Fast and High-Quality Bilateral Filtering Using Gauss-Chebyshev Approximation

Sanjay Ghosh  Kunal N. Chaudhury
Department of Electrical Engineering, Indian Institute of Science, Bangalore, India
e-mail: {sgosh,kunal}@ee.iisc.ernet.in

Abstract—The bilateral filter is an edge-preserving smoother that has diverse applications in image processing, computer vision, computer graphics, and computational photography. The filter uses a spatial kernel along with a range kernel to perform edge-preserving smoothing. In this paper, we consider the Gaussian bilateral filter where both the kernels are Gaussian. A direct implementation of the Gaussian bilateral filter requires $O(\sigma^2)$ operations per pixel, where $\sigma$ is the standard deviation of the spatial Gaussian. In fact, it is well-known that the direct implementation is slow in practice. We present an approximation of the Gaussian bilateral filter, whereby we can cut down the number of operations to $O(1)$ per pixel for any arbitrary $\sigma$, and yet achieve very high-quality filtering that is almost indistinguishable from the output of the original filter. We demonstrate that the proposed approximation is few orders faster in practice compared to the direct implementation. We also demonstrate that the approximation is competitive with existing fast algorithms in terms of speed and accuracy.

Index Terms—Fast algorithm, bilateral filter, Gaussian kernel, Chebyshev polynomial, minimax approximation.

I. INTRODUCTION

The bilateral filter was proposed by Tomasi and Maduchi [1] as a non-linear extension of the classical Gaussian filter. It is an instance of an edge-preserving filter that can simultaneously smooth homogenous regions and preserve sharp edges. We refer the reader to [2] for a detailed survey of the working of the filter and its various applications.

Before proceeding further, we set up the necessary notation and terminology. Consider an image $f: I \rightarrow \mathbb{R}$, where $I$ is some finite rectangular domain of $\mathbb{R}^2$; we extend the image outside $I$ using symmetry, if required. Consider the Gaussian kernels $g_\sigma: \mathbb{R}^2 \rightarrow \mathbb{R}$ and $g_\sigma: \mathbb{R} \rightarrow \mathbb{R}$ given by

$$g_\sigma(i) = \exp \left( -\frac{||i||^2}{2\sigma^2} \right) \quad \text{and} \quad g_\sigma(t) = \exp \left( -\frac{t^2}{2\sigma^2} \right).$$

The former is called the spatial kernel and the latter is called the range kernel [1]. The output of the Gaussian bilateral filter is the image $B[f]: I \rightarrow \mathbb{R}$ given by

$$B[f](i) = \sum_{j\in\Omega} g_\sigma(j) g_\sigma(f(i-j) - f(i)) f(i-j) \sum_{j\in\Omega} g_\sigma(j) g_\sigma(f(i-j) - f(i)). \quad (1)$$

In practice, the domain of the spatial kernel $\Omega$ is set to be $\Omega = [-W,W] \times [-W,W]$, where $W = 3\sigma$. [1]

It is clear that the direct computation of (1) requires $O(\sigma^2)$ operations per pixel. In general, the direct implementation is slow for practical settings of $\sigma$. [2], [3]. To overcome this, researchers have come up with several fast algorithms; e.g., see [3], [4], [5], [6], [7], [8]. We refer the interested reader to [2], [8] for a survey of fast algorithms. The fast algorithms in [4], [6], [7] are particularly relevant to the present work. The authors here approximate the Gaussian range kernel using polynomial and trigonometric functions respectively, and demonstrate how the resulting filter can be decomposed into a series of spatial Gaussian filtering. The degree (or order) of the spatial filtering is used to control the filtering accuracy; the larger the degree, the better is the accuracy. On the other hand, since the Gaussian filter can be implemented using $O(1)$ operations per pixel (e.g., using separability and recursion [9]), the overall approximation has $O(1)$ complexity with respect to $\sigma$, as a result. Here and henceforth, by filtering accuracy, we will refer to the quality of the filtering in relation to (1) that can be obtained using a given approximation. More recently, a novel approach was proposed in [8], where the author propose to approximate the range kernel using the so-called Gauss-polynomials. Similar to [4], [7], the Gauss-polynomial approximation allows one to decompose the bilateral filter into a series of Gaussian filtering. In particular, it was demonstrated in [8] that the Gauss-polynomial approximation offers better accuracy compared to the polynomial approximation [4], and is more amenable to hardware implementation in contrast to the trigonometric approximations [6], [7].

