

Constant-time filtering using shiftable kernels

Kunal Narayan Chaudhury

Abstract

It was recently demonstrated in [1] that the non-linear bilateral filter [2] can be efficiently implemented using a constant-time or $O(1)$ algorithm. At the heart of this algorithm was the idea of approximating the Gaussian range kernel of the bilateral filter using trigonometric functions. In this letter, we explain how the idea in [1] can be extended to few other linear and non-linear filters [2], [3], [4]. While some of these filters have received a lot of attention in recent years, they are known to be computationally intensive. To extend the idea in [1], we identify a central property of trigonometric functions, called *shiftable*, that allows us to exploit the redundancy inherent in the filtering operations. In particular, using shiftable kernels, we show how certain complex filtering can be reduced to simply that of computing the *moving sum* of a stack of images. Each image in the stack is obtained through an elementary pointwise transform of the input image. This has a two-fold advantage. First, we can use fast recursive algorithms for computing the moving sum [5], [6], and, secondly, we can use parallel computation to further speed up the computation. We also show how shiftable kernels can also be used to approximate the (non-shiftable) Gaussian kernel that is ubiquitously used in image filtering.

Index Terms

Filtering, shiftable, kernel, $O(1)$ complexity, approximation, constant-time algorithm, moving sum, neighborhood filter, spatial filter, bilateral filter, non-local means.

Constant-time filtering using shiftable kernels

I. INTRODUCTION

The function $\cos(x)$ has the remarkable property that, for any translation τ , $\cos(x - \tau) = \cos(\tau)\cos(x) + \sin(\tau)\sin(x)$. That is, we can express the translate of a sinusoid as the linear combination of two fixed sinusoids. More generally, this holds true for any function of the form $\varphi(x) = c_1 \exp(\alpha_1 x) + \dots + c_N \exp(\alpha_N x)$. This follows from the addition-multiplication property of the exponential. As special cases, we have the pure exponentials when α_i is real, and the trigonometric functions when α_i is imaginary.

The translation of not all functions can be written in this way, e.g., consider the functions $\exp(-x^2)$ and $(1+x^2)^{-1}$. The other class of functions that have this property are the polynomials, $\varphi(x) = c_0 + c_1 x + \dots + c_N x^N$. Functions with this property can be realized in higher dimensions using higher-dimensional polynomials and exponentials, or simply by taking the tensor product of the one-dimensional functions.

More generally, we say that a function $\varphi(\mathbf{x})$ is *shiftable* in \mathbf{R}^d if there exists a fixed (finite) collection of functions $\varphi_1(\mathbf{x}), \dots, \varphi_N(\mathbf{x})$ such that, for every translation τ in \mathbf{R}^d , we can write

$$\varphi(\mathbf{x} - \tau) = c_1(\tau)\varphi_1(\mathbf{x}) + \dots + c_N(\tau)\varphi_N(\mathbf{x}). \quad (1)$$

We call the fixed functions $\varphi_1(\mathbf{x}), \dots, \varphi_N(\mathbf{x})$ the *basis functions*, $c_1(\tau), \dots, c_N(\tau)$ the *interpolating coefficients*, and N the *order* of shiftability. Note that the coefficients depend on τ , and are responsible for capturing the translation action.

Shiftable functions, and more generally, steerable functions [7], have a long history in signal and image processing. Over the last few decades, researchers have found applications of these special functions in various domains such as image analysis [8], [9] and motion estimation [10], and pattern recognition [11], to name a few. In several of these applications, the role of steerability and its formal connection with the theory of Lie groups was only recognized much later. We refer the readers to the exposition of Teo [12] for an excellent account on the theory and practice of steerable functions.

