DO WE REALLY NEED GAUSSIAN FILTERS FOR FEATURE POINT DETECTION?

Lee-kang Liu[*]*, Truong Nguyen*

University of California, San Diego
http://videoprocessing.ucsd.edu

Stanley H. Chan[†]

School of Engineering and Applied Science
Harvard University

ABSTRACT

This paper studies the issue of which filters should be used for feature point detection. Classical feature point detection methods, e.g., SIFT, are based on the scale-space theory in which Gaussian filters are proven to be optimal under the scale-space axiom. However, the recent method SURF demonstrates empirically that a box filter can also achieve good performance even though it violates the scale-space axiom. This leads to the question: Is Gaussian filters necessary for feature point detection? Based on the analysis using filter bank and detection theory, we show that theoretically it is possible for a box filter to perform better than the Gaussian filter. Additionally, we show that a new filter, pyramid filter, performs better than both box and Gaussian filters in some situations.

***Index Terms*—** feature point detection, SIFT, SURF

1. INTRODUCTION

1.1. Feature Extraction

Extracting feature points of an image is a fundamental problem in computer vision. Good feature points allow one to perform correspondence match across different viewing angles, hence making it possible to extract camera parameters, calibrate cameras, rectify, interpolate-extrapolate and render images, etc. Features of interests depend on the application and the availability of a mathematical model. Common examples include edges, texture and blobs.

Feature extraction is a well-studied subject under the framework of multiscale image representation, also known as the scale-space theory [1]. Feature extraction consists of two major components: *detection*, which aims at finding the location of a feature, and *description*, which aims at finding a set of labels to describe the feature. The focus of this paper is the detection part.

[*]Corresponding author: L. Liu, email: l7liu@ucsd.edu. This work is supported in part by NSF grant CCF-1065305.

[†]S. Chan performed the work while at UC San Diego.

1.2. Scale-space Representation

To detect a feature point, we form a scale-space volume [2]

$$L(x, y, \sigma) = g(x, y; \sigma) * I(x, y), \tag{1}$$

where $I(x, y)$ is the input image, $g(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$ is a Gaussian filter, and "$*$" denotes convolution. A precise definition of a feature point depends on the application. In this paper, we focus on detecting blobs. In this case, we define a feature point as the local minimum of the Hessian determinant

$$\det\mathcal{HL}(x, y, \sigma) = L_{xx}L_{yy} - L_{xy}L_{yx}, \tag{2}$$

where $L_{xx}, L_{yy}, L_{xy}, L_{yx}$ are the second-order directional derivatives. In literature, Gaussian filters are used in (1) because they optimal under the scale-space axioms [3]. However, convolution with a Gaussian filter is computationally intensive. Thus, Bay *et al*. [4] proposed a zero-order approximation, known as SURF, and later, Hussein *et al*. [5] generalized the idea to higher order approximations. Interestingly, their experimental results show that the non-Gaussian filters perform not much worse, sometimes even better, than the Gaussian filters. This leads to the question: except for satisfying the scale-space axioms, what are reasons of using Gaussian filters? Or simply, do we really need Gaussian filters for feature detection?

1.3. Objectives

The objective of this paper is to study the performance of feature detection algorithms. We show the following results:

Unified Framework: Popular feature detection methods, SIFT [2] and SURF [4], can be generalized under the proposed filter bank framework.

Detection Analysis: Performance of a feature detection method depends on the match between the object and the filter. Therefore, although Gaussian filters are *usually* good due to the intrinsic smoothness of natural images, there are failure cases.

Pyramid Approximation: Gaussian filter has a high computational complexity. We propose an approximation scheme that preserves properties of a Gaussian filter, but achieves similar complexity as the box filter.

1.4. Notation

For notational simplicity we consider the analysis in one-dimension, although the results are directly applicable to higher dimensions. Input signal is denoted by $I(t)$, and the scale-space volume is denoted by $L(t,\sigma)$. First and second derivatives of $L(t,\sigma)$ along the t-direction is given by $L'(t,\sigma)$ and $L''(t,\sigma)$, respectively. Thus, the Hessian determinant is $\det\mathcal{HL}(t,\sigma) = L''(t,\sigma)$. The one-dimensional Gaussian filter is denoted by $g(t;\sigma)$.

2. FILTER BANK FRAMEWORK

In this section we present a common framework of SIFT [2] and SURF [4] for computing $\det\mathcal{HL}(t;\sigma)$.