As will be demonstrated shortly, it turns out that the Gauss-polynomial approximation can result in low-accuracy filtering for images with lot of sharp edges, unless the degree of the polynomial is very high (the run-time, however, scales with the degree). We trace the source of this problem to the Taylor polynomials used in [8] to approximate the exponential function over an interval. In particular, while a Taylor polynomial of fixed degree provides good approximation near the origin, the approximation tends to degrade as one moves away from the origin. In other words, the error incurred by the approximation is not evenly distributed over the interval. As is well-known, this problem can be fixed by replacing the Taylor polynomial with a Chebyshev polynomial of the same degree. Indeed, the Chebyshev polynomials have the property that they balance out the error over the whole interval [10]. We demonstrate that the filtering accuracy can indeed be increased (often substantially) by replacing the Taylor polynomial with a
Chebyshev polynomial of the same degree. While the resulting algorithm structurally resembles the one in [9], the filtering now is shown to be much more accurate for images with sharp edges. Importantly, the run-time of the proposed algorithm is almost identical to that of the algorithm in [8], and we retain the advantage of using polynomials that is advantageous for hardware implementations.

The rest of the paper is organized as follows. We recall the Gauss-polynomial approximation proposed in [8] and the resulting fast algorithm in Section II. We next propose the Gauss-Chebyshev approximation in Section III. We present some simulation results obtained using the proposed fast algorithm in Section IV before concluding the paper.

II. FAST BILATERAL FILTERING

The proposed fast algorithm is an extension of the algorithm in [8]. We now review the main ideas behind the fast algorithm in [8]. The author here proceeds by factoring the range kernel $g_{\sigma_r}(t-\tau)$ in (1), where $t=f(i-j)$ and $\tau=f(i)$, as follows:

$$g_{\sigma_r}(t-\tau) = \exp\left(-\frac{t^2}{2\sigma_r^2}\right) \exp\left(-\frac{t^2}{2\sigma_t^2}\right) \exp\left(\frac{t\tau}{\sigma_t^2}\right).$$

The variables $t$ and $\tau$ take on values in some intensity range $[L, U]$. For example, $L=0$ and $U=255$ for a 8-bit grayscale image. The proposal in [8] is to use Taylor polynomials to approximate the third exponential term in (2), the advantage of doing so will be evident shortly. In particular, setting $x = \tau t / \sigma_t^2$, the following Taylor approximation is considered

$$\exp(x) \approx \sum_{n=0}^N \frac{x^n}{n!}.$$  

The key observation in this context is that there is nothing special about Taylor polynomials, and that any arbitrary polynomial could be used. Of course, the advantage with Taylor polynomials is that the coefficients have a particularly simple expression. Continuing with our observation, we consider a general polynomial approximation

$$\exp(x) \approx \sum_{n=0}^N c_n x^n,$$

where $N$ is the degree of the polynomial. Substituting (3) in (2), we arrive at the following approximation of (2):

$$\exp\left(-\frac{\tau^2}{2\sigma_r^2}\right) \exp\left(-\frac{t^2}{2\sigma_t^2}\right) \sum_{n=0}^N c_n \left(\frac{\tau t}{\sigma_t^2}\right)^n.$$  

(4)

Being the product of two Gaussians and a polynomial, (4) was referred to as the Gauss-Polynomial approximation in [8]. We substitute (4) into (1), and compute the numerator and the denominator of the resulting approximation. Recalling that $t=f(i-j)$ and $\tau=f(i)$, the numerator is given by

$$\sum_{j \in \Omega} g_{\sigma_r}(j)f(i-j) \exp\left(-\frac{(f(i-j))^2}{2\sigma_r^2}\right) \exp\left(-\frac{(f(i))^2}{2\sigma_t^2}\right) \sum_{n=0}^N c_n \left(\frac{f(i)f(i-j)}{\sigma_t^2}\right)^n.$$  

(5)

Next, for $n = 0, \ldots, N$, we construct the images $G_n : I \rightarrow \mathbb{R}$ and $F_n : I \rightarrow \mathbb{R}$ given by

$$G_n(i) = \left(\frac{f(i)}{\sigma_r}\right)^n \quad \text{and} \quad F_n(i) = \exp\left(-\frac{f(i)^2}{2\sigma_r^2}\right)G_n(i).$$  

(6)

