighborhood (NN) search, a target interest point $\boldsymbol{y}_{\widehat{m}}$ is matched to the reference interest point $\boldsymbol{x}_m$ by [7]:

$$\widehat{m} = \arg\min_n \|\boldsymbol{D}_r(\boldsymbol{x}_m, \eta_m) - \boldsymbol{D}_t(\boldsymbol{y}_n, \sigma_n)\|_2. \tag{4.9}$$

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:

$$\|\boldsymbol{D}_r(\boldsymbol{x}_m, \eta_m) - \boldsymbol{D}_t(\boldsymbol{y}_{\widehat{m}}, \sigma_{\widehat{m}})\|_2 < \tau \tag{4.10}$$

for some threshold τ. The second condition is known as nearest neighborhood distance ratio (NNR) [2, 7]:

$$\frac{\|\boldsymbol{D}_r(\boldsymbol{x}_m, \eta_m) - \boldsymbol{D}_t(\boldsymbol{y}_{\widehat{m}}, \sigma_{\widehat{m}})\|_2}{\|\boldsymbol{D}_r(\boldsymbol{x}_m, \eta) - \boldsymbol{D}_t(\boldsymbol{y}_n, \sigma_n)\|_2} < t, \forall n \neq \widehat{m}. \tag{4.11}$$

The intuition behind the NNR is that descriptor error $\|\boldsymbol{D}_r(\boldsymbol{x}_m, \eta_m) - \boldsymbol{D}_t(\boldsymbol{y}_{\widehat{m}}, \sigma_{\widehat{m}})\|_2$ must be sufficiently smaller than the error of a non-match $\|\boldsymbol{D}_r(\boldsymbol{x}_m, \eta_m) - \boldsymbol{D}_t(\boldsymbol{y}_n, \sigma_n)\|_2$. Hence, varying $t \in [0, 1]$ gives us a sense of distinctiveness or discriminativity of the descriptors. In [2], the optimum NNR threshold value t was experimentally determined to be 0.85 for sharp images.

Consider now the scenario where the target image is blurred. Define $\widetilde{\boldsymbol{D}}_t(\cdot, \cdot)$ 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 σ is known, there exists $\widetilde{\sigma}$ such that the following approximation holds:

$$\widetilde{\boldsymbol{D}}_t(\boldsymbol{y}_{\widehat{m}}, \widetilde{\sigma}) \approx \boldsymbol{D}_t(\boldsymbol{y}_{\widehat{m}}, \sigma_{\widehat{m}}), \tag{4.12}$$

25
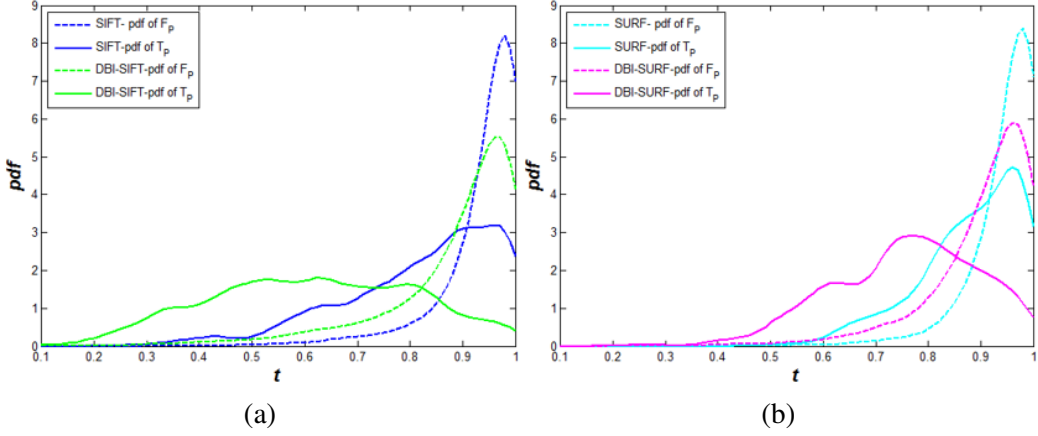
Figure 4.2: Probability density functions (pdf) of true (T_p) and false (F_p) matches admitted by the choice of NNR threshold value t in (4.11). The proposed DBI-SIFT/DBI-SURF matching criteria restores the distinctiveness of the interest point descriptor, making it easier to differentiate between the positive and false matches.

that is, a descriptor $\widetilde{\boldsymbol{D}}_t$ computed with a scale other than the detected scale ζ. In practice, the detected blur size $\zeta = \sqrt{v^2 + \sigma^2}$ is larger than the σ 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 [21] that interprets descriptor matching criteria as a function of $\widetilde{\sigma}$. We modify NN search as a minimization of the form

$$\widehat{m} = \arg\min_n \left\{ \min_{\widetilde{\sigma}} \| \boldsymbol{D}_r(\boldsymbol{x}_m, \eta_m) - \widetilde{\boldsymbol{D}}_t(\boldsymbol{y}_n, \widetilde{\sigma}) \|_2 \right\}. \tag{4.13}$$

That is, we draw on the distinctiveness of the descriptors to claim that $\widetilde{\boldsymbol{D}}_t(\boldsymbol{y}_{\widehat{m}}, \widetilde{\sigma})$ with the minimum descriptor error suffices as a proxy for $\boldsymbol{D}_t(\boldsymbol{y}_{\widehat{m}}, \sigma_{\widehat{m}})$. A one-dimensional search over $0 < \widetilde{\sigma} < \zeta$ can be executed in a reasonable time.

When both the reference and the target images are blurred, there also exists a scale $\widetilde{\eta}$ such that

$$\widetilde{\boldsymbol{D}}_r(\boldsymbol{x}_m, \widetilde{\eta}) \approx \boldsymbol{D}_r(\boldsymbol{x}_m, \eta_m), \tag{4.14}$$

26

where the descriptor $\widetilde{\boldsymbol{D}}_r$ is computed from a blurry reference image. Drawing on the distinctiveness of the descriptors again, the following is proxy for (4.9):

$$\widehat{m} = \arg\min_n \left\{ \min_{\widetilde{\eta},\widetilde{\sigma}} \|\widetilde{\boldsymbol{D}}_r(\boldsymbol{x}_m,\widetilde{\eta}) - \widetilde{\boldsymbol{D}}_t(\boldsymbol{y}_n,\widetilde{\sigma})\|_2 \right\}. \tag{4.15}$$

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 Chapter 6 verifies. The proposed defocus blur-invariant descriptor matching criteria of (4.13) and (4.15) are used in conjunction with thresholdings in (4.10) and (4.11).

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 (4.13) and (4.15). This is clearly evidenced by Figure 4.2. 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 (4.11) have significant overlap. As clearly shown by Figure 4.2, 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 (4.9) for SIFT/SURF clearly suffers from false rejection/matching. Based on this analysis, the thresholding value of $t = 0.80$ as empirically determined in [2] remains to be an appropriate choice for blurred scale-space feature matching.

CHAPTER V

PERFORMANCE EVALUATION

5.1 Data Set

The proposed algorithms are tested on real images taken by Nikon D9 with Nikkor f/1.8D 50mm 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 [22], color correction, and white balance [23] before feature extraction was performed. Gamma correction was applied after feature extraction owing to the fact that the blur model of (3.2) will not hold after nonlinear tone mapping. Example images are shown in Figure 5.2. (Upon acceptance of this paper, we will make this dataset available on our research website.)

Similar to [3, 7], each scene is photographed multiple times with zoom, rotation, and/or view angles of varying degrees. But for each zoom/rotation/angle setting, we captured 7 images while varying the aperture so that we obtain sharp and blurry images—we refer to the grouping of these 7 images as "*aperture stack*." We claim this procedure results in variation in lighting and noise, since large f-number images are very dark. Iin fact, the light efficiency of f/1.8 (minimum Nikkor setting) is $149.38 = (\frac{22}{1.8})^2$ times more than the f/22 (maximum Nikkor setting)—for display, images in Figure 5.2 are shown with the lighting variation normalized, however. We note that there were no pixels that are saturated other than the sky region. The scenes captured with viewing angle

28

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. Figure 6.1 shows a detailed example of images with zoom and blur variations—the columns are organized by a zoom setting (zoom set 1 through 4), while rows are organized by aperture stack (images 1 and 7 are taken with f/22 and f/1.8, and therefore are the sharpest and blurriest images, respectively). The remainder of the images are shown in the supplementary material.

A homography matrix describes the coordinate transforms between different zoom/rotation/view angle images. It takes the following form:

$$H = \begin{pmatrix} z\cos\phi & -z\sin\phi & z\,t_x \\ z\sin\phi & z\cos\phi & z\,t_y \\ 0 & 0 & 1 \end{pmatrix},$$ (5.1)

where z is the zoom, ϕ is the rotation angle, and t_x and t_y are the x- and y-direction translations, respectively. In our dataset, homography matrices were computed using the f/22 images (top row of Figure 6.1, for example), 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 (e.g. all images within a column of Figure 6.1).

