Filtering algorithms using shiftable kernels

Kunal Narayan Chaudhury *

Abstract

It was recently demonstrated in [4] that the non-linear bilateral filter [18] can be efficiently implemented using an \(O(1)\) or constant-time 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 [4] can be extended to few other linear and non-linear filters [18, 21, 2]. 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 [4], we identify a central property of trigonometric functions, called \textit{shiftability}, 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 \textit{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 [19, 8], 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.

\textbf{Keywords:}

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

*Kunal N. Chaudhury (kchaudhu@princeton.edu) is at Princeton University, NJ 08544, USA. The author was supported by a fellowship from the Swiss National Science Foundation.
1 Introduction

The function $\cos(x)$ has the remarkable property that, for any translation $\tau$, $\cos(x - \tau) = \cos(\tau) \cos(x) + \sin(\tau) \sin(x)$. 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 $\varphi(x) = c_1 \exp(\alpha_1 x) + \cdots + c_N \exp(\alpha_N x)$. This follows from the addition-multiplication property of the exponential. As special case, we have the pure exponentials (when $\alpha_i$ is real), and the trigonometric functions (when $\alpha_i$ 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, $\varphi(x) = c_0 + c_1 x + \cdots + c_N x^N$. Note that functions having such properties can be realized in higher dimensions using higher-dimensional polynomials and exponentials, or simply by taking the tensor product of one-dimensional functions.

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

$$\varphi(x - \tau) = c_1(\tau) \varphi_1(x) + \cdots + c_N(\tau) \varphi_N(x). \quad (1)$$

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

Shiftable functions, and more generally, steerable functions [15], have a long history in signal and image processing. Over the last few decades, researchers have found wonderful applications of such special functions in various domains such as image analysis [7, 12, 10], motion estimation [11, 20], and pattern recognition [6, 14], 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 [17] for an excellent account on the theory and practice of steerable functions.

In a recent paper [4], 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 2, where we also propose some $O(1)$ algorithms. To the best of our knowledge, such algorithms have not been reported in the community. The problem of designing shiftable kernels is addressed in Section 3. Here we also propose a scheme for approximating the ubiquitous Gaussian kernel using shiftable functions. Finally, in Section 4, we present some thoughts on how shiftability could be used for reducing the complexity of the non-local means filter [2].

2 Filtering using moving sums

We now show how $O(1)$ algorithms for image filtering can be obtained using shiftable kernels. The idea is that by using shiftability, 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

$$\bar{f}(x) = \frac{1}{\eta} \int_{\Omega} \varphi(y) f(x - y) \, dy$$

where

$$\eta = \int_{\Omega} \varphi(y) \, dy.$$  

Here $\varphi(x)$ is the kernel, and $\Omega$ 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 could construct such kernels using shiftable functions. We will address this problem in the sequel.

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

$$\int_{\Omega} \varphi(y) f(x - y) \, dy = \sum_{n=1}^{N} c_n(x) \int_{\Omega} \varphi_n(x - y) f(x - y) \, dy.$$
Similarly,
\[ \eta = \sum_{n=1}^{N} c_n(x) \int_{\Omega} \varphi_n(x-y) \, dy. \]

Now consider the case when \( \Omega \) is a square, \( \Omega = [-T, T]^2 \). Then the above integrals are of the form
\[ \int_{[-T,T]^2} F(x-y) \, dy. \]

This is easily recognized as the \textit{moving sum} of \( F(\cdot) \) evaluated at the point \( x \). We will use the notation \text{Sum}(F, x, T)\) to denote this integral. As is well-known, this can be efficiently computed using recursion \cite{19, 8, 5}.