We then consider the output of the Gaussian filtering $\tilde{F}_n : I \rightarrow \mathbb{R}$ of each $F_n$, given by

$$\tilde{F}_n(i) = \sum_{j \in \Omega} g_{\sigma_r}(j)F_n(i-j).$$  

(7)

By exchanging the order of the summations, and after some manipulation, we can write (5) as $\exp\left(-f(i)^2/2\sigma_r^2\right)P(i)$, where

$$P(i) = \sigma_t \sum_{n=0}^N c_n G_n(i) \tilde{F}_{n+1}(i).$$  

(8)

In an identical manner, we can approximate the denominator using $\exp\left(-f(i)^2/2\sigma_r^2\right)Q(i)$, where

$$Q(i) = \sum_{n=0}^N c_n G_n(i) F_n(i).$$  

(9)

In other words, the approximation of (1) obtained using the kernel approximation in (4) is given by

$$\hat{B}[f](i) = \frac{P(i)}{Q(i)}.$$  

(10)

As remarked in [8], using the range approximation in (4), we have effectively transferred the non-linearity of the bilateral filter to the intermediate images in (6), which are obtained from the input image using simple pointwise transforms. The computational edge that we get from the above manipulation is that the filtering in (7) can be computed using $O(1)$ operations per pixel for any arbitrary $\sigma_r$ [9]. The overall cost of computing (10) is therefore $O(1)$ per pixel with respect to $\sigma_r$. On the other hand, the complexity is $O(N)$ with respect to $N$, since we are required to compute the $N+2$ images $\tilde{F}_0, \ldots, \tilde{F}_{N+1}$ in [8] and [9]. Needless to say, we can obtain better approximation (and hence better accuracy) by increasing $N$, but at the expense of the run-time.

III. GAUSS-CHEBYSHEV APPROXIMATION

In this section, we present the so-called Gauss-Chebyshev approximation in which we use the Chebyshev approximation of $\exp(x)$. This essentially amounts to computing the coefficients $c_n$ in (3). The motivation behind this choice comes from the well-known fact in approximation theory that, for a fixed degree, the Chebyshev approximation outperforms the Taylor approximation [10], [11]. In fact, Taylor polynomials can be fairly poor to approximate a function, except perhaps on very small intervals. The ideal choice would have been the so-called minimax polynomial that minimizes the maximum absolute error ($L_\infty$ error). However, finding such a polynomial is computationally expensive, and hence it is a common practice to settle for a near-minimax approximation such as that obtained using Chebyshev polynomials [10], [12]. We now present some basic facts about Chebyshev polynomials before turning to the proposed Gauss-Chebyshev approximation.
A. Chebyshev Approximation

The Chebyshev polynomials \( T_0(x), T_1(x), \ldots \) are defined over the interval \([-1, 1]\). They are given by the formula \[ T_l(x) = \cos(l \arccos x), \quad x \in [-1, 1]. \] (11)
The lowest-order polynomials are explicitly given by: \( T_0(x) = 1, T_1(x) = x \), \( T_2(x) = 2x^2 - 1 \), and so on. The higher-order
polynomials can be computed using the recurrence
\[ T_{l+1}(x) = 2xT_l(x) - T_{l-1}(x). \] (12)
This follows from definition (11). Let \( a_{l,n} \) be the coefficient of \( x^n \) in \( T_l(x) \), so that
\[ T_l(x) = \sum_{n=0}^{l} a_{l,n} x^n. \] (13)
Notice that (12) induces a recursion on the coefficients \( (a_{l,n}) \).
This can be used to compute \( (a_{l,n}) \) in an efficient and stable manner (12). The coefficients can also be computed offline and stored in a look-up table. It follows from (11) that \( T_l(x) \) has \( l \) zeros over the interval \([-1, 1]\), given by
\[ \xi_k = \cos \left( \frac{\pi(2k-1)}{2l} \right) \quad (k = 1, 2, \ldots, l). \]
The Chebyshev polynomials satisfy a type of “discrete” orthogonality relationship (12) in that
\[ \sum_{k=1}^{l} T_i(\xi_k)T_j(\xi_k) = \begin{cases} 0 & i \neq j, \\ \frac{l}{2} & i = j \neq 0, \\ l & i = j = 0. \end{cases} \] (14)
A given function \( h(x), x \in [-1, 1] \) can be approximated in terms of Chebyshev polynomials using the linear expansion
\[ h(x) \approx \sum_{l=0}^{N} d_l T_l(x), \] (15)
where \( (d_l) \) are the coefficients of the expansion. Note that the function on the right is a polynomial of degree \( N \). A particularly simple means of fixing the coefficients is to sample \( h(x) \) at the zeros \( \{\xi_k : k = 1, \ldots, N+1\} \) of \( T_{N+1}(x) \). Then, by exploiting the orthogonality relationship in (14), we get
\[ d_l = \begin{cases} \frac{1}{N+1} \sum_{k=1}^{N+1} h(\xi_k), & l = 0 \\ \frac{2}{N+1} \sum_{k=1}^{N+1} h(\xi_k)T_l(\xi_k), & l \neq 0. \end{cases} \] (16)
The approximation in (15) obtained using (16) is very close to the minimax approximation of \( h(x) \) (12).

