Global Color Sparseness and a Local Statistics Prior for Fast Bilateral Filtering

Mikhail G. Mozerov, Member, IEEE and Joost van de Weijer, Member, IEEE

Abstract—The property of smoothing while preserving edges makes the bilateral filter a very popular image processing tool. However, its non-linear nature results in a computationally costly operation. Various works propose fast approximations to the bilateral filter. However, the majority does not generalize to vector input as is the case with color images. We propose a fast approximation to the bilateral filter for color images. The filter is based on two ideas. Firstly, the number of colors which occur in a single natural image is limited. We exploit this color sparseness to rewrite the initial non-linear bilateral filter as a number of linear filter operations. Secondly, we impose a statistical prior to the image values which are locally present within the filter window. We show that this statistical prior leads to a closed-form solution of the bilateral filter. Finally, we combine both ideas into a single fast and accurate bilateral filter for color images. Experimental results show that our bilateral filter based on the local prior yields an extremely fast bilateral filter approximation, but with limited accuracy, which has potential application in real-time video filtering. Our bilateral filter which combines color sparseness and local statistics yields a fast and accurate bilateral filter approximation and obtains state-of-the-art results.

Index Terms—Bilateral filter, color image filtering, image enhancement.

I. INTRODUCTION

The edge preserving bilateral filter has been independently proposed in [5], [35], [36], and later generalized in [31], [30]. It is a very popular filter in image processing, computer graphics, and computer vision, and has been applied to a wide variety of applications such as denoising [7], edit propagation [4], optical flow and stereo matching [41], [26], [27], video abstraction and demosaicing [40], [34]. It has been shown to be closely related to anisotropic diffusion, robust estimation and the mean shift operation [38], [6]. Because the computational complexity of the bilateral filter is demanding, fast approximations of it are highly desirable.

The bilateral filter has the property of smoothing while preserving sharp edges, which makes it one of the most popular image enhancement filters. This is achieved by combining two kernels: a spatial kernel and a range kernel. The spatial kernel $\phi$ assigns pixels which are close to the center pixel a higher weight. The range kernel $\theta$ assigns more weight to pixels which are close to the center pixel value. The bilateral filtering can be expressed as

$$\tilde{f}(x) = \eta^{-1}(x) \sum_{y \in \Omega} \theta(|r(x) - r(x - y)|) \varphi(|y|) f(x - y)$$

$$\eta(x) = \sum_{y \in \Omega} \theta(|r(x) - r(x - y)|) \varphi(|y|),$$

where we use bold to indicate vector values. In case of a standard color image the function $f(x)$ represents the color channels $f(x) = (R(x), G(x), B(x))$, and $\eta(x)$ is a normalization factor. Here $x$ and $y$ are coordinates in the image coordinate system domain $\Omega$; $f(x)$ and $\tilde{f}(x)$ represent the input and output of the filter respectively; and $r(x)$ is a range image, which in the canonical bilateral filter coincides with the input function $f(x)$. The most common bilateral filter has Gaussian kernels for both the spatial and range domain [36], [28].

In this paper we propose two approaches to fast bilateral filtering in color images. The first is based on the observation of color sparseness, referring to the fact that only a limited set of colors suffices to represent the colors in an image. Other than Paris and Durand [29] who proposed a grid sampling of the combined image coordinate-color space, we propose to decompose the non-linear bilateral filter into a set of linear filter operations along a limited, but representative, set of colors. We show that exploiting color sparseness can reduce significantly the number of linear filter operations which are required to achieve an accurate bilateral filter approximation. Our second approach to fast bilateral filtering is derived by assuming a prior on the local statistics. We show that, assuming a uniform distribution over the pixel values in a local histogram, it is possible to derive a closed-form solution to this 'approximated' bilateral filter. Note that, other than methods which apply histograms for bilateral filtering [33], [38], our method never explicitly computes a histogram. The approximation is extremely fast, but is not as accurate as the approximation based on color sparseness. Therefore, we combine both methods into a single fast and accurate bilateral filter which we compare to state-of-the-art bilateral filter approximations in the experimental section.