In a recent paper [1], we showed how specialized trigonometric kernels could be used for fast bilateral filtering. This work was inspired by the work of Porikli [13], who had earlier shown how polynomials could be used for the same purpose. We now realize that it is the shiftability of the kernel that is at the heart of the matter, and that this can be applied to other forms of filtering and

using more general kernels. We will provide a general theory of this in Section II, where we also propose some algorithms that have a constant-time complexity per pixel, independent of the size of the filtering kernels. To the best of our knowledge, such algorithms have not been reported in the community in its full generality. The problem of designing shiftable kernels is addressed in Section III. Here we also propose a scheme for approximating the ubiquitous Gaussian kernel using shiftable functions. Finally, in Section IV, we present some thoughts on how shiftable functions could be used for reducing the complexity of the non-local means filter [4].

II. FILTERING USING MOVING SUMS

We now show how certain constant-time algorithms for image filtering can be obtained using shiftable kernels. The idea is that by using shiftable functions, we can decompose the local kernels (obtained using translations) in terms of “global” functions – the basis functions. The local information gets encoded in the interpolating coefficients. This allows us to explicitly take advantage of the redundancy in the filtering operation.

To begin with, we consider the simplest neighborhood filter, namely the spatial filter. This is given by

$$\bar{f}(\mathbf{x}) = \frac{1}{\eta} \int_{\Omega} \varphi(\mathbf{y}) f(\mathbf{x} - \mathbf{y}) d\mathbf{y} \quad (2)$$

where $\eta = \int_{\Omega} \varphi(\mathbf{y}) d\mathbf{y}$. Here $\varphi(\mathbf{x})$ is the kernel, and Ω is the neighborhood over which the integral (sum) is taken. Note that, henceforth, we will use the term *kernel* to specifically mean that the function is symmetric, non-negative, and unimodal over its support (peak at the origin). It is not immediately clear that one can construct such kernels using shiftable functions. We will address this problem in the sequel.

For the moment, suppose that the kernel $\varphi(\mathbf{x})$ is shiftable, so that (1) holds. We use this, along with symmetry, to center the kernel at \mathbf{x} . In particular, we write $\varphi(\mathbf{y}) = \varphi(\mathbf{x} - \mathbf{y} - \mathbf{x}) = \sum_{n=1}^N c_n(\mathbf{x}) \varphi_n(\mathbf{x} - \mathbf{y})$. Using linearity, we have

$$\int_{\Omega} \varphi(\mathbf{y}) f(\mathbf{x} - \mathbf{y}) d\mathbf{y} = \sum_{n=1}^N c_n(\mathbf{x}) \int_{\Omega} \varphi_n(\mathbf{x} - \mathbf{y}) f(\mathbf{x} - \mathbf{y}) d\mathbf{y}.$$

Similarly, $\eta = \sum_{n=1}^N c_n(\mathbf{x}) \int_{\Omega} \varphi_n(\mathbf{x} - \mathbf{y}) d\mathbf{y}$.

Now consider the case when Ω is a square, $\Omega = [-T, T]^2$. Then the above integrals are of the form

$$\int_{[-T, T]^2} F(\mathbf{x} - \mathbf{y}) d\mathbf{y}.$$

This is easily recognized as the *moving sum* of $F(\cdot)$ evaluated at \mathbf{x} . We will use the notation $\text{Sum}(F, \mathbf{x}, T)$ to denote this integral (this is simply the “moving-average”, but without the normalization). As is well-known, this can be efficiently computed using recursion [5], [6].

The main idea here is that, by using shiftable kernels, we can express (2) using moving sums and these in turn can be computed efficiently. In particular, note that the number of computations required for the moving sum is independent of T , that is, the size of the neighborhood Ω . These are referred to as the constant-time or $O(1)$ algorithms in the image processing community. The main steps of our algorithm are summarized in Algorithm 1.

Algorithm 1 Constant-time spatial filtering

Input: $f(x)$, $\varphi(x)$ as in (1), and T .