The main idea here is that, by using shiftable kernels, we can express \eqref{eq: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 \( \Omega \). These are referred to as the constant-time or \( O(1) \) algorithms in the image processing community. The main steps of our \( O(1) \) algorithm are summarized in Algorithm \textbf{1}.

\begin{algorithm}[h]
\caption{Spatial filtering}
\begin{algorithmic}
\Input Image \( f(x) \), shiftable kernel \( \varphi(x) \) as in \eqref{eq:1}, and \( T \).
\State For \( 1 \leq n \leq N \), use recursion to compute \text{Sum}(f\varphi_n, x, T) \) and \text{Sum}(\varphi_n, x, T).
\State Set \( \tilde{f}(x) = \sum_{n=1}^{N} c_n(x) \text{Sum}(f\varphi_n, x, T) / \sum_{n=1}^{N} c_n(x) \text{Sum}(\varphi_n, x, T). \)
\Endalgorithmic
\end{algorithm}

The above idea can also be extended to non-linear filters such as the edge-preserving bilateral filter \cite{18, 21, 16}. The bilateral filtering of an image \( f(x) \) is given by the formula
\begin{equation}
\tilde{f}(x) = \frac{1}{\eta(x)} \int_{\Omega} \varphi(y) \phi(f(x-y) - f(x)) f(x-y) \, dy \end{equation}

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

\footnote{This is the so-called “moving average”, but without the normalization.}
In this formula, the bivariate function $\varphi(x)$ is the spatial kernel, and the one-dimensional function $\phi(t)$ is 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

$$\phi(t - \tau) = \sum_{n=1}^{N} d_n(\tau) \phi_n(t).$$

As in the earlier case, plugging these into (3), we can write the numerator as

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

Using linearity, we can simplify this to

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

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** Bilateral filtering

**Input**: Image $f(x)$, shiftable kernels $\varphi(x)$ 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 $\text{Sum}(g_{m,n}, x, T)$ and $\text{Sum}(h_{m,n}, x, T)$.

3. Set $\tilde{f}(x) = \sum_{m,n} a_{m,n}(x) \text{Sum}(g_{m,n}, x, T) / \sum_{m,n} a_{m,n}(x) \text{Sum}(h_{m,n}, x, T)$.

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

As before, we again recognize the moving sums when $\Omega = [-T, T]^2$. This gives us a new $O(1)$ algorithm for bilateral filtering, where the number
of computations per pixel is independent of both the size of the spatial and the range kernel. This is summarized in Algorithm 2. We note that this is an extension of the algorithm in [4], where we used shiftable kernels only for the range filter.

3 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 [11, 9]. It is well-known that the polynomials and the exponentials are essentially the only shiftable functions.

Theorem 3.1 (The class of shiftable functions). 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 3.2 (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 $N = 0, 1, 2, \ldots$, consider the polynomial

$$p_N(t) = \left(1 - \frac{t^2}{T^2}\right)^N.$$
and the trigonometric function

\[ q_N(t) = \left[ \cos \left( \frac{\pi t}{2T} \right) \right]^N, \]

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 \( N + 1 \). As we will see shortly, this order is closely related to the approximation order of these kernels. Note that it is the kernel \( q_N(t) \) that was used in [4].

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 [3], 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 \([-T, T]^2\):

\[ \phi(x_1, x_2) = q_1(x_1)q_1 \left( \frac{x_1 + x_2}{\sqrt{2}} \right) q_1(x_2)q_1 \left( \frac{x_1 - x_2}{\sqrt{2}} \right). \] (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 \( p_2(x_1)p_2(x_2) \).
3.1 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. The good news is that we can instead use shiftable kernels \( p_N(t) \) and \( q_N(t) \) to approximate Gaussians to any arbitrarily precision (in practice, all algorithms do use some form of approximation or the other). By plotting these functions (see [4, Fig. 1]), we see that they become more Gaussian-like as the degree increases. The fact, however, is that, as \( N \) gets large, they degenerate at all points to zero, excepting the origin. This problem can be addressed by suitably rescaling the kernels.

**Theorem 3.3 (Gaussian approximation).** For every \(-T \leq t \leq T\),

\[
\lim_{N \to \infty} p_N \left( \frac{t}{\sqrt{N}} \right) = \exp \left( -\frac{t^2}{T^2} \right),
\]

and

\[
\lim_{N \to \infty} q_N \left( \frac{t}{\sqrt{N}} \right) = \exp \left( -\frac{\pi^2 t^2}{8T^2} \right).
\]

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 [4, 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. We refer the authors to [4, 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,

\[
\lim_{N \to \infty} q_N \left( \frac{x_1}{\sqrt{N}} \right) q_N \left( \frac{x_2}{\sqrt{N}} \right) = \exp \left( -\frac{\pi(x_1^2 + x_2^2)}{8T^2} \right). \tag{5}
\]

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 [3, Appendix A], we can show the following.