This paper is organized as follows. In section II we discuss related work. In section III we propose an approximation to the bilateral filter based on the cluster decomposition. In section IV, we derive a closed-form solution to the bilateral filter by imposing a prior on the local statistic. In section V we combine both contributions into a single fast and accurate bilateral filter approximation. Experimental results are provided in section VI. Finally, conclusions are presented in section VII.
II. RELATED WORK

One of the first bilateral filter approximations was proposed by Durand and Dorsey in [14] and later improved in [28]. The main idea of the method is to approximate the image by sampled copies and to filter these copies with discrete kernels. This approach was further improved by Yang et al. [44] by introducing a piecewise-linear approximation of the range kernel. This approach achieves O(1) computational complexity. More recently another constant-time algorithm was proposed by Chaudhury et al. [10]. This method exploits the shiftability property of trigonometric functions. It was further improved in [9]. The recently published paper by Galiano and Velasco [18] proves the equivalence between the standard pixel-based version and a rearranged version of the filter. Restricted to the discrete setting, this reformulation of the bilateral filter allows for a fast implementation of the bilateral filter. Pham and Van Vliet proposed a separable implementation of the bilateral filter [32]. The local histogram approach was first proposed by Weiss [39]. This approach was significantly improved by Porikli [33]. He demonstrated a constant-time realization of the bilateral filter using polynomial range kernels. Similar approach was developed by Gunturk [20]. The method uses multiple uniform box kernels, which are optimally combined to approximate an arbitrary domain kernel. A drawback of these approaches is that they apply to grayscale images and their extension to the vector domain is problematic. In this paper we focus on bilateral filtering of color images.

Most approaches to bilateral filtering for color images are based on the splat-blur-slice paradigm. A comprehensive overview of such methods has been provided by Adams [1]. The principle idea can be traced back to the work of Yang et al. [42] who proposes a clustering approach for fast high-dimensional filtering. This idea was further developed in [3] and applied to fast bilateral filtering. Paris and Durand [29] proposed to sample the high dimensional color-range space on a regular grid. This extension is however problematic, and their approach which involves a convolution in the 5D combined image coordinates-color space, is only feasible for kernels with a large spatial extent. This sampling approach was improved by introducing the compact permutohedral lattice with sparsely used nodes [2]. This permutohedral lattice algorithm considerably outperforms the earlier KD-tree method [3]. Gastal and Oliveira [18] propose a similar clustering approach, but for interpolation nodes the manifolds are used.

Excellent work on edge-preserving filtering based on geodesic distance has been proposed in recent years [17]. The usage of the geodesic distance instead of the combined Euclidean tonal-spatial distance results in filters with different properties. A fundamental difference between the two is that bilateral filters communicate over edges within the spatial extend of the filter, whereas geodesic distance-based filters do not. Depending on the application that property of bilateral filtering can be desirable. Also geodesic distance based filters suffer from a lack of rotation invariance while our approximations do not. In addition, bilateral filtering has theoretical links with the mean-shift operation and inference in fully connected fields [23]; fields which potentially benefit from fast implementations. The geodesic distance was also used in [12] for the purpose of image and video editing. This paper uses clustering to speed-up computation. However, other than us, they ignore the vectorial nature of color images and only applied to clustering to the luminance domain. A recent work [43] proposed a recursive implementation of the bilateral filter and also uses the geodesic distance approach. It should be noted however that this implementation does not provide an approximation of the canonical bilateral filter with Gaussian kernels which we study in this paper.

There exist a considerable amount of research on edge preserving filtering which are not using the bilateral filter. For example, several works [24], [21] use for this purpose a local linear regression (the guided filter class). However, like in the case of the geodesic distance based filters the filtered output possesses different property in comparison with the canonical bilateral filter with Gaussian kernels.

Within the context of this paper it is interesting to point out that the usage of clustering has been applied outside bilateral filter theory for various non-linear filter operations [8], [25], [13]. Mostly, the purpose of these filters is denoising or image reconstruction problems. Such techniques explicitly exploit self-similarities in natural images. As a result filters average out the noise among similar patches, while sparse coding
encodes each image patch into a linear combination of a few elements from a set called a dictionary. In this paper, we apply clustering for the purpose of fast bilateral filtering.

III. BILATERAL FILTERING BY CLUSTER DECOMPOSITION

As mentioned above, the extension of several fast implementations of the bilateral filter to color is problematic. Here we propose an alternative approach to bilateral filtering of color images. Throughout this paper we use the canonical bilateral filter with the Gaussian kernel:

\[ \varphi(x) = \exp \left( -\frac{x^2}{2\sigma^2} \right), \]

where \( \sigma \) is the intrinsic parameter of the filter. Similarly, the kernel of the range domain kernel \( \theta(x) \) is Gaussian with intrinsic parameter \( \sigma_r \).

A. Cluster decomposition principle

The main idea of this method is to replace the computationally expensive non-linear filtering operation with a set of fast linear filter operations. Given \( K \) cluster centers with mean value \( \mu_k \) we assign every \( f(x) \) to the closest cluster; hence if \( m_k(x) = 1 \) then \( f(x) \) has been assigned to cluster \( k \). Here we consider only hard assignment, meaning \( m_k(x) \in \{0, 1\} \), and \( \forall x \in \Omega \) holds that \( \sum_{k \in K} m_k(x) = 1 \). Then, our bilateral filter approximation is based on what we term the cluster decomposition principle: any function can be decomposed as follows

\[ f(x) = \sum_{k \in K} f_k(x) = \sum_{k \in K} m_k(x) f(x), \]

where we define \( f_k(x) = m_k(x) f(x) \).