For computing the homograaphy matrix, we followed the procedure of [3, 7]: first, an approximated homography matrix H_0 between the reference image and the target image was computed by selecting correspondence points manually. We then warped the target image to match the referenc image coordinates, and estimated the residual homography H_1 by matching SURF interest points. The overall homography is obtained as $H = H_1 \cdot H_0$.
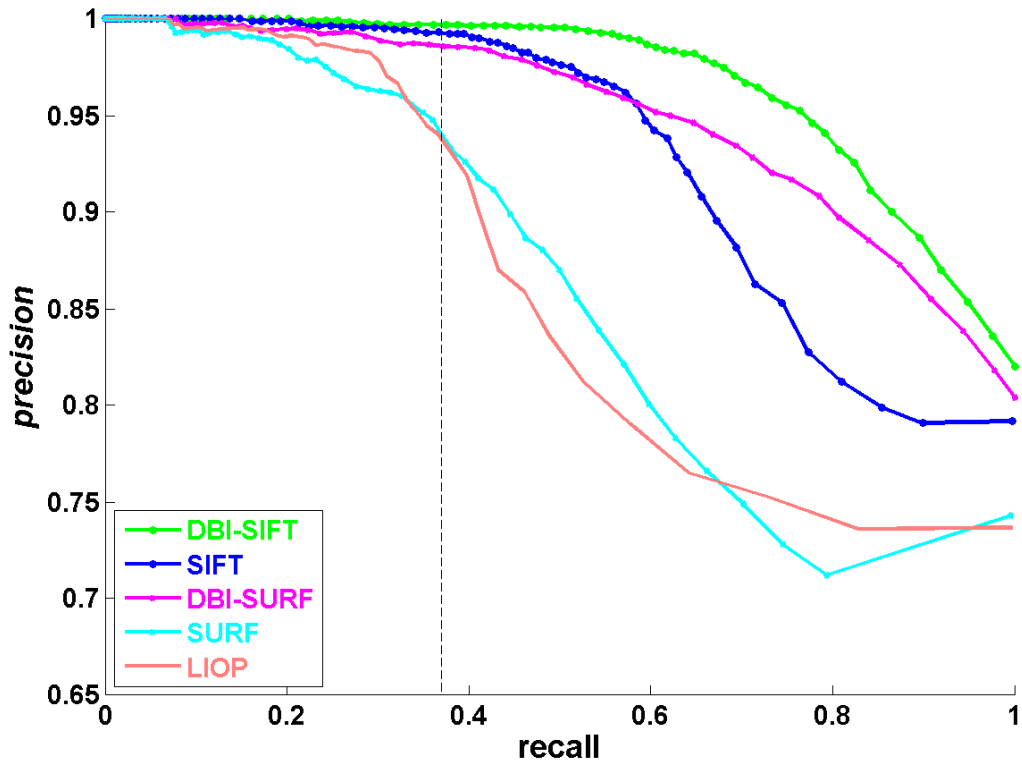
Figure 5.1: Recall-precision curve computed as a parametric function of the NNR threshold in (4.11). This example is showing results for matching descriptors computed over sharp reference image with descriptors computed on blurred target images. The vertical line shows precision of each descriptors for a given recall rate. In Figure 6.2 we plot this as a function of defocus blur severity.

5.2 Repeatability Criterion:

Pixel coordinates $\boldsymbol{x}$ and $\boldsymbol{y}$ of the detected interest points in reference and target images, respectively, are said to be "repeated" or "corresponding" if for some threshold ε,

$$||\boldsymbol{x} - H\boldsymbol{y}||_2 < \varepsilon, \tag{5.2}$$

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. In [3], the *repeatability* of the detector is defined as the ratio between the number of interest point locations repeated across two images and the total number of interest points detected in the images. We evaluate the influence defocus blur has on repeatability by plotting the number of repeated interest points in the target image as a function of image index within an aperture stack.

5.3 Recall-Precision Criterion:

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 $\boldsymbol{x}_m$ and $\boldsymbol{y}_{\widehat{m}}$ established by (4.10) and (4.11) is "correct" if they are also repeated (i.e. satisfying (5.2)). Recall describes the ability for descriptors to correctly match the corresponding interest points by the ratio:

$$recall = \frac{\#\ correct\ match}{\#\ correspondences}. \tag{5.3}$$

Since recall does not penalize against false matches made by the matching algorithm, a separate quantity to assess precision is needed. In [7], "precision" score is a percentage of interest points

31

satisfying (4.10) and (4.11) that are correct (i.e. satisfy (5.2)):

$$precision = \frac{\#\ correct\ match}{\#\ total\ match}.$$ (5.4)

If the value of $t \in [0, 1]$ in (4.11) increased, the number of correctly and incorrectly matched interest points increase. 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 the one shown in Figure 5.1—are popular ways to characterize the discriminative trends of the descriptors. We computed the precision-recall curve for varying severities of blur. To understand the impact of defocus blur on the descriptors, we summarize the overall precision-recall trend by plotting the precision at a fixed recall rate (a vertical line in Figure 5.1) as a function of the image index within an aperture stack.

5.4   Matching Score:

To assess the overall quality of feature detection, a matching score describing the combined behavior of interest point detection and descriptors is needed. In [7], the matching score is a percentage of detected interest points that are correctly matched by the descriptor.:

$$matching = \frac{\#\ correct\ match}{\#\ detected\ interest\ points}.$$ (5.5)

We evaluate the influence of blur on the overall feature extraction by plotting the matching score at a fixed number of interest points as a function of image index within an aperture stack. plotting the matching score at a fixed number of interest points as a function of image index within an aperture stack. In computing the matching score, we used NNR threshold of $t = 1$ in (4.11) because (5.5) does not penalize the false match.

zoom set 1, image 1   zoom set 1, image 7   zoom set 4, image 1   zoom set 4, image 7

zoom set 1, image 1   zoom set 1, image 7   zoom set 4, image 1   zoom set 4, image 7

rotation set 1, image 1   rotation set 1, image 7   rotation set 4, image 1   rotation set 4, image 7

rotation set 1, image 1   rotation set 1, image 7   rotation set 4, image 1   rotation set 4, image 7

view angle set 1, image 1   view angle set 1, image 7   view angle set 4, image 1   view angle set 4, image 7

view angle set 1, image 1   view angle set 1, image 7   view angle set 4, image 1   view angle set 4, image 7

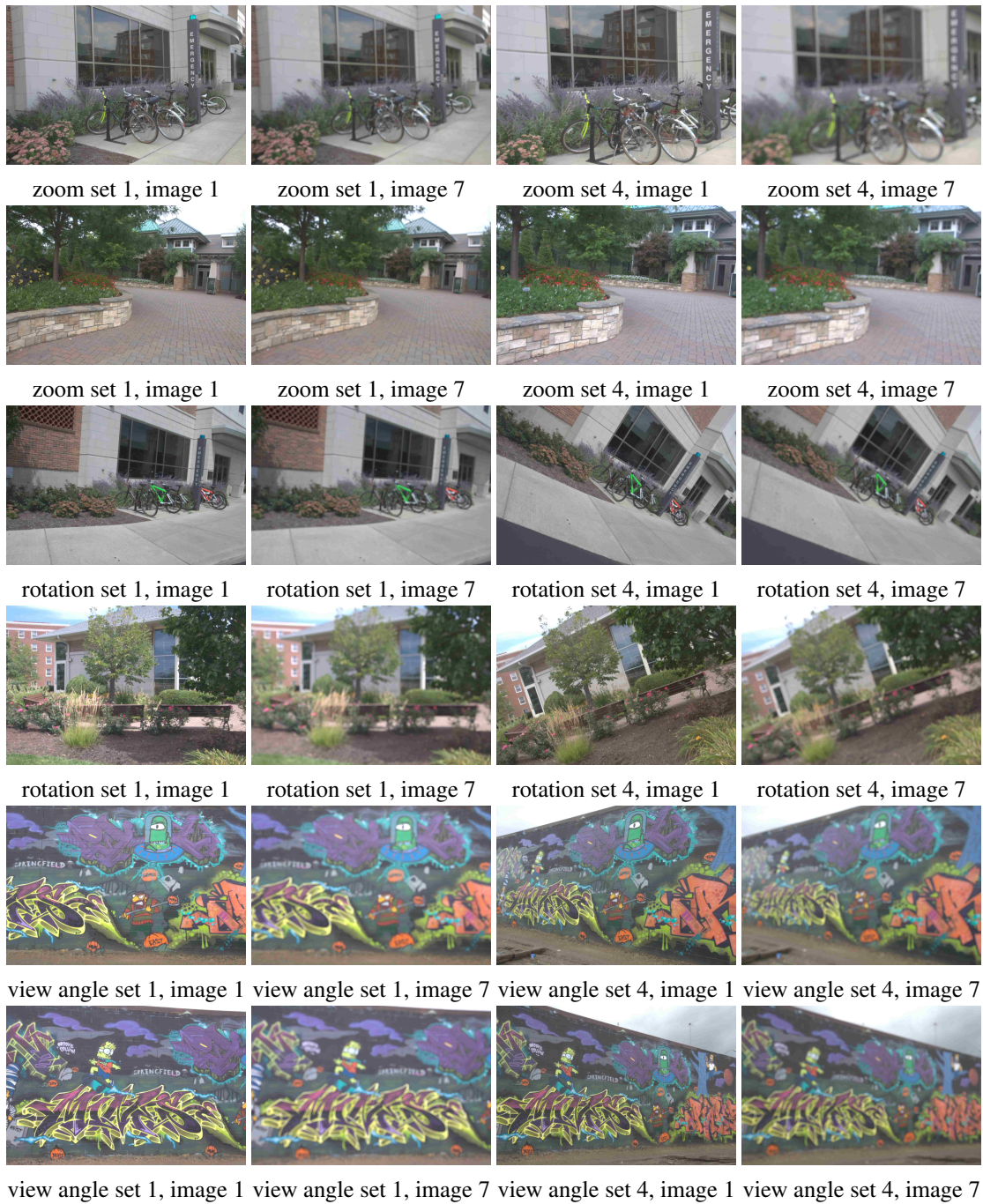Figure 5.2: Examples of images from our data set. Image 1's are the sharpest pictures taken with f/22. Image 7's are the blurriest pictures taken with f/1.8.