SIFT and SURF can be summarized as

$$\text{SIFT:}\quad L''(t,\sigma) = \sigma^2 g''(t;\sigma) * I(t), \tag{3}$$

$$\text{SURF:}\quad L''(t,\sigma) = \mathcal{T}[\sigma^2 g''](t;\sigma) * I(t), \tag{4}$$

where $\mathcal{T}$ a non-linear function defined as $\mathcal{T} : C^\infty(\mathbb{R}) \to C^1(\mathbb{R})$

$$\mathcal{T}[\sigma^2 g''](t;\sigma) = \begin{cases} 1, & \sigma^2 g''(t;\sigma) \geq 0, \\ -2, & \sigma^2 g''(t;\sigma) < 0, \end{cases} \tag{5}$$

and σ^2 is the γ-normalization factor [1]. Note that (4) becomes (3) when $\mathcal{T}$ is the identity operator.

In practice, (3) and (4) are evaluated using the difference of Gaussian (DoG) method [2]. DoG states that for any $\sigma \in [\sigma_k, \sigma_{k+1}]$,

$$\sigma^2 g''(t;\sigma) \approx g(t;\sigma_{k+1}) - g(t;\sigma_k).$$

Given a sequence $\{\sigma_k\}_{k=0}^n$ where $\sigma_0 = 0$, we define $\Delta\sigma_k = \sigma_{k+1} - \sigma_k$. Then using the fact that $g(t;\sigma_{k+1}) = g(t;\sigma_k) * g(t;\Delta\sigma_k)$, the filters in (3) and (4) can be realized using a filter bank structure shown in Fig. 1.

Three important observations can be drawn from the filter bank structure. First, ignoring the non-linear function $\mathcal{T}$, the filter bank is a *perfect reconstruction* system, because the summation of all output channels is the dirac delta function. Second, in the kth stage, $g(t;\sigma_k)$ and $\delta(t) - g(t;\sigma_k)$ are lowpass and highpass filters, respectively. Therefore, except for the top and bottom channel, all other outputs are bandpass filters with bandwidth controlled by the closeness between adjacent σ_k's. Third, the frequency response of a Gaussian filter is Gaussian.

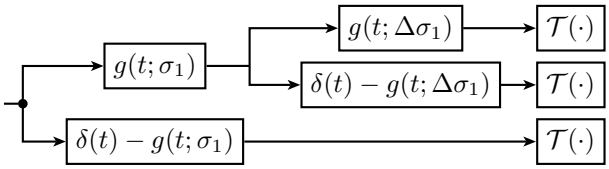


Fig. 1. Filter bank implementation of the filters in SIFT [2] (with $\mathcal{T} = 1$) and SURF [4] (with $\mathcal{T}$ defined in (5)). This figure shows a two-stage example.

Thus a Gaussian filter is non-negative and monotonically decreasing. However, when $\mathcal{T}$ is used, the frequency response becomes a sinc function, which violates the non-negativity and non-increasing condition of scale-space axiom.

3. DETECTION ANALYSIS

From the framework discussed in Section 2, we observe that the box filters of SURF are approximations of the Gaussian filters of SIFT. Historically, Gaussian filters are preferred because it is the necessary and sufficient condition for the scale-space axiom. However, if our goal is only to detect the correct feature location and not to accommodate the scale-space axiom, then Gaussian filters may not be the best filter. We show a counterexample in this section to verify our claims.

To start with, we consider an input signal

$$I(t) = \begin{cases} 1, & -T \leq t \leq T, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

$I(t)$ is a one-dimensional box function centered at the origin. Therefore, the feature point should be ideally located at the origin. Now, suppose that white Gaussian noise $n(t) \sim \mathcal{N}(0,\sigma_N^2)$ is added to the input signal so that the signal becomes $s(t) = I(t) + n(t)$. The new Hessian determinants are

$$G(t,\sigma) \stackrel{\text{def}}{=} \sigma^2 g''(t;\sigma) * I(t) + \sigma^2 g''(t;\sigma) * n(t),$$

$$H(t,\sigma) \stackrel{\text{def}}{=} h''(t;\sigma) * I(t) + h''(t;\sigma) * n(t),$$

where $h''(t;\sigma) = \mathcal{T}[\sigma^2 g''](t;\sigma)$, with $\mathcal{T}$ defined in (5).

3.1. Analysis for Gaussian Filters (SIFT)

The following result gives the statistical description of the Hessian determinant using Gaussian filter in the presence of noise.