The range image in Eq. 1 can similarly be decomposed

\[ r(x-y) = \sum_{k \in K} m_k(x-y) r(x-y) = \sum_{k \in K} m_k(x-y) (\mu_k + \delta(x-y)) \approx \sum_{k \in K} m_k(x-y) \mu_k. \]

It is easy to see that the residual \( \delta(x-y) = r(x-y) - \mu_k \) is equal to zero when the number of clusters is equal to the full color range (considering a discrete color space) and the mask function \( m_k(x) \) is non zero. In other cases, the formula is an approximation.

We can now apply the cluster decomposition to propose a fast approximation to the bilateral filter. This bilateral filter is based on linear filter operations for each of the \( K \) clusters. This sparse approximation is denoted by \( \hat{f}^S \) and is derived by plugging Eq. 3 and Eq. 4 into Eq. 1.

\[ \hat{f}^S(x) = \eta^{-1}(x) \sum_{k \in K} \theta(|r(x) - \mu_k|) \hat{f}_k(x) \]

\[ \eta(x) = \sum_{k \in K} \theta(|r(x) - \mu_k|) \overline{m}_k(x), \]

where we use the dash above the function to denote the spatial filtered version, e.g. \( \hat{f}_k(x) = \sum_{y \in \Omega} \varphi(|y|) f_k(x-y) \).

A derivation of Eq. 5 is provided in the Appendix.

The scheme of the bilateral filtering process based on the proposed cluster decomposition is illustrated in Fig. 1. Eq. 5 approximates the initial non-linear filter problem in Eq. 1 with \( K \) linear filtering sub-problems. After decomposing the original image into \( K \) color images according to the cluster assignments, the spatial kernel is applied to these images separately. Next, the range-kernel is applied when combining the images into the final bilateral filter approximation.

In Eq. 5 we use the global cluster mean \( \mu_k \), which, in case of for example k-means, is based on the global image according to

\[ \mu_k = \frac{\sum_{x \in \Omega} f_k(x)}{\sum_{x \in \Omega} m_k(x)}, \]

instead we could also apply a local cluster mean by using

\[ \overline{m}_k(x) = \frac{\hat{f}_k(x)}{\overline{m}_k(x)}. \]

Note that both numerator and denominator have already been computed; therefore this change comes at negligible computational cost. Especially, for a low number of clusters it is expected that the local cluster mean provides a much better approximation of the actually range distance. The two methods are compared in the Fig. 2 (left graph) where we show the PSNR between the true bilateral filter output and the fast approximations. The numbers are based on 1000 random images from the COREL dataset. One can observe that the local averaging in Eq. 7 allows to decrease the computational cost almost by a factor of two: for example to achieve 40 dB the global cluster mean requires 14 clusters while the local cluster mean achieves these results already for 8 clusters.

B. Global Color Statistics

Bilateral filtering based on the clustering decomposition is exact when the number of clusters is equal to the color values present in the image (then the residual error in Eq. 4 is zero). Hence, for this algorithm to be successful it is important that the color statistics of images has a sparse behavior, and can be approximated by few cluster centers.

To investigate sparseness of global color statistics, consider a simple cluster quantization experiment: we replace all the pixels in an image with a limited set of colors (assigning each to the closest color in an Euclidian sense), after which we evaluate the quantization error. We compare three strategies here to fix the limited set of colors. For the first approach we choose a regular grid in the RGB space, leading to sets with \( 2^3, 3^3 \), etc number of colors. This strategy is applied for fast bilateral filtering on grey images [29], however as we will see it is not optimal for color images. In the second approach, we compute the color statistics of a large set of images (here we use 1000 random images of the COREL data set), and apply a k-means algorithm to obtain a fixed set of colors which we will use for all images. In the third approach, we apply k-means to the statistics of each single image, i.e. we will change the set of colors with each image.

In Fig. 2 (bottom graph) the quantization error in dB is given for the three methods. The experiment is done on the 40,000 images of the COREL dataset. Comparing the three methods it shows - none surprisingly - that choosing the color
that the derivation in this section is only valid for canonical bilateral filtering when \( r(x) = f(x) \).