CHAPTER VI

EXPERIMENTS RESULTS

6.1   Experiments

In this Chapter, we discuss in detail an example provided in Figure 6.1. The analysis of the experimental results in the supplementary materials follows with similar behavioral trends.

6.1.1   Setup

In our first experiment, we evaluated the proposed detectors and descriptors (DBI-SIFT and DBI-SURF) by establishing correspondence between sharp reference image and blurred target images with varying zoom/rotation/angle changes. SIFT and SURF interest points are first detected on the sharp reference images (e.g. zoom set 1, image 1 in Figure 6.1), while DBI-SIFT and DBI-SURF interest points are computed on the blurred target images (e.g. zoom sets 1-4, images 1-7 in Figure 6.1). The descriptors of the SIFT/SURF interest points computed on the sharp reference image is matched to the descriptors of the DBI-SIFT/DBI-SURF interest points computed on the blurred target image using the matching criteria of (4.13).

In the second experiment, both reference and target images are blurred (with zoom/rotation/angle changes as before). In this case, DBI-SIFT/DBI-SURF interest points and their descriptors were

computed in the reference image (e.g. zoom set 1, image 5 in Figure 6.1), and matched to DBI-SIFT/DBI-SURF detection/descriptors in the target image (e.g. zoom sets 1-4, images 1-7 in Figure 6.1). Matching criteria used was (4.15).

The quality of interest point detection was evaluated by the repeatability test [3]. The descriptor quality was evaluated by recall/precision test [7]. Overall detection/descriptor score is produced as well [3, 7]. Our results were compared to the ordinary SIFT [2], ordinary SURF [4], Harris-Laplace [3], and LIOP [6].

6.1.2 Results of Experiment #1

In Figure 6.1, four zoom variations are shown in the columns of the figures—i.e. images/plots appearing within the same column correspond exactly to the same zoom setting—while the first four rows show images with increasing degrees of blur (image 1=f/22, image 7=f/1.8). In each image, the colored points are the detected SIFT (blue) and DBI-SIFT (green) interest points, respectively. The images in rows 5-8 are identical to the rows 1-4; however, the SIFT interest points are replaced by the SURF (cyan) and DBI-SURF (magenta) interest points.

As a point of reference, the SIFT (red) and SURF (yellow) interest points computed on the "zoom set 1, image 1" (sharp and least zoomed). These points were projected to each zoomed pixel coordinates via homography transformation, and they are shown in every image. By comparing blue/green points to red (or cyan/magenta points to yellow), one can appreciate the stability of the proposed DBI-SIFT (green) and DBI-SURF (magenta) interest points compared to the ordinary SIFT (blue) and SURF (cyan) interest points. In particular, in the "zoom set 1, image 1" image, the red and yellow points coincide perfectly with the blue and cyan points, respectively, since they are indeed calculated in an identical manner. However, ordinary SIFT and SURF interest points

become randomly scattered as the blur strength increases. Contrast this to the DBI-SIFT and DBI-SURF interest points that remain close to the interest points detected by the "zoom set 1, image 1" interest points marked in red/yellow, even with severe blur and zoom.

In our first experiment, we chose "zoom set 1, image 1" (sharp and least zoomed; red bounding box in Figure 6.1) as the reference image. The remainder images in the first four rows of Figure 6.1 are the target images that differ from the reference image in terms of zoom and blur. In row 1 of Figure 6.2, the repeatability scores for each of the four zoom settings are reported for increasing levels of image blur. Intuitively, the repeatability score can be understood in Figure 6.1 as thresholding on the distance between red points (reference SIFT) and the nearest blue/green points (SIFT/DBI-SIFT target); or between the yellow points (reference SURF) and the nearest cyan/magenta points (target SURF/DBI-SURF).

While it is clear that blur deteriorates repeatability for all interest points detected (blur increased from f/22 to f/1.8 coinciding with the image index in each plot), DBI-SIFT and DBI-SURF demonstrates considerable robustness against blur compared to the ordinary SIFT and SURF. In particular, the ordinary SIFT/SURF outperform DBI-SIFT/DBI-SURF for sharp images, but the repeatability of SIFT/SURF rapidly decays with blur while DBI-SIFT/DBI-SURF enjoy a far slower decay rate. In fact, the repeatability of DBI-SIFT/DBI-SURF improved with a moderate degree of blur—this is consistent with our understanding since the fourth derivative is inconsistent with the sharp images. We emphasize that the proposed DBI-SIFT and DBI-SURF remain zoom-invariant also, as evidenced by the stable trends along the columns of Figure 6.2. We included Harris-Laplace corner detectors—though the original intention of the Laplace operator was for detecting corners through different scales [3], our analysis in Chapter 4 led us to suspect that higher order derivative may also

increase robustness against defocus blur. Indeed, Harris-Laplace is competitive against the proposed DBI-SIFT and DBI-SURF interest point detection.

Plots in row 2 of Figure 6.2 assess the quality of the interest descriptor in terms of precision and recall scores. For each zoom settings, SIFT interest points $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$ were first detected on the sharpest image (taken with f/22) within each aperture stack (e.g. images in the first row of Figure 6.1). SIFT/SURF/DBI-SIFT/DBI-SURF/LIOP descriptors were then computed on every image of the aperture stack over the same interest points $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$. These descriptors are matched to the SIFT/SURF/DBI-SIFT/DBI-SURF/LIOP descriptors computed on the interest points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$, where $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ correspond to the SIFT interest points detected on the sharp reference image (rows 5, zoom set 1 in Figure 6.1). With reference ($\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$) and target ($\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$) interest point locations fixed, this procedure ensures that precision-recall analysis is decoupled from the quality of interest point detection.

Since both precision and recall scores vary with the NNR thresholding value of $t \in [0, 1]$ in (4.11), plots in row 2 of Figure 6.2 report the precision values for a fixed recall score value (recall vertical line in Figure 5.1). Although SIFT/SURF and DBI-SIFT/DBI-SURF are competitive for sharp images, it is clear from this plot that the blur is detrimental to the precision of descriptor matching based on NN-search criteria. By contrast, the rate at which the precision of the multi-scale matching criteria of DBI-SIFT/DBI-SURF decay with the increased blur is slower. We conclude that DBI-SIFT/DBI-SURF retain discriminativity more effectively under the influence of severe defocus blur.

The matching score shown by the plot in row 3 of Figure 6.2 assesses the combined interest point detection/descriptor performance. The performance evaluation based on matching score follow the general trends of the repeatability and precision. The proposed DBI-SIFT/DBI-SURF are less

effective than SIFT/SURF when the images are sharp, but they exhibit performance advantages when the target image suffers from substantial defocus blur.

6.1.3 Results of Experiment #2

We chose "zoom set 1, image 5" (blurred and least zoomed; yellow bounding box in Figure 6.1) as the reference image for our second experiment. The remainder images in the first four rows of Figure 6.1 are the target images that differ from the reference image in terms of zoom and sharpness/blurriness. Figure 6.3 report the results of the second experiment, organized in a similar manner to Figure 6.2. As an overall trend, the repeatability, precision-recall, and matching scores of SIFT/SURF/DBI-SIFT/DBI-SURF/Harris-Laplace/LIOP typically peak near images 4 and 5, since the level of blur is well matched between the reference and the target images. The performance of all interest point detection/descriptor deteriorates when the target images are blurrier or sharper.

The performance of DBI-SIFT/DBI-SURF was significantly better than SIFT/SURF/Harris-Laplace/LIOP for all target images (sharp and blurry), except of course when the target image was identical to the reference image. This is not surprising, as the ordinary SIFT/SURF interest point detection and descriptor for the blurred reference image were less meaningful. However, we also note that the computational complexity of an increased search dimension in (4.15) was also significant.

6.1.4 Discussion

The rest of the figures follow the same format as the Figures 6.1-6.3(i-vi) SIFT (blue) and DBI-SIFT (green) interest points detected from each image, respectively. (v-viii) SURF (cyan) and DBI-SURF (magenta) interest points detected from each image, respectively. For reference, SIFT (red) and SURF (yellow) are interest points computed over "view angle set 1, image 1" (in red bounding

38

box), respectively. By homography transformation, the red points are projected into the coordinates of each view angle setting. By comparing the positions of blue/green to red points (or cyan/magenta to yellow points), one can assess the robustness of the interest point detection to defocus blur. Red and yellow bounding boxes indicate sharp and blurred reference images in experiments 1 and 2, respectively. where the detected interest points and repeatability/precision-recall/matching scores of the two experiments are shown. Each scene set is comprised of at least 28 images—at least four zoom/rotation/angle changes, and there are seven aperture stack images of varying blur severities for each zoom/rotation/angle setting. The general performance trends of DBI-SIFT/DBI-SURF agrees with our analysis above of Figures 6.1—6.3. In the first experiment (with sharp reference image), DBI-SIFT/DBI-SURF showed an improvement over SIFT/SURF when the blur in the target image was severe. The extent of the improvement depended on the image context, however. In the second experiment (with blurred reference image), DBI-SIFT/DBI-SURF demonstrated substantial improvements over SIFT/SURF since the interest point detection/descriptor computed on the blurred reference images is more meaningful. We may conclude that the proposed DBI-SIFT/DBI-SURF tolerate significantly more defocus blur than the ordinary SIFT/SURF, even while retaining invariance to zoom/rotation/angle and robustness to noise/lighting.

Table 6.1 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 Chapter4 depends on the approximation in (3.19) hold better.