1. For $1 \leq n \leq N$, use recursion to compute $\text{Sum}(f\varphi_n, \mathbf{x}, T)$ and $\text{Sum}(\varphi_n, \mathbf{x}, T)$.
2. Set $\bar{f}(x)$ as the ratio of $\sum_{n=1}^N c_n(x)\text{Sum}(f\varphi_n, \mathbf{x}, T)$ and $\sum_{n=1}^N c_n(x)\text{Sum}(\varphi_n, \mathbf{x}, T)$.

Return: Filtered image $\bar{f}(x)$.

The above idea can also be extended to non-linear filters such as the edge-preserving bilateral filter [2], [3], [14]. The bilateral filtering of an image $f(x)$ is given by the formula

$$\tilde{f}(x) = \frac{1}{\eta(x)} \int_{\Omega} \varphi(y) \phi(f(x-y) - f(x)) f(x-y) dy \quad (3)$$

where $\eta(x) = \int_{\Omega} \varphi(y) \phi(f(x-y) - f(x)) dy$.

In this formula, the bivariate function $\varphi(x)$ is called the *spatial kernel*, and the one-dimensional function $\phi(t)$ is called the *range kernel*. Suppose that both these kernels are shiftable. In particular, let $\varphi(x - \tau) = \sum_{m=1}^M c_m(\tau)\varphi_m(x)$, and $\phi(t - \tau) = \sum_{n=1}^N d_n(\tau)\phi_n(t)$.

Plugging these into (3) and using linearity, we can write the numerator as

$$\sum_{m,n} c_m(x) d_n(f(x)) \int_{\Omega} \varphi_m(y) \phi_n(f(x-y)) f(x-y) dy.$$

Similarly,

$$\eta(x) = \sum_{m,n} c_m(x) d_n(f(x)) \int_{\Omega} \varphi_m(x-y) \phi_n(f(x-y)) dy.$$

Algorithm 2 Constant-time bilateral filtering

Input: $f(x)$, $\varphi(x)$ and $\phi(s)$ as in (3), and T .

1. For $1 \leq m \leq M$ and $1 \leq n \leq N$, set $a_{m,n}(x) = c_m(x)d_n(f(x))$, $g_{m,n}(x) = \varphi_m(x)\phi_n(f(x))f(x)$, and $h_{m,n}(x) = \varphi_m(x)\phi_n(f(x))$.
2. Use recursion to compute $\text{Sum}(g_{m,n}, \mathbf{x}, T)$ and $\text{Sum}(h_{m,n}, \mathbf{x}, T)$.
3. Set $\tilde{f}(x)$ as the ratio of $\sum_{m,n} a_{m,n}(x)\text{Sum}(g_{m,n}, \mathbf{x}, T)$ and $\sum_{m,n} a_{m,n}(x)\text{Sum}(h_{m,n}, \mathbf{x}, T)$.

Return: Filtered image $\tilde{f}(x)$.

As before, we again recognize the moving sums when $\Omega = [-T, T]^2$. This gives us a new constant-time algorithm for bilateral filtering, which is summarized in Algorithm 2. We note that

this algorithm is an extension of the one in [1], where we used shiftable kernels only for the range filter.

The main computational advantage of Algorithms 1 and 2 is that the number of computations per pixel does not depend on the size of the spatial or the range kernel. It only depends on the order of shiftability. More precisely, the computation time scales linearly with the number of basis functions. For example, in case of the bilateral filter, the computation size is roughly MN times the computations required to perform a single moving sum of the image, plus the overhead time of initializing the basis images and recombining the outputs. To get an estimate of the running time, we implemented Algorithm 2 in MATLAB on an Intel machine with a quad-core 2.83 GHz processor. We found that a single recursive implementation of the moving sum requires roughly 10 milliseconds on a 512×512 image when $T = 5$. Considering an instance where $M \times N = 3^2 \times 5$, the total computation time was roughly 2.5 seconds, 1 seconds to initialize the basis images and combine the results, and 1.5 seconds for the moving sums. This indeed looks promising since we can definitely bring down the time using a JAVA or C compiler. Moreover, we can also use multithreading (parallel computation) to further accelerate the implementation.