Let us first consider the one channel luminance case. Later we will extend the theory to color. Eq. 1 can be rewritten as:

\[
\tilde{f}(x) = \eta^{-1}(x) \sum_{z \in Z} z \theta(|f(x) - z|) h(x, z),
\]

with the normalization term being,

\[
\eta(x) = \sum_{z \in Z} \theta(|f(x) - z|) h(x, z),
\]

where \( z \) is the intensity value and \( Z \) its domain, e.g. [0,255]. The histogram function \( h \) is given by

\[
h(x, z) = \sum_{y \in \Omega} \varphi(|y|) \delta(f(x - y) - z),
\]

where \( \delta \) is the Kronecker delta function. The function \( h \) can be interpreted as a local histogram. Intensity values, \( f(x) \), are weighted with the spatial kernel \( \varphi(|y|) \) and collected together over the local neighborhood. The bilateral filter operation, according to Eq. 8, can be interpreted as taking a weighted average of the intensity in the local histogram \( h \). The weighting dependents on the distance from the luminance of the center pixel according to \( \theta(|f(x) - z|) \). The histogram \( h \) represents the local distribution of luminance values given the spatial kernel \( \varphi(|y|) \).

To prevent the costly computation of \( h \) we propose to enforce a prior on the local distribution. For the prior to yield a good approximation of the actual bilateral filter, it should have some of the same properties as the true histogram \( h \). The mean of the local distribution \( h \) can be computed very fast, and we propose to use a prior distribution with similar mean. The mean is computed with:

\[
\hat{h}(x) = \int_{\Omega} z h(x, z) dz = \int_{\Omega} z \sum_{y \in \Omega} \varphi(|y|) \delta(f(x - y) - z) dz = \sum_{y \in \Omega} f(x - y) \varphi(|y|).
\]

This is convenient since, if we consider the Gaussian smoothed version of the image to be \( \tilde{f} = f \otimes \varphi \), where we use \( \otimes \) to denote spatial convolution, then we can write \( \hat{h}(x) = \tilde{f}(x) \).

As the local statistics prior we propose to use the uniform distribution. We choose the uniform distribution to share some of the properties of the actual local distribution \( h \), namely that it has the same mean as \( h \) and that it includes the actual image value \( f(x) \). The family of uniform which fulfills these requirements is given by

\[
\hat{h}^\Delta(z, x) = \begin{cases} 
\frac{1}{2(|u(x)| + \Delta)} & \text{for } \tilde{f}(x) - u(x) - \Delta \leq z \leq \tilde{f}(x) + u(x) + \Delta \\
0 & \text{else} \end{cases}
\]

where we apply \( u(x) = \tilde{f}(x) - f(x) \) to denote the signed distance between the image value and its spatial smoothed version. The local prior \( \hat{h}^\Delta \) assumes that the local distribution of luminance values is equal to a uniform distribution from \( \tilde{f}(x) - u(x) - \Delta \) to \( \tilde{f}(x) + u(x) + \Delta \), where \( \Delta \) is a constant, whose choice we outline later. Choosing \( \Delta = 0 \) we have

IV. A LOCAL STATISTICS PRIOR

Here we investigate the usage of a local statistics prior to approximate bilateral filtering. With local statistics we refer to the distribution of image values within the local window. First we show that the bilateral filter can be understood as operating on a local histogram. Then we propose a local statistics prior and show that this prior leads to a closed-form solution and a fast \( O(1) \) bilateral filter approximation, consisting of only a single Gaussian filter operation combined with pixel operations. In Section V this idea will be combined with the cluster decomposition proposed in the previous section.
the narrowest distribution which fulfills the requirements (see Fig. 3(top) for an illustration of the local statistics prior). Note that in Eq. 12 we assume that \( u(x) \geq 0 \). In case \( u(x) \leq 0 \) replace \( \leq \) by \( \geq \) in Eq. 12.

Having defined the local statistics prior \( \hat{h} \), we now proceed to compute an estimate of the bilateral filter \( \hat{f}^p \) based on the local prior. Combining Eq. 8 and Eq. 12 and using Eq. 2 for \( \theta \) we get:

\[
\hat{f}^p(x) = \frac{\int f(x) + u(x) + \Delta}{\int f(x) + u(x) - \Delta} \cdot \frac{\int f(x) - u(x) - \Delta}{\int f(x) - u(x) + \Delta} = \frac{\int f(x) + u(x) + \Delta \cdot e^{-\frac{1}{2\sigma^2}(f(x)-z)^2} \, dz}{\int f(x) + u(x) - \Delta \cdot e^{-\frac{1}{2\sigma^2}(f(x)-z)^2} \, dz}
\]

which can be simplified to

\[
\hat{f}^p(x) = f(x) + \frac{2u(x) + \Delta}{-\Delta} \cdot \frac{\int f(x) - u(x) - \Delta \cdot e^{-\frac{1}{2\sigma^2}(f(x)-z)^2} \, dz}{\int f(x) - u(x) + \Delta \cdot e^{-\frac{1}{2\sigma^2}(f(x)-z)^2} \, dz}
\]