Table 6.1: Execution average time for matching image sharp image and transformed blurred target images (21 images).

average	SIFT	DBI-SIFT	SURF	DBI-SURF
# points in ref. image	4700	4700	4700	4700
# points in target image	4355	4355	4355	4355
# of correct match	1247	1928	998	1558
execution time	23s	80s	17s	63s

It is clear from Table 6.1, 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 match for the proposed descriptors is significantly more than the number of correct match 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 [4] and faster searches for matches inspired by KD-trees [24], FLANN [25], and IVFADC [26]. Since our multi-scale descriptor matching criteria is similar to that of [21], we also expect speed improvements by leveraging the optimization strategies taken in [21].

6.2 Conclusion

In this article, we proved that scale-space feature extraction techniques of SIFT and SURF are not invariant to defocus blur. The proposed DBI-SIFT and DBI-SURF detectors use Gaussian fourth derivative filters on blurred blobs, restoring the response that Gaussian second derivative filter 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.
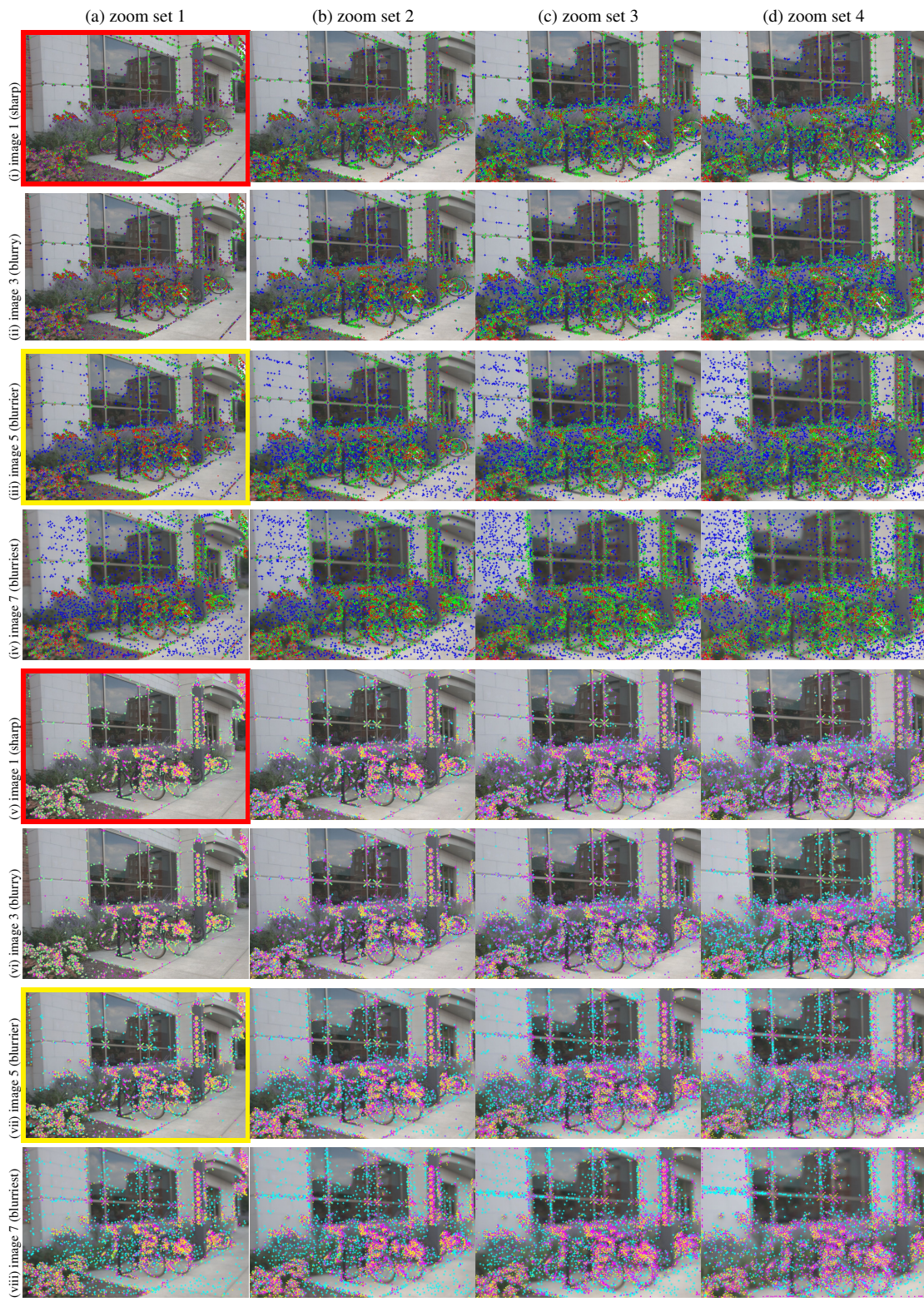
Figure 6.1: Stability of interest point detection under the influence of zoom and defocus blur.
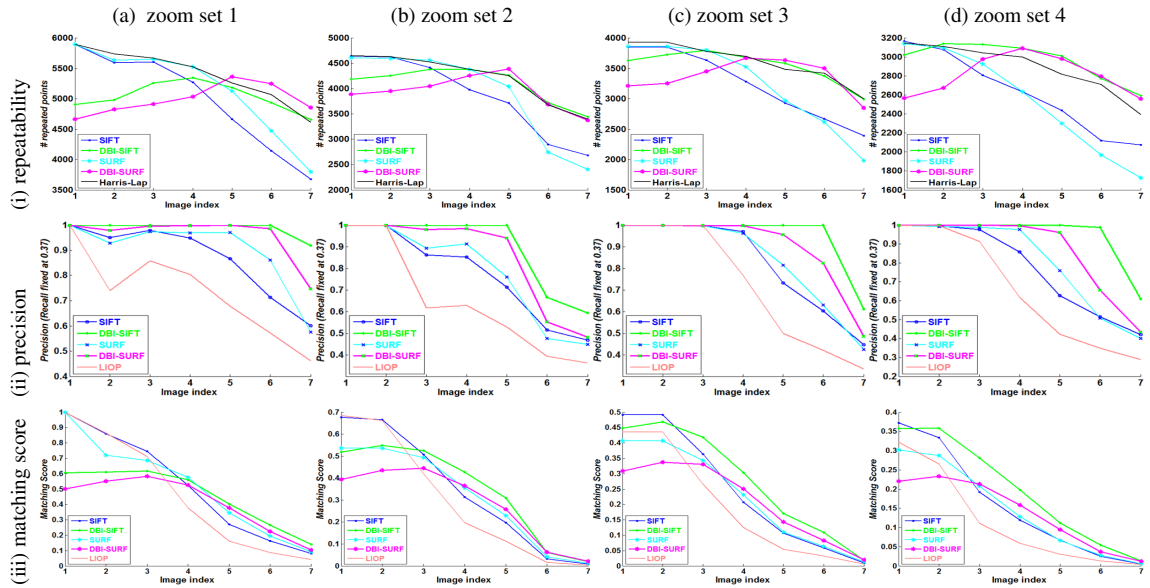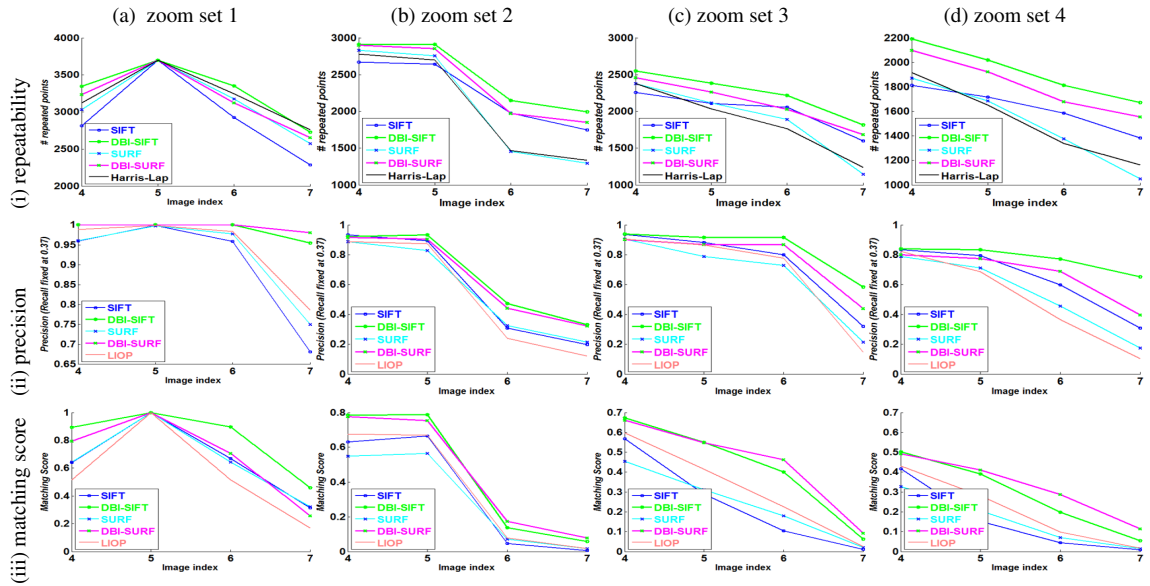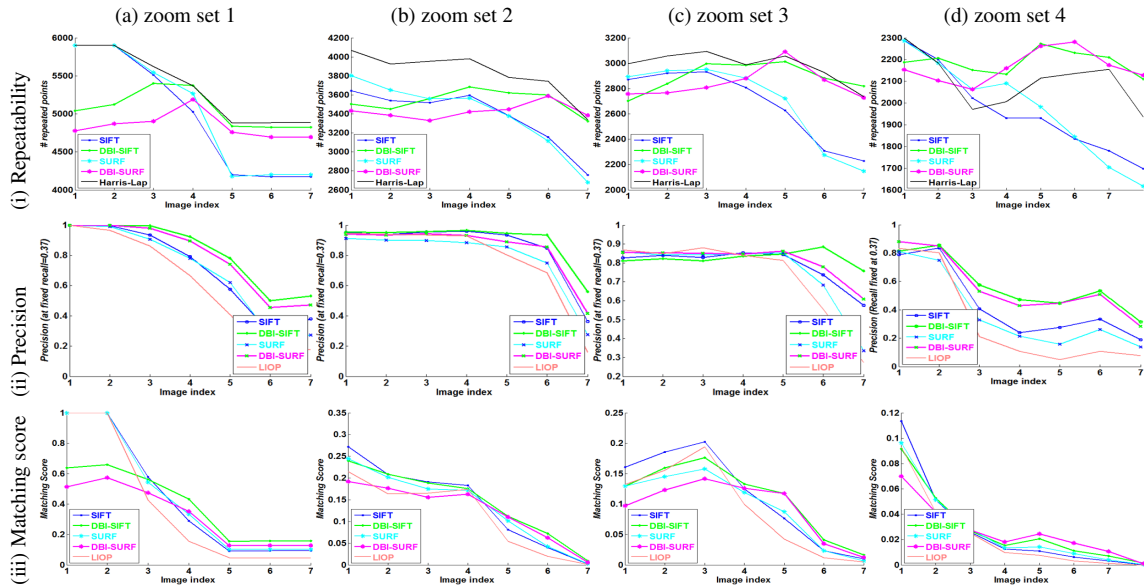
Figure 6.2: Evaluation of detectors and descriptors under the influence of zoom and defocus blur. The curves represent the relation between the features of sharp reference image (red bounding box in Figure 6.1) and the target images from all zoom/blur settings. The plots in the second row are taken from precision-recall curve, where the recall value fixed to 0.37 (see Figure 5.1).