III. SHIFTABLE KERNELS

We now address the problem of designing kernels that are shiftable. The theory of Lie groups can be used to study the class of shiftable functions, e.g., see discussions in [12], [15]. It is well-known that the polynomials and the exponentials are essentially the only shiftable functions.

Theorem III.1 (The class of shiftable functions, [12]). *The only smooth functions that are shiftable are the polynomials and the exponentials, and their sum and product.*

We recall that a kernel is a smooth function that is symmetric, non-negative, and unimodal. A priori, it is not at all obvious that there exists a kernel that is a polynomial or exponential. Indeed, the real exponentials cannot form valid kernels since they are neither symmetric nor unimodal. On the other hand, while there are plenty of polynomials and trigonometric functions that are both symmetric and non-negative, it is impossible to find a one that is unimodal over the entire real line. This is simply because the polynomials blow up at infinity, while the trigonometric functions are oscillatory.

Proposition III.2 (Conflicting properties). *There is no kernel that is shiftable on the entire real line.*

Nevertheless, unimodality can be achieved at least on a bounded interval, if not the entire real line, using polynomial and trigonometric functions. Note that, in practice, a priori bounds on the data are almost always available. For the rest of the paper, and without loss of generality, we fix

the bounded interval to be $[-T, T]$.

We now give two concrete instances of shiftable kernels on $[-T, T]$. The reason for these particular choices will be clear in the sequel. For every integer $N = 0, 1, 2, \dots$, consider the functions

$$p_N(t) = \left(1 - \frac{t^2}{T^2}\right)^N \quad \text{and} \quad q_N(t) = \left[\cos\left(\frac{\pi t}{2T}\right)\right]^N,$$

where t takes values in $[-T, T]$. It is easily verified that these are valid kernels on $[-T, T]$. The crucial difference between the above kernels is that the order of shiftability of $p_N(t)$ is $2N + 1$, while that of $q_N(t)$ is much lower, namely $N + 1$. We note that it is the kernel $q_N(t)$ that was used in [1].

The fact that the sum and product of shiftable kernels is also shiftable can be used to construct kernels in higher dimensions. For example, we can set $\varphi(x_1, x_2) = p_N(x_1)p_N(x_2)$, or $\varphi(x_1, x_2) = q_N(x_1)q_N(x_2)$. We call these the *separable kernels*. In higher dimensions, one often requires the kernel to have the added property of isotropy. In two dimensions, we can achieve near-isotropy using these separable kernels provided that N is sufficiently large. We will give a precise reason later in the section. However, as shown in a different context in [16], it is worth noting that kernels other than the standard separable ones (of same order) can provide better isotropy, particularly for low orders. Indeed, consider the following kernel on the square $[-T, T]^2$:

$$\phi(x_1, x_2) = q_1(x_1)q_1\left(\frac{x_1 + x_2}{\sqrt{2}}\right)q_1(x_2)q_1\left(\frac{x_1 - x_2}{\sqrt{2}}\right). \quad (4)$$

The kernel is composed of four cosines distributed uniformly over the circle. It can be verified that $\phi(x_1, x_2)$ is more isotropic than the separable kernel of same order, $q_2(x_1)q_2(x_2)$. However, note that the non-negativity constraint is violated on the corners of the square $[-T, T]^2$ in this case. This is unavoidable since we are trying to approximate a circle within a square. Nevertheless, this does not create much of a problem since the negative overshoot is well within 2% of the peak value. Following the same argument, the polynomial

$$\phi(x_1, x_2) = p_1(x_1)p_1\left(\frac{x_1 + x_2}{\sqrt{2}}\right)p_1(x_2)p_1\left(\frac{x_1 - x_2}{\sqrt{2}}\right)$$

tends to be more isotropic on $[-T, T]^2$ than $p_2(x_1)p_2(x_2)$.