The integral has the following closed-form solution

\[
\hat{f}^p(x) = f(x) + \frac{\sqrt{2}\sigma_r u(x)}{\sqrt{\pi} |u(x)|} \cdot e^{-\varepsilon_1^2} - e^{-\varepsilon_2^2} \cdot erf(\varepsilon_1) + erf(\varepsilon_2),
\]

where we used the fact that

\[
\int e^{-z^2} \, dz = \frac{1}{2} \sqrt{\pi} erf(z) + c, \quad \int e^{-z^2} \, dz = \frac{1}{2} e^{-z^2} + c
\]

with \( c \) a constant, and

\[
\varepsilon_1 = \frac{\Delta}{\sqrt{2}\sigma_r}, \quad \varepsilon_2 = \frac{\Delta + 2u(x)}{\sqrt{2}\sigma_r}.
\]

In conclusion, the local statistics prior which we proposed in Eq. 12 only needs the computation of the smoothed image \( f \) to arrive at a closed-form estimation of the bilateral filter. To the best of our knowledge this is the first bilateral filter approximation which is based upon only a single filter operation, namely the computation of the Gaussian smoothed image. An example of bilateral filtering based on a local image statistics prior is given in Figure 3.

For color images we use the following fast approximation to the bilateral filter:

\[
\hat{f}^p(x) = f(x) + \frac{\sqrt{2}\sigma_r u(x)}{\sqrt{\pi} |u(x)|} \cdot e^{-\varepsilon_1^2} \cdot erf(\varepsilon_1) + erf(\varepsilon_2),
\]

with

\[
\varepsilon_2 = \frac{\Delta}{\sqrt{2}\sigma_r}, \quad \varepsilon_1 = \frac{\Delta + 2|u(x)|}{\sqrt{2}\sigma_r},
\]

where \( |u(x)| \) denotes the Euclidean distance of the vector. Note that the \( u(x) = f(x) - f(x) \) is now a vector. The prior distribution, which is assumed here, is 1D and is the uniform distribution which goes from \( f(x) - \Delta \frac{|u(x)|}{|u(x)|} \) to \( f(x) + 2u(x) + \Delta \frac{|u(x)|}{|u(x)|} \). This color extension seems a ruder approximation than its luminance counter part, assuming all values to lie on the line defined by the actual RGB value and its locally averaged mean RGB value. However, this ensures that the final solution does not introduce any chromaticities which are absent in the input image, something which would certainly happen if one would apply the local statistics prior to the three color channels separately.

In a similar experiment as in Section III we evaluate the quality of the here presented bilateral filter based on local statistics prior. On 1000 random images from the COREL we compare the proposed filter \( \hat{f}^p \) to the true bilateral filter (again \( \sigma_r = 0.1 \) and \( \sigma_s = 5 \) are used). We measured the impact of \( \Delta \). We found its behavior to be reasonably stable when varying the bilateral filter parameters (\( \sigma_r \) and \( \sigma_s \)). Based on the COREL data set we found \( \Delta = 6.25\% \) to obtain optimal results, where the percentage is over the maximum range of the image. Applying this value in the above experiment we obtained an PSNR = 35.11dB, almost 3.5dB above the baseline with \( \Delta = 0 \). In our experiments we will therefore apply \( \Delta = 6.25\% \).

The fact that bilateral filtering can be interpreted as an operation on local histogram space [22] has been understood in several earlier works [38], [28], [37]. In these works histogram

\[\text{Fig. 3. Illustration of bilateral filtering based on local statistics prior (\( \Delta = 0 \)). (top) Line-plot from 2D step edge signal with noise in blue and its Gaussian smoothed version in dotted blue (‘-‘). The blue area indicates the local prior, within this region the local distribution \( h(x) \) is assumed to be uniform. (bottom) Original signal in blue, and smoothed in blue dotted. The bilateral filter is given in green (‘-‘), and the bilateral filter approximation based on the local prior in red (‘-‘). Note, that the green line nicely follows the bilateral filter output in red.}\]
space is actually the smoothed version of our histogram, 
\( H = h \otimes \theta \), where the smoothing is performed with the range kernel \( \theta \). Van de Weijer et al. [38], [37] show that bilateral filtering can be interpreted as fixed point iteration to the local modes in the \( H \) histogram, and is closely related to the mean shift filter [11]. Paris et al. [28] used the low bandpass nature of \( H \) to propose a speeded up implementation of the bilateral filter. However, their method differs from ours in that no local statistics where exploited.