**Theorem 3.4 (Approximation of isotropic Gaussian).** For every $(x_1, x_2)$ in $[-T, T]^2$,

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

Note that the target Gaussian in this case is the same as in (5). 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.

### 4 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. Moreover, we can speed up the implementation of Algorithm 1 and 2 using parallel computation or multi-threading. We have implemented these algorithms on MATLAB using the parallel computing toolbox, for both the polynomial and trigonometric functions. The preliminary results look satisfactory. Our main critique is that, with the present technology, parallel computation becomes indispensable at large orders when the size of the image stack is large. It is therefore practically difficult to achieve very close approximations of the Gaussian. The algorithms, however, are well suited for applications for which low-order approximations suffice. The reader would appreciate that a detailed analysis of the performance of the algorithms is beyond the scope of this short communication. 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, including our algorithm in [4].

To conclude, we note that the main idea can also be extended to some other forms of neighborhood filtering [21, 16]. What is even more interesting is that we can extend the idea for doing non-local means [2], at least in principle. 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

\[
\hat{f}(x) = \frac{\int f(x - y)w(x, y) \, dy}{\int w(x, y) \, dy}
\]

(6)

where

\[
w(x, y) = \exp \left[ -\frac{1}{h^2} \int g(u) \left( f(x + u) - f(x - y + u) \right)^2 \, du \right].
\]

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 \([2]\).

It is possible to express (6) 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 \( u_1, \ldots, u_p \), where, say, \( u_1 = 0 \). That is, we consider the formula

\[
\hat{f}(x) = \frac{1}{\eta(x)} \int_{[-T,T]^2} f(x - y) \varphi(f(x + u_1) - f(x - y + u_1), \ldots, f(x + u_p) - f(x - y + u_p)) \, dy,
\]

(7)

where  \( \varphi(t_1, \ldots, t_p) = \exp(-h^2 \sum_{k=1}^{p} g(u_k) t_k^2) \), and

\[
\eta(x) = \int_{[-T,T]^2} \varphi(f(x + u_1) - f(x - y + u_1), \ldots, f(x + u_p) - f(x - y + u_p)) \, dy.
\]

Note that  \( \varphi(t_1, \ldots, t_p) \) is simply an anisotropic Gaussian in  \( p \) variables, whose covariance is  \( \text{diag}(h^2/2g(u_1), \ldots, h^2/2g(u_p)) \). Now, using a shiftable approximation (we continue using the same symbol) of  \( \varphi(t_1, \ldots, t_p) \) as in (1), we can write  \( \varphi(f(x + u_1) - f(x - y + u_1), \ldots, f(x + u_p) - f(x - y + u_p)) \) as

\[
\sum_{n=1}^{N} c_n(f(x + u_1), \ldots, f(x + u_p)) \varphi_n(f(x - y + u_1), \ldots, f(x - y + u_p))
\]

as

\[
\sum_{n=1}^{N} c_n(f(x + u_1), \ldots, f(x + u_p)) \varphi_n(f(x - y + u_1), \ldots, f(x - y + u_p)).
\]
Plugging this in (7), we can write the numerator as

\[
\sum_{n=1}^{N} c_n(f(x + u_1), \ldots, f(x + u_p)) \text{Sum}(G_n, x, T),
\]

where \(G_n(x) = f(x)\varphi_n(f(x+u_1), \ldots, f(x+u_p))\). Similarly, setting \(H_n(x) = \varphi_n(f(x + u_1), \ldots, f(x + u_p))\), we have

\[
\eta(x) = \sum_{n=1}^{N} c_n(f(x + u_1), \ldots, f(x + u_p)) \text{Sum}(H_n, x, T).
\]

The catch here is that it 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 \(\varphi(t_1, \ldots, t_p)\), 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 a polynomial or trigonometric approximation of the Gaussian \(\varphi(t_1, \ldots, t_p)\) whose order grows slowly with \(p\). It would indeed be interesting to see if such approximations exist at all.

5 Acknowledgment

The author would like to thank M. Unser for reading the document and for his fruitful comments.

References