A. Approximation of Gaussian kernels

The most commonly used kernel in image processing is the Gaussian kernel. This kernel, however, is not shiftable. As a result, we cannot directly apply our algorithm for the Gaussian kernel. One straightforward option is to instead approximate the Gaussian using its Fourier series or its Taylor polynomial, both of which are shiftable. The difficulty with either of these is that they do not yield valid kernels. That is, it is not guaranteed that the resulting approximation is non-negative or

unimodal; see [1, Fig. 1]. This is exactly the problem with the polynomial approximation used for the bilateral filter in [13]. As against these, we can instead use the specialized shiftable kernels $p_N(t)$ and $q_N(t)$ to approximate the Gaussian to any arbitrarily precision, based on the following results.

Theorem III.3 (Gaussian approximation). *For every $-T \leq t \leq T$,*

$$\lim_{N \rightarrow \infty} p_N\left(\frac{t}{\sqrt{N}}\right) = \exp\left(-\frac{t^2}{T^2}\right), \quad (5)$$

and

$$\lim_{N \rightarrow \infty} q_N\left(\frac{t}{\sqrt{N}}\right) = \exp\left(-\frac{\pi^2 t^2}{8T^2}\right). \quad (6)$$

In either case, the convergence takes place quite rapidly. The first of these results is well-known. For a proof of the second result, we refer the readers to [1, Sec. II-D]. The added utility of these asymptotic formulas is that we can control the variance of these kernels using the variance of the target Gaussian. This is particularly useful because no simple closed-form expressions for the variance of $p_N(t)$ and $q_N(t)$ are available. We refer the authors to [1, Sec. II-E] for details on how one can exactly control the variance of $q_N(t)$. The same idea applies to $p_N(t)$.

We now discuss how to approximate isotropic Gaussians in two dimensions using shiftable kernels. Doing this using separable kernels is straightforward. For example,

$$\lim_{N \rightarrow \infty} q_N\left(\frac{x_1}{\sqrt{N}}\right) q_N\left(\frac{x_2}{\sqrt{N}}\right) = \exp\left(-\frac{\pi(x_1^2 + x_2^2)}{8T^2}\right). \quad (7)$$

There is yet another form of convergence which is worth mentioning. Consider the following kernel defined on $[-T, T]^2$:

$$\phi_N(x_1, x_2) = \prod_{k=1}^N q_1\left(\sqrt{\frac{6}{N}}(x_1 \cos \theta_k + x_2 \sin \theta_k)\right),$$

where $\theta_k = (k-1)\pi/N$. The kernel $\phi_4(x_1, x_2)$ is simply the (rescaled) kernel $\phi(x_1, x_2)$ in (4). By slightly adapting the proof of Theorem 2.2 in [16, Appendix A], we can show the following.

Theorem III.4 (Approximation of isotropic Gaussian). *For every (x_1, x_2) in $[-T, T]^2$,*

$$\lim_{N \rightarrow \infty} \phi_N(x_1, x_2) = \exp\left(-\frac{\pi^2(x_1^2 + x_2^2)}{8T^2}\right).$$

Note that the target Gaussian in this case is the same as in (7). The key difference, however, is that for low orders N , e.g. $N = 4$, $\phi_N(x_1, x_2)$ looks more isotropic than $q_N(x_1/\sqrt{N})q_N(x_2/\sqrt{N})$. However, as noted earlier, the non-negativity requirement of a kernel is mildly violated by $\phi_N(x_1, x_2)$ at the corners of the square domain.