V. COMBINING COLOR SPARSENESS AND LOCAL STATISTICS PRIOR

Above we have proposed two fast approximations to bilateral filtering. The approximation based on the cluster decomposition (\( \hat{f}^p \)) yields a highly accurate bilateral filter, and depending on the number of clusters this approximation is more accurate than the bilateral filter based on the local statistics prior (\( \hat{f}^p \)). However, the latter is significantly faster than the former. Here we combine the two approaches into a single fast and accurate bilateral filter approximation. This extension is valid for canonical filtering when \( r(x) = f(x) \). The combination of the two approximations is based on two observations:

1) Suppose that within the neighborhood (e.g. \( 3\sigma_x \)) of a point \( x \) all values of the range image belong to the same color cluster \( k \). Then the best approximation to the original bilateral filter in this point can be reached if the local cluster mean is equal to the real bilateral filter \( \hat{f} \). In other words the cluster-wise bilateral filtering or a good approximation to this filter is highly desirable.

2) When \( m_k(x) = 0 \) (a point \( x \) does not belong to a \( k \) cluster) the distance \( |f(x) - f_k(x - y)| \) inside the neighborhood of the point \( x \) varies relatively less than distance \( |f(x) - f_l(x - y)| \) under condition \( m_l(x) = 1 \). Thus, the correction for the cluster \( k \) with \( m_k(x) = 0 \) is less important and can be omitted.

The first observation suggests that it would be preferable to replace \( \bar{I}_k(x) \) in Eq. 5 with the bilateral filter approximation \( \hat{I}_k(x) \). The second observation suggests that this replacement is especially important for \( m_k(x) = 1 \). This prevents us from computing the more expensive \( \hat{I}_k(x) \) for all clusters at every location. We combine the observations with:

\[
\bar{I}_k(x) = \begin{cases} \hat{I}_k(x) & \text{if } m_k(x) = 1 \\ \hat{I}_k(x) & \text{else} \end{cases}
\]

This leads to the following bilateral filter approximation which combines color sparseness and the local statistics prior:

\[
\hat{f}^{sp}(x) = \eta^{-1}(x) \sum_{k \in K} \theta(\|f(x) - \bar{I}_k(x)\|) \bar{I}_k(x) \]

\[
\eta(x) = \sum_{k \in K} \theta(\|f(x) - \bar{I}_k(x)\|) m_k(x),
\]

where

\[
m_k(x) = \frac{\bar{I}_k(x)}{m_k(x)}.
\]

The optimal \( \Delta \) varies with the number of clusters \( K \). We found empirically that good results were obtained by applying \( \Delta = \frac{0.25\sqrt{K}}{\sqrt{K}} \) and we use this in all our experiments. The results of the bilateral filter approximation \( \hat{f}^{sp}(x) \) as a function of the number of clusters are shown in the left graph of Fig. 2 (green line). One can observe that for example to obtain 40dB PSNR on average only five clusters are required for \( \hat{f}^{sp}(x) \). The complexity of the here presented filter is \( O(K) \).

VI. EXPERIMENTS

Here we evaluate the performance of our bilateral filter approximations against several state-of-the-art filters. We address both accuracy and speed of the approximations.

Experimental settings: The experiments are performed on a set of six color images (courtesy Paris and Durand [29]) which are commonly used for filter evaluation. The average size of the images used is 370,000 pixels. We evaluate the accuracy of the approximation by means of the PSNR error between the bilateral filter approximation and the real bilateral filter for these 24 settings on all six images. We report the average PSNR in dB. The PSNR of one color image is computed with:

\[
PSNR = 20 \log_{10} \left( \frac{MAX_f}{\sqrt{MSE}} \right)
\]

where \( MSE \) is the least mean square error between the estimated image and the true bilateral filtered image, and \( MAX_f \) refers to the maximum image value which is 255 \( \times \sqrt{3} \) for the color images in our experiments.

To obtain the \( K \) clusters which are at the bases of the cluster decomposition (Eq. 3) a fast clustering algorithm is required. In our experiments we have used the minimum variance quantization (implemented in matlab as rgb2ind) which took less than 0.02 seconds for one Megapixel image. For the implementation of the Gaussian filters we use the O(1) recursive implementation [45][19]. In addition Gaussian smoothing is performed in an image which is downscaled with a factor equal to \( \sigma_s \). For both the down and upscaling we perform bilinear interpolation. The convolutional operation applied after downscaling is corrected for the downsampling factor which has been applied. All references to execution time of algorithms refer to processing time on an Intel Xeon 3.60 GHz processor. The code of our bilateral filter approximations is available at http://www.cvc.uab.es/LAMP/joost/BilateralFiltering/.