B. Kernel Approximation and Filtering

The present idea is to approximate the exponential function (third factor) in (2) using Chebyshev polynomials. In particular, we use the Chebyshev approximation in (19) for the polynomial in (3). This choice is motivated using a numerical example in Figure 1 where we compare the approximations achieved using Taylor and Chebyshev polynomials of the same degree. We see that the Taylor polynomial tends to perform poorly as one moves away from the origin. On the other hand, the Chebyshev counterpart works well over the entire interval.

There are, however, some details that we must pay attention prior to applying the approximation. As mentioned earlier, the expansion in (15) is valid only for \( x \in [-1, 1] \). In the present case, the target function is \( h(x) = \exp(x) \), where \( x = \tau t/\sigma^2 \). Recall that \( t = f(i-j) \) and \( \tau = f(i) \), which take values in the intensity range \([L, U]\). Since the Chebyshev polynomials are defined on a symmetric domain, we first apply the following
where \( l_e = (L + U)/2 \). Let \( \mu = (U - L)^2/4\sigma_r^2 \). One can verify that \( g(i)g(j)/\sigma_r^2 \in [\mu, \mu] \) for all \( i \in I \) and \( j \in \Omega \). Thus, we are required to approximate \( h(x) \) over the interval \([\mu, \mu] \). To do so, we consider the following rescaled version of (15):

\[
h(x) \approx \sum_{l=0}^{N} d_l T_l(x/\mu).
\]

For this expansion, formula (16) becomes

\[
d_l = \begin{cases} 
\frac{1}{N+1} \sum_{k=1}^{N+1} h(\mu \xi_k), & l = 0 \\
\frac{2}{N+1} \sum_{k=1}^{N+1} h(\mu \xi_k) T_l(\xi_k), & l \neq 0.
\end{cases}
\]

Plugging (13) into (18), we obtain

\[
h(x) \approx \sum_{n=0}^{N} c_n x^n, \quad c_n = (1/\mu)^n \sum_{l=0}^{N} d_l a_{l,n}.
\]

In summary, the sequence of operations for computing using the Gauss-Chebyshev approximation, which we will henceforth refer to as the Gauss-Chebyshev Filter (GCF), is as follows:

- Center the image using (17).
- Compute \( B[h] \) using the Chebyshev approximation in (19).
- This is done using the fast algorithm in section II.
- Undo the centering: \( B[f] = B[h] + c_e \).

### IV. SIMULATIONS

We first present a couple of representative results in Figures 2 and 3 to demonstrate that the Gauss-Chebyshev approximation is indeed better that the Gauss-Polynomial approximation. Indeed, the Gauss-Chebyshev approximation of the translated kernels are more accurate, particularly for large \( \tau \).

**TABLE I**

| Method \( \sigma_r \) | Exact | GPF | GCF |
|----------------------|-------|-----|-----|
| \( 2 \)            | 103   | 101 | 0.46 |
| \( 3 \)            | 238   | 105 | 0.58 |
| \( 4 \)            | 370   | 7.8 | 0.65 |
| \( 5 \)            | 570   | 8.3 | 0.69 |
| \( 10 \)          | 2250  | 10.2| 0.69 |
| \( 15 \)          | 5207  | 11.4| 0.71 |

We next present results concerning the accuracy and the run-time. The simulations were performed using Matlab on a 3.4 GHz Intel 8-core machine with 32 GB memory. To quantify accuracy, we have used the mean-squared-error (MSE) between (1) and the filtering obtained using the corresponding fast algorithm. The MSE between two images \( f : I \rightarrow \mathbb{R} \) and \( g : I \rightarrow \mathbb{R} \) is defined to be \( 10 \log_{10}(\text{MSE}) \) dB, where \( \text{MSE} = \|I\|^{-1} \sum_{i \in I} (f(i) - g(i))^2 \). For the simulation, we have used the Checker and Testpat images from [13], and the popular Barbara image. The former images have lot of sharp edges, while the latter has a fair mix of texture and edges. The accuracies offered by the Gauss-Polynomial and the Gauss-Chebyshev Filter are compared in Table I. It is not surprising that the latter performs significantly better. This is also evident from the visual results in Figure 4.