The significance of the order N is that it allows one to arbitrarily control the accuracy of the Gaussian approximation. As discussed in [1, Sec. II-E], N has to be greater than a threshold N_0 for

the approximating kernels in (5) and (6) to be non-negative and unimodal. In particular, if σ^2 is the variance of the target Gaussian, then N_0 is of the order $O(T^2/\sigma^2)$. Thus, a large N_0 is required to approximate a narrow Gaussian on a large interval. For the spatial kernel, the ratio T^2/σ^2 is usually small, since T is small in this case. This is, however, not the case for the range kernel of the bilateral filter, where T can almost be as large as the dynamic range of the image. The good news is that, by allowing for mild (and controlled) violations of the non-negativity constraint, one can closely approximate the Gaussian using a significantly lower number of terms. This is done by discarding the less significant basis functions in (1); see [17] for detailed discussion and results. The same idea can also be applied to the polynomials.

IV. DISCUSSION

We have shown how, using shiftable kernels, we can express two popular forms of image filters (and possibly many more) in terms of simple moving sums. We note that we can speed up the implementation of Algorithm 1 and 2 using parallel computation or multithreading. In future work, we plan to implement these algorithms in C or JAVA, and make extensive comparison of the result and execution time with the state-of-the-art algorithms.

To conclude, we note that the main idea can also be extended to some other forms of neighborhood filtering, e.g., the ones in [3], [14]. What is even more interesting is that we can extend the idea for approximating the non-local means filter [4]. The non-local means is a higher order generalization of the bilateral filter, where one works with patches instead of single pixels. The filtered image $\hat{f}(x)$ in this case is given by

$$\hat{f}(x) = \frac{\int f(x-y)w(x,y) dy}{\int w(x,y) dy} \quad (8)$$

where

$$w(x,y) = \exp(-b^{-2} \int g(u)(f(x+u) - f(x-y+u))^2 du)$$

Here $g(u)$ is a two-dimensional Gaussian, and the integrals (sums) are taken over the entire image domain. In practice, though, the sum is performed locally [4].

It is possible to express (8) in terms of moving sums, using shiftable approximates of the Gaussian. First, we approximate the domain of the outer integral by a sufficiently large square $[-T, T]^2$, and the inner integral by a finite sum over p neighborhood pixels u_1, \dots, u_p , where, say, $u_1 = 0$. In this case, $\hat{f}(x)$ is given by

$$\frac{1}{\eta(x)} \int_{[-T, T]^2} f(x-y) \varphi(f(x+u_1) - f(x-y+u_1), \dots, f(x+u_p) - f(x-y+u_p)) dy \quad (9)$$

where $\varphi(t_1, \dots, t_p) = \exp(-h^{-2} \sum_{k=1}^p g(\mathbf{u}_k) t_k^2)$, and where $\eta(\mathbf{x})$ is given by

$$\int_{[-T, T]^2} \varphi(f(\mathbf{x} + \mathbf{u}_1) - f(\mathbf{x} - \mathbf{y} + \mathbf{u}_1), \dots, f(\mathbf{x} + \mathbf{u}_p) - f(\mathbf{x} - \mathbf{y} + \mathbf{u}_p)) d\mathbf{y}. \quad (10)$$

Note that $\varphi(t_1, \dots, t_p)$ is an anisotropic Gaussian in p variables with covariance $\text{diag}(h^2/2g(\mathbf{u}_1), \dots, h^2/2g(\mathbf{u}_p))$. Now, using a shiftable approximation (we continue using the same symbol) of $\varphi(t_1, \dots, t_p)$ as in (1), we can write the integrand in (10) as $\sum_{n=1}^N c_n(f(\mathbf{x} + \mathbf{u}_1), \dots, f(\mathbf{x} + \mathbf{u}_p)) \varphi_n(f(\mathbf{x} - \mathbf{y} + \mathbf{u}_1), \dots, f(\mathbf{x} - \mathbf{y} + \mathbf{u}_p))$.