Results: Here we evaluate the accuracy and speed of the proposed bilateral filter approximations and compare them to several existing approximations. We compare our method to the methods from Chaudhury [9] and Paris and Durand [29]. These methods apply a naive color filter approximation. This involves applying the bilateral filter approximation separately to the three color channels (thereby ignoring the vectorial nature of color images). This is common practice, see for example [10], [29], [33]. Paris and Durand also provide a vectorial extension to the color domain but they observe that this extension is not viable for \( \sigma_s < 10 \). For the implementation of [29] we have used the default settings proposed by

\[\text{MATLAB code available from http://people.csail.mit.edu/jiawen/#code}\]

\[\text{The dragon image is significantly larger than the other images and is subsampled by a factor of two in all our experiments.}\]
The results of the bilateral filter approximation $\hat{f}$, which is based on a combination of the local prior and the cluster decomposition, are presented in Table I. We see that already for as few as four clusters the method outperforms the methods which are based on the naive bilateral filter [10], [29]. Our average results are almost $2\text{dB}$ better than the method of [2] when we use eight clusters per image (average PSNR over the six images is $50.9\text{dB}$ compared to $49.0\text{dB}$ for [2]) at a faster average processing time. As expected, $\hat{f}$ outperforms $\hat{f}$ for the same number of clusters at the cost of a small increase in time. To provide an insight in the PSNR as a function of the parameters we have provided the results for $\hat{f}$ with 4 clusters in Fig. 4(left). Given the growing availability of large image sizes, we have analyzed the performance and processing time of $\hat{f}$ as a function of the image size and number of clusters. The results are summarized in Table II. The $\hat{f}$ takes between 3 and 9 seconds depending on the desired accuracy for an image of 10 megapixels. Also note that the accuracy remains constant as a function of the image size.

In Fig. 4 (middle and right) we do a more detailed comparison between with the best state-of-the-art methods for color bilateral filtering, namely [2] and [18]. For various settings for the parameters of the bilateral filter we compare at equal processing time the quality of the approximation. This is done by picking for each parameter setting the maximum number of clusters for $\hat{f}$ such that the processing time of $\hat{f}$ remains below the processing time of [2], respectively [18]. The graph shows the difference between the average PSNR over the six images between the two methods. For this set of parameters our method always outperforms the method of [18]. Compared with [2] we obtain similar average PSNR at equal processing time, however the parameter settings for which the methods obtain good results vary. The approximation $\hat{f}$ does significantly better for small $\sigma_s$ whereas for larger $\sigma_s$ it is beneficial to use the permutohedral

\[
\text{TABLE I}
\]  

We report the average PSNR over a set of standard values for $\sigma_s$ and $\sigma_r$ (we consider $\sigma_s = \{2.5, 7.5, 10\}$ and $\sigma_r = \{10, 20, \ldots, 60\}$), yielding a total of 24 settings. Results are reported in dB. An $N$ between brackets signifies ‘naive implementation’ and refers to methods which originally were proposed for grayscale images and have been extended to color by applying it separately to the channels. Penultimate column shows average processing time per image in seconds, and the last column indicates if it is based on MATLAB (M) or C-code (C). In the bottom six rows we note the number of clusters which are applied between brackets.

<table>
<thead>
<tr>
<th>method</th>
<th>dragon</th>
<th>done</th>
<th>house</th>
<th>Polin</th>
<th>swamp</th>
<th>tulip</th>
<th>time</th>
<th>code</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="N">29</a></td>
<td>30.1</td>
<td>46.7</td>
<td>42.0</td>
<td>41.1</td>
<td>44.4</td>
<td>45.3</td>
<td>0.30</td>
<td>M</td>
</tr>
<tr>
<td><a href="N">29</a></td>
<td>34.7</td>
<td>41.8</td>
<td>38.9</td>
<td>37.3</td>
<td>37.9</td>
<td>40.1</td>
<td>0.85</td>
<td>M</td>
</tr>
<tr>
<td>[18]</td>
<td>41.9</td>
<td>49.6</td>
<td>45.6</td>
<td>43.1</td>
<td>45.2</td>
<td>47.1</td>
<td>0.62</td>
<td>M</td>
</tr>
<tr>
<td>[2]</td>
<td>45.9</td>
<td>51.6</td>
<td>48.1</td>
<td>47.5</td>
<td>50.5</td>
<td>50.5</td>
<td>0.27</td>
<td>C</td>
</tr>
</tbody>
</table>

\[
\text{TABLE II}
\]  

PSNR results and processing time (in s) for $\hat{f}$ as a function of number of clusters and image size. The $\sigma_r = 15\%$ and the $\sigma_s = \{40, 20, 10, 5\}$ descending with the image size. The input image is the Mammoth image shown in Fig. 8.