Proposition 1. *Let $v(t,\sigma) = \sigma^2 g''(t;\sigma) * I(t)$ and $w(t,\sigma) = \sigma^2 g''(t;\sigma) * n(t)$, it holds that*

$$v(t,\sigma) = \frac{1}{\sqrt{2\pi}}\left[-\frac{t+T}{\sigma}e^{-\frac{(t+T)^2}{2\sigma^2}} + \frac{t-T}{\sigma}e^{-\frac{(t-T)^2}{2\sigma^2}}\right].$$

The mean and autocorrelations (over t) of $w(t,\sigma)$ *are* $\mathbb{E}[w(t,\sigma)] = 0$ *and* $\mathbb{E}[w(t,\sigma)w(t+\tau,\sigma)] = \sigma^4\sigma_N^2 g''(-\tau;\sigma) * g''(\tau;\sigma)$, *respectively. Hence, the variance of* $w(t,\sigma)$ *is*

$$Var\left[w(t,\sigma)\right] = \frac{3\sigma_N^2}{8\sqrt{\pi}\sigma}. \tag{7}$$

Proof. The deterministic part $v(t) = \sigma^2 g''(t;\sigma) * I(t)$ can be shown through direct substitution. For the noise part, since convolution and expectation are both linear, they are interchangeable. Thus $\mathbb{E}[n(t)] = 0$ implies $\mathbb{E}[w(t,\sigma)] = 0$. The autocorrelation is a standard result [6]. The variance is found by evaluating the autocorrelation at $\tau = 0$, i.e., Var $[w(t,\sigma)] = \sigma^4\sigma_N^2 g''(-\tau;\sigma) * g''(\tau;\sigma)|_{\tau=0}$. □

Since ideally the feature point is at the origin, we say that the feature is detected correctly if $G(0,\sigma)$ is a local minimum, i.e., for any $\epsilon > 0$, $G(0,\sigma) \leq G(t,\sigma)$ whenever $|t| \leq \epsilon$. Consequently, we show the following corollary.

Corollary 1. *For* $I(t)$ *given by (6) and using a Gaussian filter, the probability of correct detection is characterized by the random variable* $z(t,\sigma) = G(0,\sigma) - G(t,\sigma)$, *where*

$$z(t,\sigma) \sim \mathcal{N}\left(v(0,\sigma) - v(t,\sigma), \frac{3\sigma_N^2}{4\sqrt{\pi}\sigma}\right). \tag{8}$$

Proof. The probability of correct detection is $P[G(0,\sigma) \leq G(t,\sigma)]$, or $P[G(0,\sigma) - G(t,\sigma) \leq 0]$. Let $z(t,\sigma) = G(0,\sigma) - G(t,\sigma)$, it holds that the mean of $z(t,\sigma)$ is $v(0,\sigma) - v(t,\sigma)$, and the variance of $z(t,\sigma)$ is twice of $\frac{3\sigma_N^2}{8\sqrt{\pi}\sigma}$. □

Corollary 1 provides a quantitative argument for the comparison of SIFT and SURF which will be discussed in Section 3.3. Before moving to the SURF case, we derive one addition result to link T and σ.

Proposition 2. *Given* T, $v(0,\sigma)$ *is a local minimum of* $v(t,\sigma)$ *if and only if* $\sigma > T/\sqrt{3}$. *Furthermore, the optimal scaling parameter* σ^* *for which* $\partial v(0,\sigma)/\partial\sigma = 0$ *is* $\sigma^* = T$.

Proof. The ideal location of the minimum is at the origin. Thus $v(0,\sigma)$ must be a local minimum for $v(t,\sigma)$. By the first and second order optimality conditions, $v(0,\sigma)$ is a local minimum if and only if $\frac{\partial v(t,\sigma)}{\partial t}|_{t=0} = 0$ and $\frac{\partial^2 v(t.\sigma)}{\partial t^2}|_{t=0} > 0$. Through some calculation it can be shown that these are equivalent to requiring $\sigma > T/\sqrt{3}$. The second statement can be proved by taking $\frac{\partial v(0,\sigma)}{\partial\sigma}$=0. □

Fig. 2 illustrates the result of Proposition 2. This plot shows the function $v(t,\sigma)$ with magnitude indicated by the color. In this plot $T = 15$, and so by Proposition 2 $v(0,\sigma)$ is a local minimum (along the t-dimension) iff $\sigma > T/\sqrt{3} \approx 8.6$. Additionally, $v(0,\sigma)$ achieves minimum at $\sigma = T$.