Figure 6.3: Evaluation of detectors and descriptors under the influence of zoom and defocus blur. The curves represent the relation between the features of blurred reference image (yellow bounding box in Figure 6.1) and the target images from all zoom/blur settings. The plots in the second row are taken from precision-recall curve, where the recall value fixed to 0.37 (see Figure 5.1).
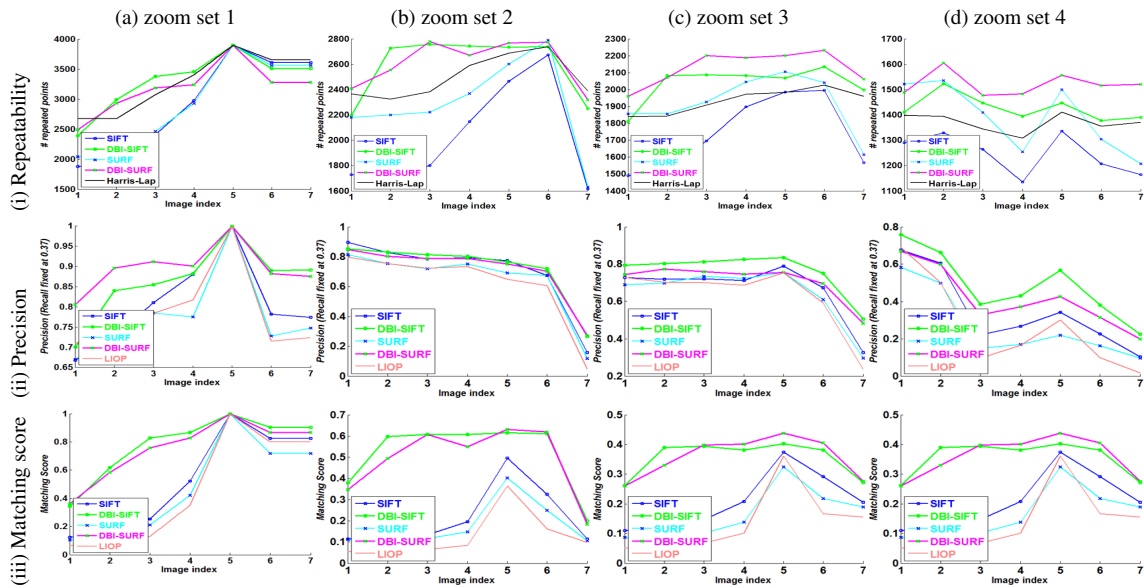
Figure 6.4: Stability of interest point detection under the influence of zoom and defocus blur.

43

Figure 6.5: Evaluation of detectors and descriptors under the influence of zoom and defocus blur. The curves represent the relation between the features of sharp reference image (red bounding box in Figure 6.4) and the target images from all zoom/blur settings. The plots in the second row are taken from precision-recall curve, where the recall value fixed to 0.37.

Figure 6.6: Evaluation of detectors and descriptors under the influence of zoom and defocus blur. The curves represent the relation between the features of blurred reference image (yellow bounding box in Figure 6.4) and the target images from all zoom/blur settings. The plots in the second row are taken from precision-recall curve, where the recall value fixed to 0.37.