Further visual results are provided in Figures 5 and 6. Notice in Figure 5 that the GCF filtering is significantly better than some of the existing fast algorithms. On the other hand, GPF and GCF have almost identical performance for the Barbara image in Figure 4. This is not surprising given that this image has far less sharp edges compared to the other images.

Finally, we compare the run times and the accuracy of various fast algorithms and report the results in Table II. We see that the accuracy of GCF is substantially better that that offered by some of the existing fast algorithms, while its run-time is comparable. In general, notice that the run-time of GCF is few orders smaller than that of the direct implementation of (1).
Fig. 5. Comparison of the direct implementation with various fast algorithms on a montage. The filter parameters are $\sigma_s = 5$ and $\sigma_r = 30$. For GPF and GCF, the approximation degree is $N = 26$.

V. Conclusion

We proposed a fast algorithm for Gaussian bilateral filtering based on the Gauss-Chebyshev approximation. In particular, we demonstrated that the algorithm is comparable with the GPF algorithm [8] in terms of run-time, but performs significantly better on images that have lot of sharp edges. We note that the GPF algorithm was already shown in [8] to be superior to existing fast algorithms for natural images, such as the Barbara image in Figure 6. The proposed algorithm can be particularly useful for regularizing depth maps arising in stereo reconstruction, which exhibits sharp transitions. For example, it was demonstrated in [14] that a local aggregation of the similarity score using bilateral weights significantly improves the disparity estimation.

Fig. 6. Comparison of the direct implementation of (1) with GPF [8] and the proposed GCF algorithm. The parameters of the filter are: $\sigma_s = 5$ and $\sigma_r = 30$. The degree of the polynomial is $N = 30$.

REFERENCES

[1] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” Proc. IEEE International Conference on Computer Vision, pp. 839-846, 1998.
[2] P. Kornprobst and J. Tumblin, Bilateral Filtering: Theory and Applications, Now Publishers Inc., 2009.
[3] S. Paris and F. Durand, “A fast approximation of the bilateral filter using a signal processing approach,” Proc. European Conference on Computer Vision, pp. 568-580, 2006.
[4] F. Porikli, “Constant time $O(1)$ bilateral filtering,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2008.
[5] Q. Yang, K. H. Tan, and N. Ahuja, “Real-time $O(1)$ bilateral filtering,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 557-564, 2009.
[6] K. N. Chaudhury, D. Sage, and M. Unser, “Fast $O(1)$ bilateral filtering using trigonometric range kernels,” IEEE Transactions on Image Processing, vol. 20, no. 12, pp. 3376-3382, 2011.
[7] K. N. Chaudhury, “Acceleration of the shiftable algorithm for bilateral filtering and nonlocal means,” IEEE Transactions on Image Processing, vol. 22, no. 4, pp. 1291-1300, 2013.
[8] K. N. Chaudhury, “Fast and accurate bilateral filtering using Gauss-polynomial decomposition,” Proc. IEEE International Conference on Image Processing, pp. 2005-2009, 2015.
[9] R. Deriche, “Recursively implementing the Gaussian and its derivatives,” Research Report, INRIA-00074778, 1993.
[10] D. Elliott, D. F. Paget, G. M. Phillips and P. J. Taylor, “Error of truncated Chebyshev series and other minimax polynomial approximations,” Journal of Approximation Theory, vol. 50, pp. 49-57, 1987.
[11] N. Brisebarre and M. Joldes, “Chebyshev interpolation polynomial-based tools for rigorous computing,” Proc. International Symposium on Symbolic and Algebraic Computation, ACM, pp. 147-154, New York, 2010.
[12] W. H. Press, B. P. Flannery, S. A. Teutolsky, and W. T. Vetterling, Numerical Recipes. The Art of Scientific Computing, Cambridge University Press, 1990.
[13] USC-SIP Image Database, http://sipi.usc.edu/database.
[14] K. Yoon and I. Kweon, “Adaptive support-weight approach for correspondence search,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 4, pp. 650 - 656, 2006.