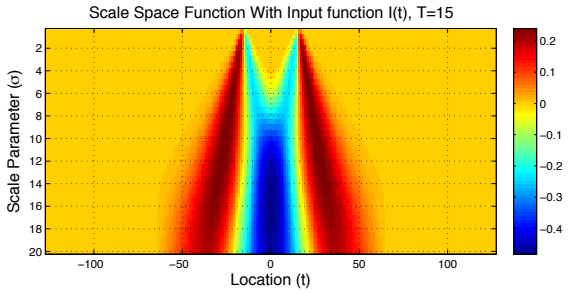


Fig. 2. This plot shows that along the t-dimension, $v(0,\sigma)$ is a local minimum of $v(t,\sigma)$ if $\sigma > 8.6$, and along the σ-dimension, $v(0,\sigma^*)$ is a local minimum of $v(0,\sigma)$ if $\sigma^* = 15$.

3.2. Analysis for Box Filters (SURF)

A similar result of (8) can be shown for the case of SURF. The filter used in SURF is a zero-order approximation of $g''(t;\sigma)$. Following [4], $h''(t;\sigma)$ is defined as

$$h''(t;\sigma) = \begin{cases} -2C/6, & |t| \leq C, \\ C/6, & C < |t| \leq 3C, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

The factor C is a function depending on σ and it controls the width of the filter [4]. The precise relationship between C and σ is given by the following result (analogous to Prop. 2).

Proposition 3. *Given* σ, *and let*

$$C^* = \underset{C}{argmin} \int_{-\infty}^{\infty} \left[\sigma^2 g''(t;\sigma) - h''(t;\sigma)\right]^2 dt,$$

then $C^* = \alpha\sigma$ *where* $\alpha \approx 0.91$ *is a constant*.

The next two results are analogous to Proposition 1 and Corollary 1, but for the case of $h''(t;\sigma)$.

Proposition 4. *If* $v(t,\sigma) = h''(t;\sigma) * I(t)$ *and* $w(t) = h''(t;\sigma) * n(t)$, *then Var* $[w(t,\sigma)] = \frac{\sigma_N^2}{3C}$ *and*

$$v(t,\sigma) = \begin{cases} -\frac{2}{3}, & t = 0, \\ -\frac{2}{3} - \frac{t}{2C}, & 0 < |t| \leq 2C, \\ \frac{1}{3} + \frac{t}{6C}, & 2C < |t| \leq 4C. \end{cases}$$

Corollary 2. *For* $I(t)$ *given by (6) and using* $h''(t;\sigma)$, *the probability of correct detection is characterized by the random variable* $z(t,\sigma) = H(0,\sigma) - H(t,\sigma)$, *where*

$$z(t,\sigma) \sim \mathcal{N}\left(v(0,\sigma) - v(t,\sigma), \frac{\sigma_N^2}{3C}\right). \tag{10}$$

3.3. Discussion

Corollaries 1 and 2 suggest a quantitative comparison between SIFT and SURF. Shown in Fig. 3 is the probability of correct detection for SIFT and SURF. Both methods show reduced performance when noise increases, implying that the local minimum becomes less likely to be located at the origin. Comparing SIFT and SURF, SURF has higher correct detection probability in all cases. This implies that Gaussian filters are *not* always the best filter for feature detection.

We want to put a remark here that the above analysis is about the distortion of noise. There are other aspects which should be further studied. The first aspect is the distortion caused by affine transformation. The second aspect is the presence of similar patterns in the image [7].

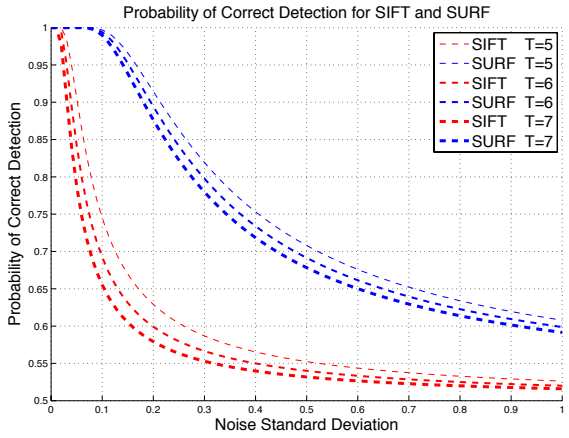


Fig. 3. Statistical comparisons between SIFT and SURF. This figure shows the probability of correct detection as a function of noise variance. T is the width of the signal defined in (6).