Figure 6.7: Stability of interest point detection under the influence of rotation and defocus blur.
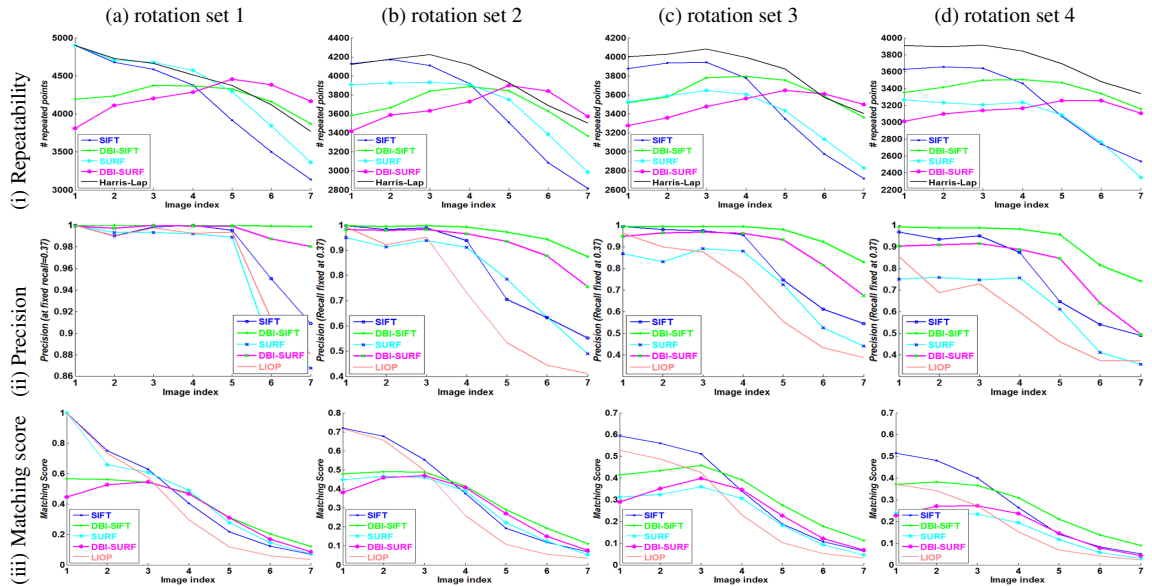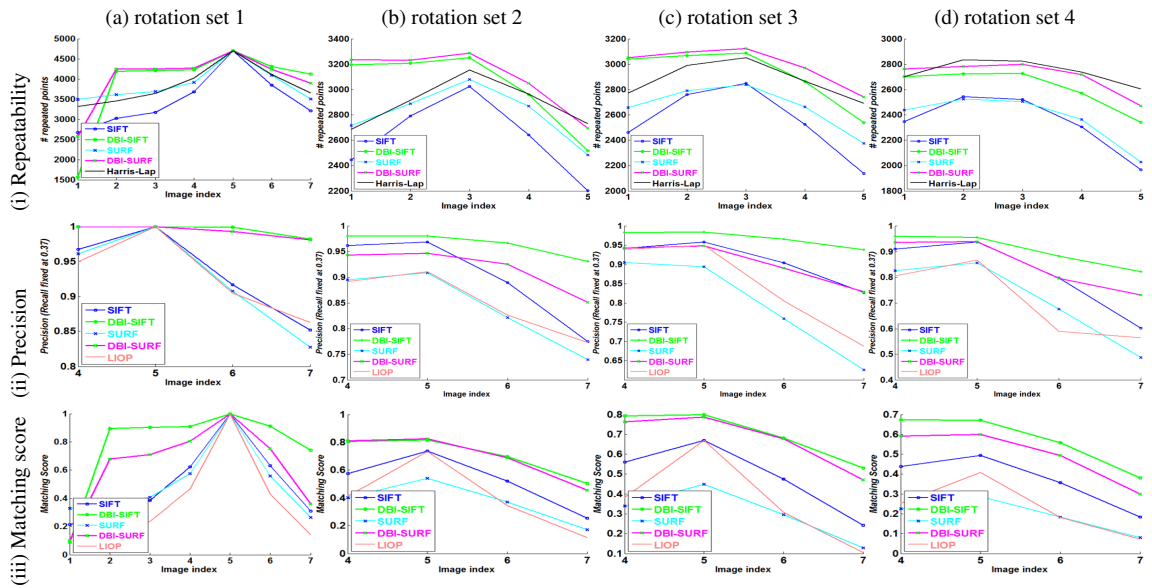
Figure 6.8: Evaluation of detectors and descriptors under the influence of rotation and defocus blur. The curves represent the relation between the features of sharp reference image (red bounding box in Figure 6.7) and the target images from all rotation/blur settings. The plots in the second row are taken from precision-recall curve, where the recall value fixed to 0.37.
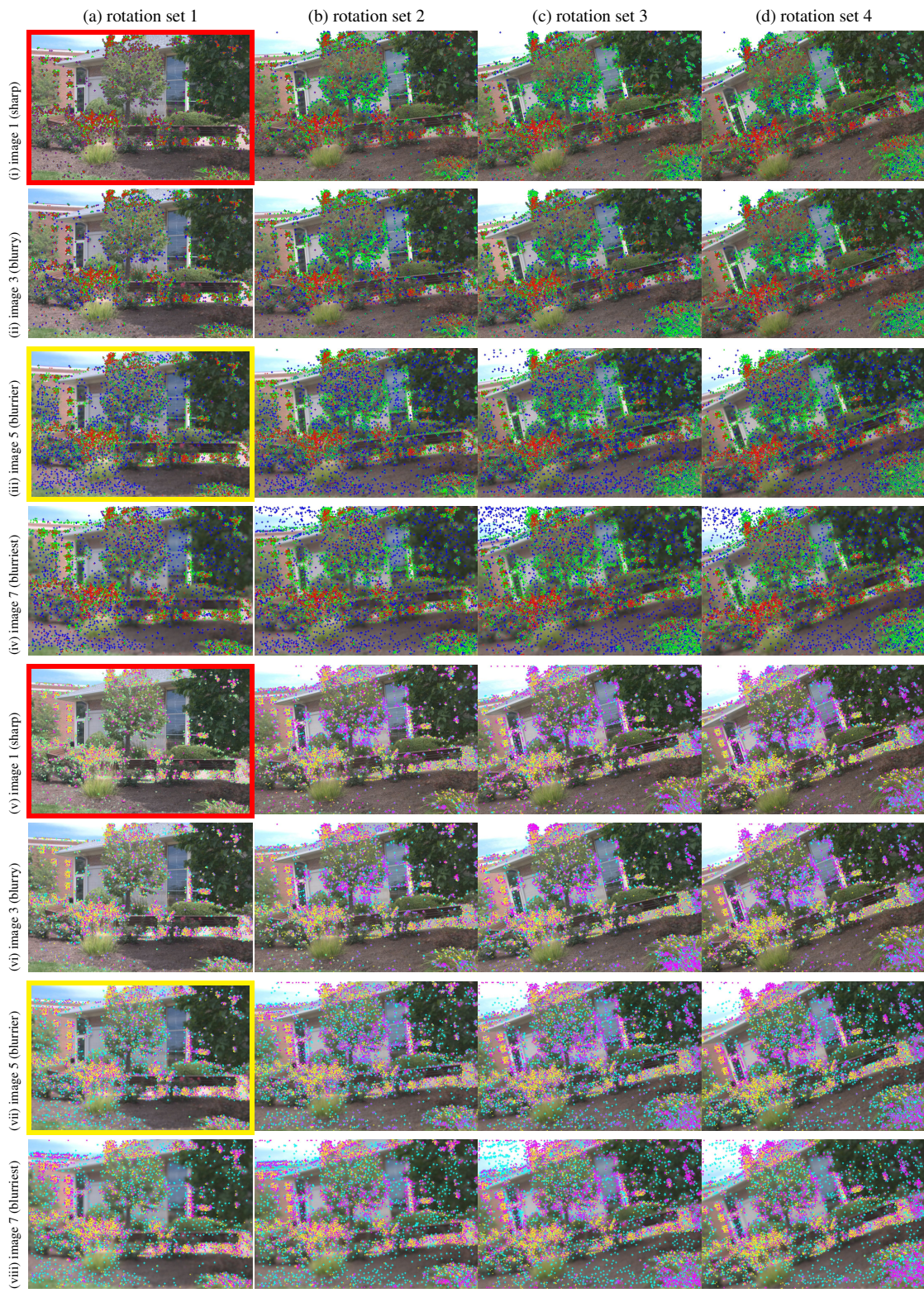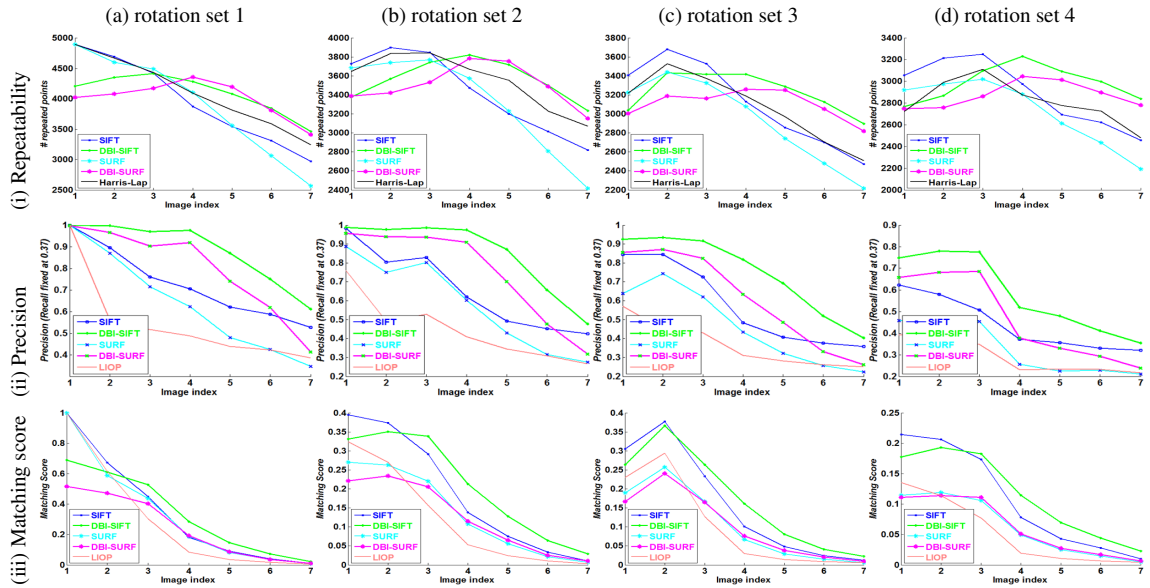

Figure 6.9: Evaluation of detectors and descriptors under the influence of rotation and defocus blur. The curves represent the relation between the features of blurred reference image (yellow bounding box in Figure 6.7) and the target images from all rotation/blur settings. The plots in the second row are taken from precision-recall curve, where the recall value fixed to 0.37.

46

Figure 6.10: Stability of interest point detection under the influence of rotation and defocus blur.

47

Figure 6.11: Evaluation of detectors and descriptors under the influence of rotation and defocus blur. The curves represent the relation between the features of sharp reference image (red bounding box in Figure 6.10) and the target images from all rotation/blur settings. The plots in the second row are taken from precision-recall curve, where the recall value fixed to 0.37.
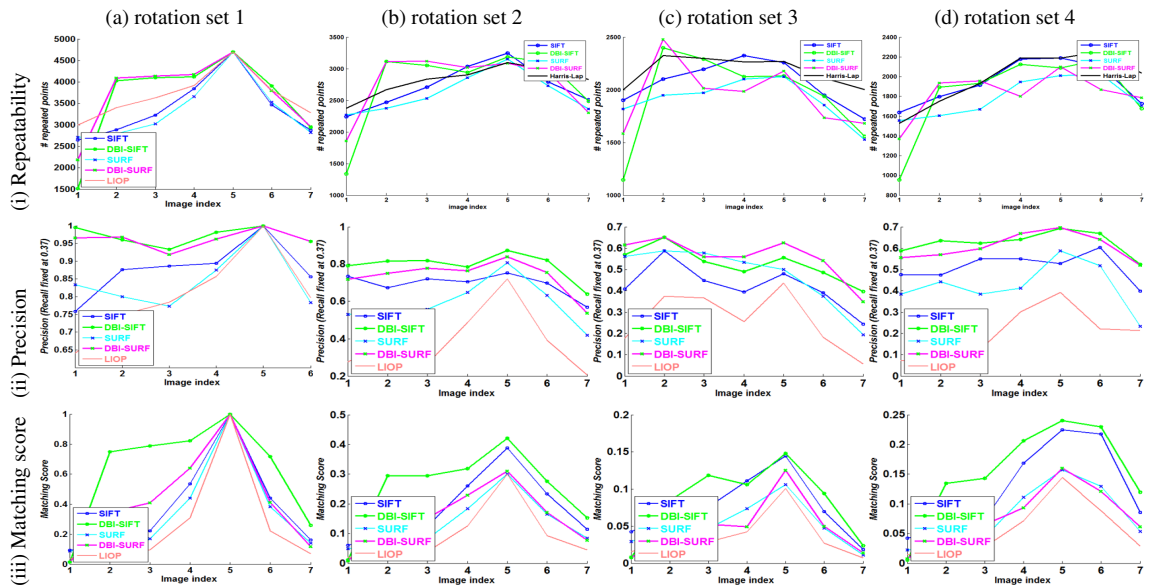