We can then write the numerator in (9) as $\sum_n c_n(f(\mathbf{x} + \mathbf{u}_1), \dots, f(\mathbf{x} + \mathbf{u}_p)) \text{Sum}(G_n, \mathbf{x}, T)$, where we set $G_n(\mathbf{x}) = f(\mathbf{x}) \varphi_n(f(\mathbf{x} + \mathbf{u}_1), \dots, f(\mathbf{x} + \mathbf{u}_p))$. Similarly, letting $H_n(\mathbf{x}) = \varphi_n(f(\mathbf{x} + \mathbf{u}_1), \dots, f(\mathbf{x} + \mathbf{u}_p))$, we have $\eta(\mathbf{x}) = \sum_n c_n(f(\mathbf{x} + \mathbf{u}_1), \dots, f(\mathbf{x} + \mathbf{u}_p)) \text{Sum}(H_n, \mathbf{x}, T)$.

The catch here is that to get a good approximation of non-local means we need to make both T and p large. While there is no problem in making T large (the cost of the moving sum is independent of T), it is rather difficult to make p large. For example, with a separable approximation of $\varphi(t_1, \dots, t_p)$, the overall order N would scale as n^p , where n is the approximation order of the Gaussian along each dimension. This limits the scheme to coarse approximations, and to small neighborhoods. To make this practical, we need a polynomial or trigonometric approximation of $\varphi(t_1, \dots, t_p)$ whose order grows slowly with p . It would indeed be interesting to see if such approximations exist at all.

V. ACKNOWLEDGMENT

The author would like to thank Michael Unser for reading the manuscript and for his useful comments, and also Yuriy Zakharov for his advice.

REFERENCES

- [1] K.N. Chaudhury, D. Sage, and M. Unser, "Fast $O(1)$ bilateral filtering using trigonometric range kernels," *IEEE Trans. Image Process.*, 2011.
- [2] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," *IEEE ICCV*, pp. 839–846, 1998.
- [3] L.P. Yaroslavsky, *Digital Picture Processing—An Introduction*, Springer-Verlag, Berlin, 1985.
- [4] A. Buades, B. Coll, and J.M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation*, vol. 4, pp. 490–530, 2005.
- [5] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *IEEE CVPR*, pp. 511–518, 2001.
- [6] F.C. Crow, "Summed-area tables for texture mapping," *ACM Siggraph*, vol. 18, pp. 207–212, 1984.
- [7] E. Simoncelli, W. Freeman, E. Adelson, and D. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Information Theory*, vol. 38(2), pp. 587–607, 1992.
- [8] W. Freeman and E. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13(9), pp. 889–906, 1991.

- [9] P. Perona, "Deformable kernels for early vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17(5), pp. 488–499, 1991.
- [10] E. Adelson and J. Bergen, "Spatiotemporal energy models for the preception of motion," *J. Optical Society of America*, vol. 2(2), pp. 284–299, 1985.
- [11] A. Watson and A. Ahumada, "Models of human visual-motion sensing," *J. Optical Society of America*, vol. 2(2), pp. 322–342, 1985.
- [12] P.C. Teo, "Theory and applications of steerable functions," Tech. Rep., Stanford University, 1998.
- [13] F. Porikli, "Constant time $O(1)$ bilateral filtering," *IEEE CVPR*, pp. 1–8, 2008.
- [14] S. M. Smith and J. M. Brady, "Susan—A new approach to low level image processing," *Int. J. Comp. Vision*, vol. 23, pp. 45–78, 1997.
- [15] Y. Hel-Or and P.C. Teo, "Canonical decomposition of steerable functions," *J. Math. Imaging Vision*, vol. 9(1), pp. 83–95, 1998.
- [16] K.N. Chaudhury, A. Muñoz-Barrutia, and M. Unser, "Fast space-variant elliptical filtering using box splines," *IEEE Trans. Image Process.*, vol. 19, pp. 2290–2306, 2010.
- [17] K.N. Chaudhury, D. Sage, and M. Unser, "Appendix to "Fast $O(1)$ bilateral filtering using trigonometric range kernels.", Tech. Rep., Ecole Polytechnique Federale de Lausanne, <http://bigwww.epfl.ch/preprints/chaudhury1101pdoc01.pdf>, 2011.