4. PYRAMID FILTER

To further illustrate our claim that Gaussian filters are not always the best filters in feature detection, we use the pyramid filter (also known as the triangle filter) and show that it achieves higher repeatability in some real images.

4.1. Derivation and Implementation

Pyramid filters are the first order approximations of the second derivatives of Gaussian filters. In the one-dimensional case, a pyramid filter $p''(t;\sigma)$ is defined by setting $\mathcal{T}$ as

$$p''(t;\sigma) = \mathcal{T}[\sigma^2 g''](t;\sigma) = \begin{cases} \frac{2t}{C} - 2, & 0 \le t \le C, \\ \frac{t}{C} - 1, & C \le t \le 2C, \\ -\frac{t}{C} + 3, & 2C \le t \le 3C, \end{cases}$$

where C is defined in (9), and $\mathcal{T}[\sigma^2 g''](t;\sigma) = \mathcal{T}[\sigma^2 g''](-t;\sigma)$. An illustration of $p''(t;\sigma)$ is shown in Fig. 4.

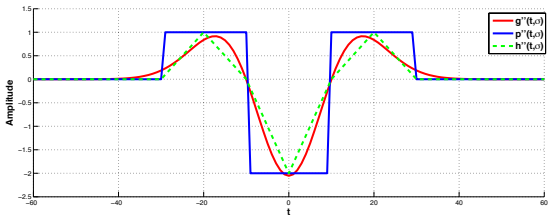


Fig. 4. Plots of $g''(t;\sigma)$, $h''(t;\sigma)$, and $p''(t;\sigma)$. Curves are scaled for visualization.

In the two-dimensional case, a pyramid filter is constructed by convolving the box filters, as shown in Fig. 5. The convolution $P(x,y,\sigma) = I(x,y) * p''(x,y;\sigma)$ is calculated using repeated integration [8], and the concept of moment integral images [9].

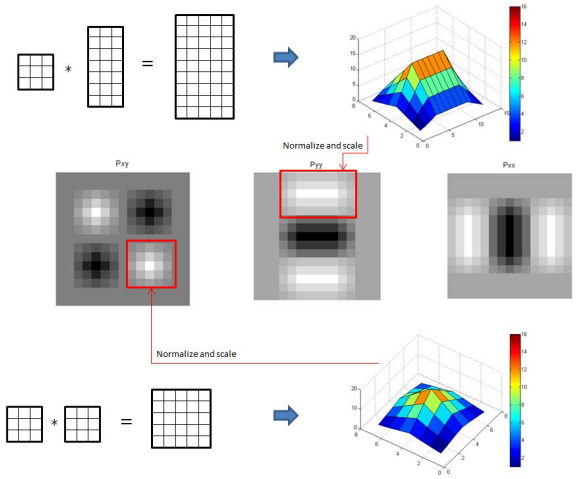


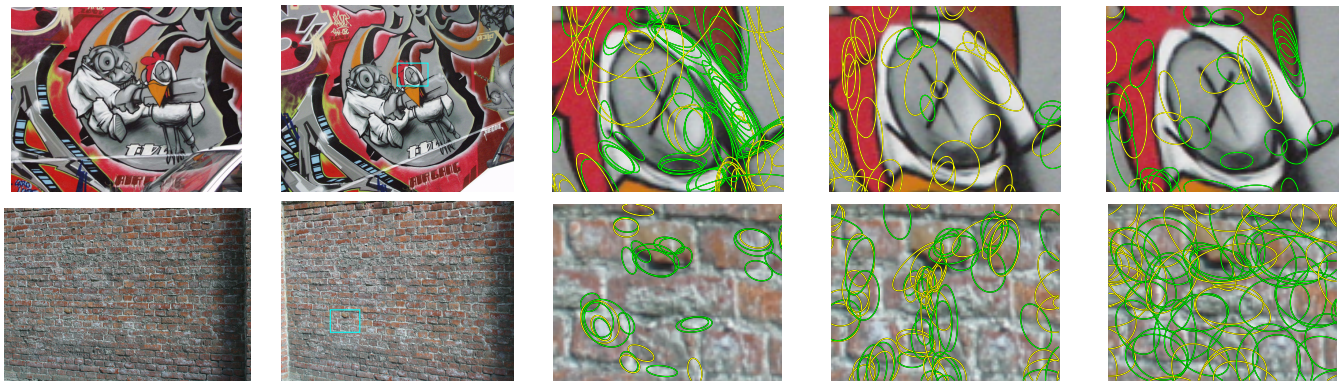
Fig. 5. Constructing the pyramid filters. There are two types of pyramid filters: rectangular and square shapes. Both can be constructed by convolving the box filters.