Figure 6.12: Evaluation of detectors and descriptors under the influence of rotation and defocus blur. The curves represent the relation between the features of blurred reference image (yellow bounding box in Figure 6.10) and the target images from all rotation/blur settings. The plots in the second row are taken from precision-recall curve, where the recall value fixed to 0.37.
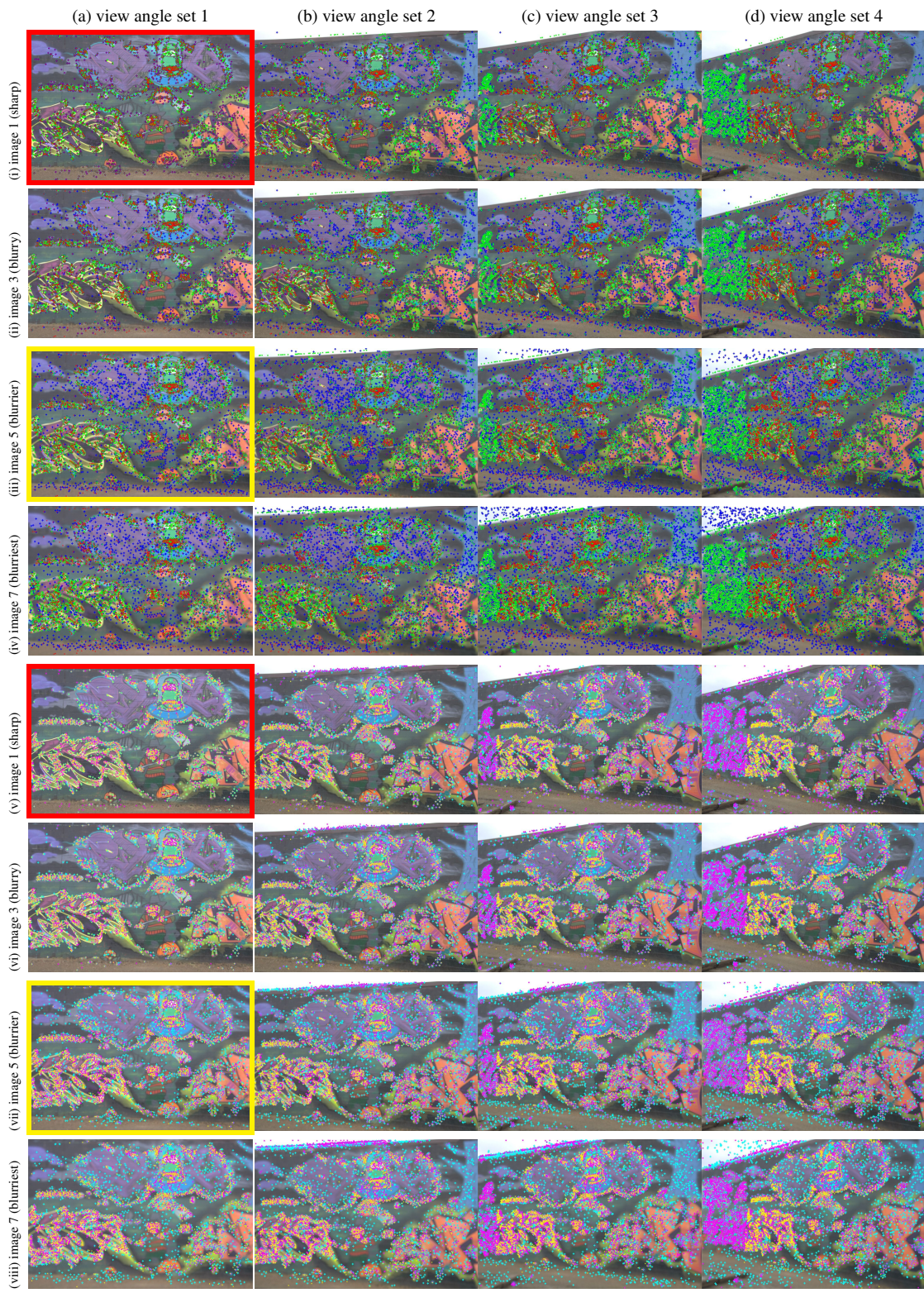
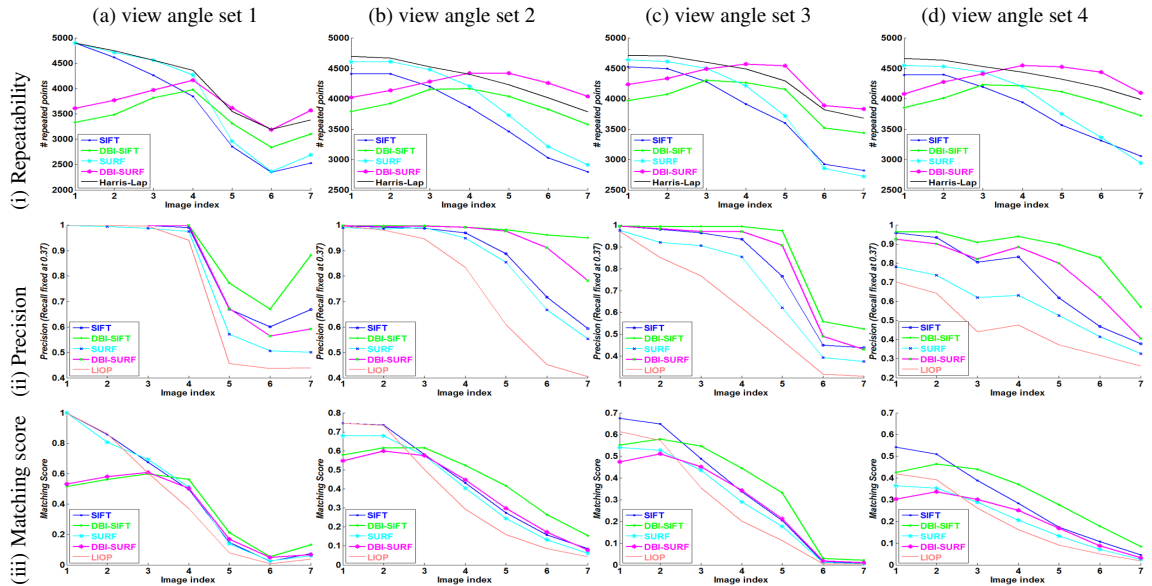Figure 6.13: Stability of interest point detection under the influence of view angle and defocus blur.

Figure 6.14: Evaluation of detectors and descriptors under the influence of view angle and defocus blur. The curves represent the relation between the features of sharp reference image (red bounding box in Figure 6.13) and the target images from all view angle/blur settings. The plots in the second row are taken from precision-recall curve, where the recall value fixed to 0.37.
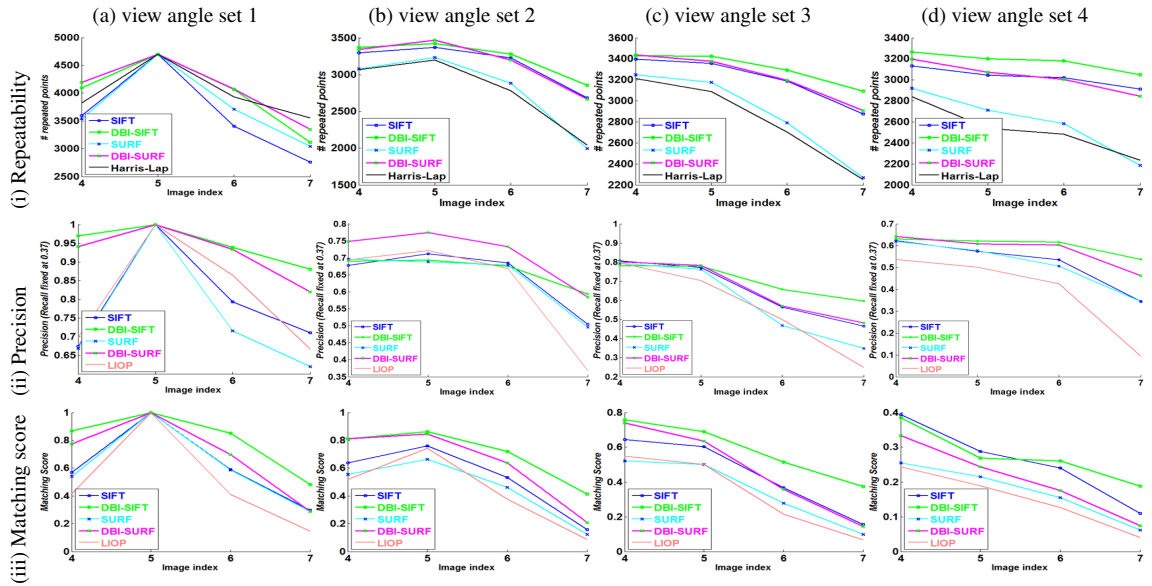

Figure 6.15: Evaluation of detectors and descriptors under the influence of view angle and defocus blur. The curves represent the relation between the features of blurred reference image (yellow bounding box in Figure 6.13) and the target images from all view angle/blur settings. The plots in the second row are taken from precision-recall curve, where the recall value fixed to 0.37.

Figure 6.16: Stability of interest point detection under the influence of view angle and defocus blur.