4.2. Evaluation

We consider two examples with homographies (viewpoint difference) in the dataset (totally 12 images), available at [10] `http://www.robots.ox.ac.uk/~vgg/research/affine/`. In the test, the repeatability score, define as [11]

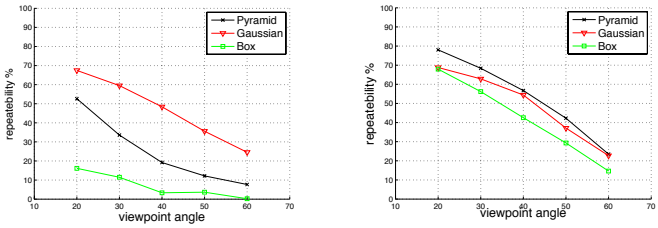
$$R = \frac{\text{number of correspondences}}{\text{minimum number of points detected}}$$

is calculated. Two points $\mathbf{x}_1 = (x_1, y_1)$ and $\mathbf{x}_2 = (x_2, y_2)$ are considered correspond if [11]



(a) Reference (b)Test (c) Gaussian (d) Box (e) Pyramid

Fig. 6. Feature points detected using different filters. The third to fifth columns are zoom in regions of test image. Green ellipses are features that have corresponding features in reference image and the yellow ellipses are features that have no corresponding features in reference image.



(a) Image 1 (b) Image 2

Fig. 7. Repeatability scores of the two images.

1. $\|\mathbf{x}_1 - \mathbf{H}\mathbf{x}_2\|_2 \leq 1.5$, where $\mathbf{H}$ is a (known) linear transformation that defines the homography,

2. The surface error for affine regions is less than a threshold, which, roughly speaking, is the amount of overlap between two ellipses - one being a region defined by $L(x_1, y_1, \sigma)$ and the other one being a region defined by $L(x_2, y_2, \sigma)$ where $L(x, y, \sigma)$ can be $G(x, y, \sigma)$, $H(x, y, \sigma)$, or $P(x, y, \sigma)$.

5. CONCLUSION

Gaussian filters are not necessary for feature point detection. Instead, the successfulness of feature point detection depends on the match between the shape of the filter and that of the signal. We showed, in particular, a box signal in the presence of noise can be better detected using a box filter than the Gaussian filter. We also proposed a pyramid approximation of the Gaussian filter to yield better detection rate than the box filter while keeping a low complexity.

One question remains open is what kinds of features are more common in an image. While not proved, we believe that in natural images there are more features favorable to Gaussian filters than the others. This claim requires further analysis.

6. APPENDIX

Detailed proofs and additional results of this paper are available at `http://videoprocessing.ucsd.edu/~LeeKang/Research/Supplementary.pdf`.

7. REFERENCES

[1] T. Lindeberg, "Scale-space," *Encyclopedia of Computer Science and Engineering*, vol. 6, pp. 2495–2504, 2009.

[2] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 3, pp. 91–110, 2004.

[3] J. Babaud, A. Witkin, M. Baudin, and R. Duda, "Uniqueness of the gaussian kernel for scale-space filtering," *IEEE PAMI*, vol. 8, no. 1, pp. 26–33, 1986.

[4] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[5] M. Hussein, F. Porikli, and L. Davis, "Kernel integral images: A framework for fast non-uniform filtering," in *CVPR*, 2008.

[6] H. Stark and J. Woods, *Probability and Random Processes with Applications to Signal Processing*, Prentice Hall, 2001.

[7] J. Fan, Y. Wu, and S. Dai, "Discriminative spatial attention for robust tracking," 2010, pp. 480–493.

[8] P. Heckbert, "Filtering by repeated integration," in *SIGGRAPH*, 1986.

[9] A. Haselhoff and A. Kummert, "A vehicle detection system based on haar and triangle features," in *IEEE Intelligent Vehicles Symposium*, 2009, pp. 261–266.

[10] K. Mikolajczyk, T. Tuytelaars, and C. Schmid et al., "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1, pp. 43–72, 2005.

[11] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.