51

Figure 6.17: Evaluation of detectors and descriptors under the influence of view angle and defocus blur. The curves represent the relation between the features of sharp reference image (red bounding box in Figure 6.16) and the target images from all view angle/blur settings. The plots in the second row are taken from precision-recall curve, where the recall value fixed to 0.37.
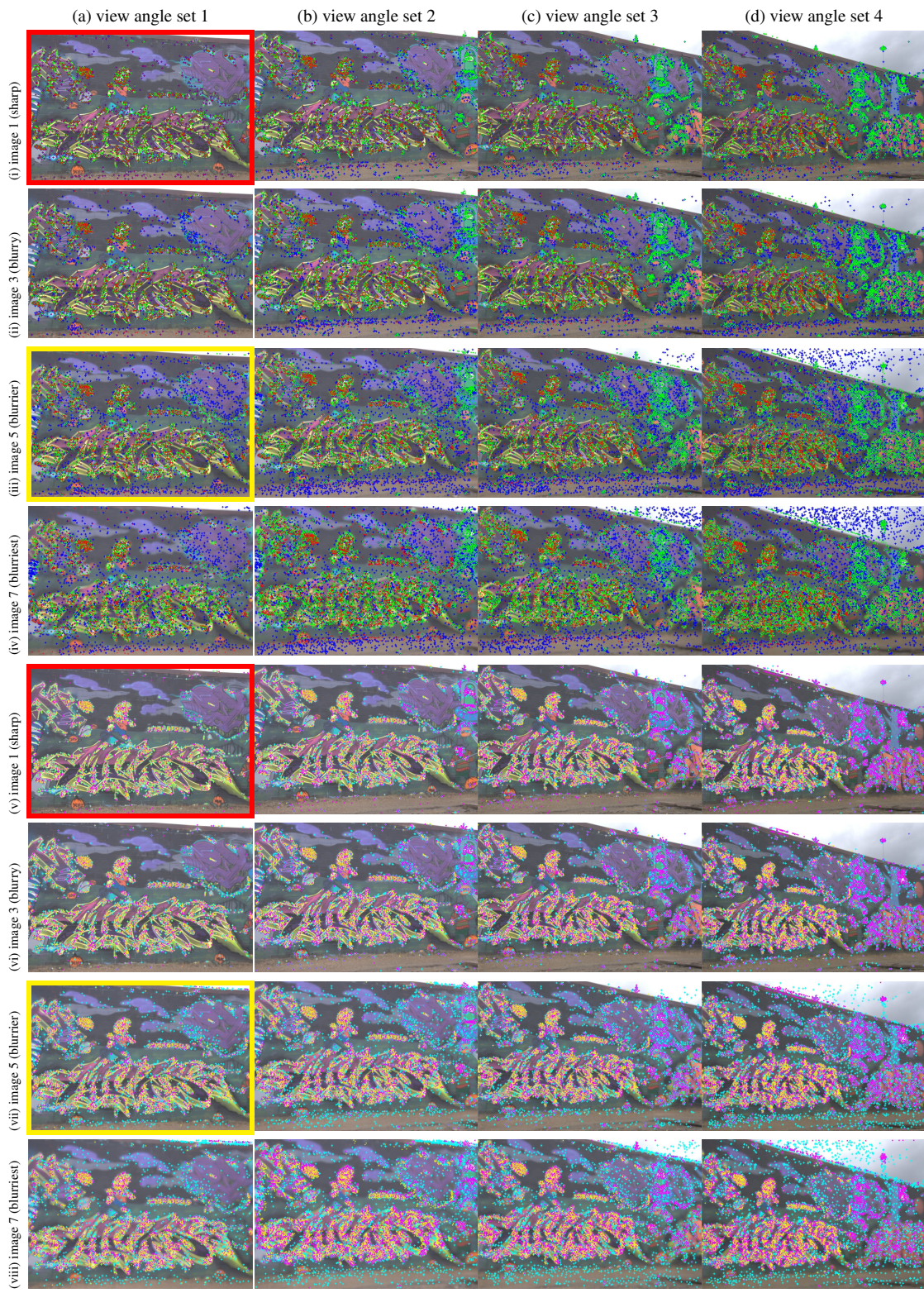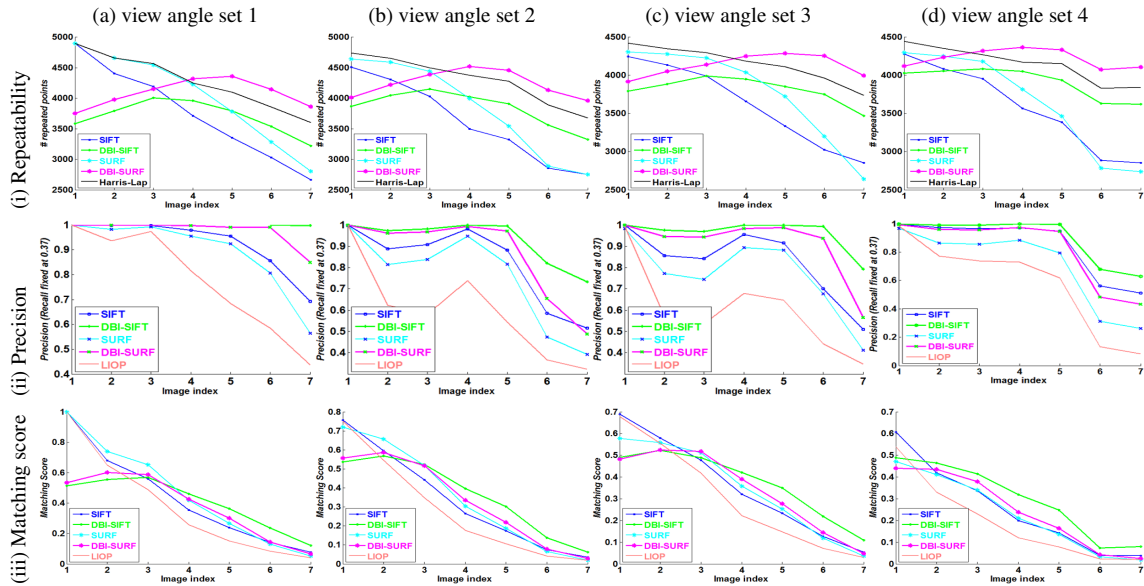
Experimental results from real images showed that the proposed DBI-SIFT and DBI-SURF detectors have repeatability much stable than regular SIFT and SURF across various degrees of photometric and geometric transformation combined with defocus blur. The proposed DBI-SIFT and DBI-SURF descriptors have robustness and distinctiveness much higher than SIFT and SURF descriptors when they extracting features from blurred images. The applications of proposed defocus blur-invariant scale-space feature extraction include detection, recognition, and tracking of blurred objects in a scene.

Acknowledgment

BIBLIOGRAPHY

[1] T. Lindeberg, "Feature detection with automatic scale selection," *International journal of computer vision*, vol. 30, no. 2, pp. 79–116, 1998.

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, pp. 91–110, 2004.

[3] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Computer Vision–ECCV 2002*. Springer, 2002, pp. 128–142.

[4] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer Vision–ECCV 2006*. Springer, 2006, pp. 404–417.

[5] C. Harris and M. Stephens, "A combined corner and edge detector." in *Alvey vision conference*, vol. 15. Manchester, UK, 1988, p. 50.

[6] Z. Wang, B. Fan, and F. Wu, "Local intensity order pattern for feature description," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 603–610.

[7] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 10, pp. 1615–1630, 2005.

[8] H. Jin, P. Favaro, and R. Cipolla, "An analysis of scale-space sampling in shift," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 18–25.

[9] P. N. Sabes and M. I. Jordan, "Advances in neural information processing systems," in *In G. Tesauro & D. Touretzky & T. Leed (Eds.), Advances in Neural Information Processing Systems*. Citeseer, 1995.

[10] P. S. H. B. K. . F. W. T. Cho, T. S., "Blur kernel estimation using the radon transform." IEEE, 2011, pp. 241–248.

[11] H. Jin, P. Favaro, and R. Cipolla, "Visual tracking in the presence of motion blur," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 18–25.

[12] A. Pretto, E. Menegatti, M. Bennewitz, W. Burgard, and E. Pagello, "A visual odometry framework robust to motion blur," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE, 2009, pp. 2250–2257.

[13] Y. Wu, H. Ling, J. Yu, F. Li, X. Mei, and E. Cheng, "Blurred target tracking by blur-driven tracker," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1100–1107.

[14] T. S. Cho, S. Paris, B. K. Horn, and W. T. Freeman, "Blur kernel estimation using the radon transform," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 241–248.

[15] H. Cai, X. Wang, M. Xia, and Y. Wang, "Entropy-based maximally stable extremal regions for robust feature detection," *Mathematical Problems in Engineering*, vol. 2012, 2012.

[16] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing*, vol. 22, no. 10, pp. 761–767, 2004.

[17] Y. Zhang and K. Hirakawa, "Blur processing using double discrete wavelet transform," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 1091–1098.

[18] H. Cai, X. Wang, M. Xia, and Y. Wang, "Evaluating performance of feature extraction methods for practical 3d imaging systems," *Proceedings of the 27th Conference on Image and Vision Computing New Zealand*, 2012.

[19] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," *Foundations and Trends® in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008.

[20] S. Ravela and R. Manmatha, "Gaussian filtered representations of images," *Wiley Encyclopedia of Electrical and Electronics Engineering*, 1999.

[21] T. Hassner, V. Mayzels, and L. Zelnik-Manor, "On sifts and their scales," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1522–1528.

[22] K. Hirakawa and T. W. Parks, "Adaptive homogeneity-directed demosaicing algorithm," *Image Processing, IEEE Transactions on*, vol. 14, no. 3, pp. 360–369, 2005.

[23] G. Buchsbaum, "A spatial processor model for object colour perception," *journal of the Franklin institute*, vol. 310, no. 1, pp. 1–26, 1980.

[24] C. Silpa-Anan and R. Hartley, "Optimised kd-trees for fast image descriptor matching," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[25] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration." in *VISAPP (1)*, 2009, pp. 331–340.

[26] H. Jegou, M. Douze, and C. Schmid, "Product quantization for nearest neighbor search," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 1, pp. 117–128, 2011.

[27] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," 2008.