<table>
<thead>
<tr>
<th>clusters</th>
<th>4</th>
<th>8</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>image size</td>
<td>PSNR</td>
<td>time</td>
<td>PSNR</td>
</tr>
<tr>
<td>1.3M pixels</td>
<td>41.8</td>
<td>2.91</td>
<td>50.6</td>
</tr>
<tr>
<td>2.5M pixels</td>
<td>41.2</td>
<td>0.68</td>
<td>50.5</td>
</tr>
<tr>
<td>5M pixels</td>
<td>41.3</td>
<td>0.18</td>
<td>50.2</td>
</tr>
<tr>
<td>10M pixels</td>
<td>41.3</td>
<td>0.05</td>
<td>50.9</td>
</tr>
</tbody>
</table>

--

C code from authors available from http://graphics.stanford.edu/papers/permutohedral/

Matlab code from authors available from http://inf.ufrgs.br/~eslagastal/AdaptiveManifolds/  

C code from authors available from http://inf.ufrgs.br/~eslagastal/AdaptiveManifolds/

When applying naive bilateral color filtering we apply $0.75 \times \sigma_s$ on the separate channels. This was found to yield optimal performance on these six images.

\[5\text{C code from authors available from http://graphics.stanford.edu/papers/permutohedral/}

\[5\text{Matlab code from authors available from http://inf.ufrgs.br/~eslagastal/AdaptiveManifolds/}

\[5\text{When applying naive bilateral color filtering we apply } 0.75 \times \sigma_s \text{ on the separate channels. This was found to yield optimal performance on these six images.}
lattice approximation. This suggests that the ideas of [2] are complementary and combination with ours might lead to further improvements.

We include Fig. 5 and Fig.6 for an illustration of the very fast bilateral filter approximation $\tilde{F}^p$. Visual comparison shows that the results are very similar to the exhaustive bilateral filter, but slight differences can be observed for example on the petals. We also include illustrations of the bilateral filter approximations $\tilde{F}^{sp}$ with four and eight clusters for the test images Dragon and Mammoth in Fig. 7 and Fig. 8. Even for as few as four clusters the approximation of the bilateral filter obtains good visual results.

Finally, we perform an experiments for edge-aware detail enhancement. A brief introduction to detail enhancement is as follows (for more details see [15]). For a given input image $f$, its edge-preserving smoothed output $\tilde{f}$ is used as a base layer. The difference between the input signal and the base layer $f - \tilde{f}$ is considered as the detail layer, which is magnified to boost details. Then the enhanced signal is the combination of the boosted detail layer and the base layer. In general enhancement techniques work well with the bilateral filter, but its implementation can produce the gradient reversal artifacts. In Fig. 9 (a) we see that for the Tulip image the true bilateral filter produces the gradient reversal halos. Near some edges (the largest zoom at the bottom of the images) these thin stroke-like artifacts do not correspond to any real details. Implementation of our approximate filter produces almost the same output as the true bilateral filter for $K > 30$, resulting in the same artifacts. Surprisingly, the less clusters the less gradient reversal artifacts are observed in the filter output (see Fig. 9 (b) with $K = 3$). Finally, our approximation based on the local prior ($\tilde{F}^p$) does not produce any gradient reversal artifacts. Of course, choosing between the results is a subjective choice but when running time is an important factor, $\tilde{F}^p$ provides a good alternative for edge-aware detail enhancement.

VII. Conclusions

In this paper we have presented two fast approximations to bilateral filtering for color images. The first approach is based on the color sparseness of images: even though there exist a large number of colors in the world, in a single image only a limited set of colors are present. To exploit color sparseness, we propose the cluster decomposition which allows to approximate the non-linear bilateral filter with a limited number of linear filters.

Our second approach is based on a local statistics prior, which assumes that the image values within the filter window form a uniform distribution which is described completely by its mean and center pixel value. This assumption leads to a closed-form solution for the bilateral filter which results in an extremely fast approximation of the bilateral filter. To the best of our knowledge, this is the first method which exploits a prior on the local distribution for fast image filtering. The filter takes 0.10 seconds on a Intel Xeon 3.6 GHz CPU for a one Megapixel image. Even though the accuracy is limited, this approximation could be interesting for application where speed is crucial. It would be interesting to further investigate this approach: considering other distributions which could lead to closed-form solutions, and applying the same principle to other filter operations.

Our bilateral filter which combines both ideas, i.e. local prior and color sparsity, is shown to obtain similar results as the bilateral filter based on permutohedral lattice [2]. A closer analysis shows that our method is faster for applications which require smaller spatial scale. As future work we are interested to combine our work with ideas from the work of [2], which is expected to lead to further speed-ups. Finally, we are interested in applying the theory to speed-up computation of non-local means, as studied in [2], [18].

ACKNOWLEDGEMENTS

This work is funded by the Projects TIN2013-41751 and TIN2012-39051 of the Spanish Ministry of Science and the Catalan project 2014 SGR 221. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the GPU used in this research. We acknowledge Andrew D. Bagdanov for his advice on GPU implementation of the algorithm.

APPENDIX

Here we provide the derivation of the bilateral filter approximation based on color sparseness as proposed in Eq. 5. By
combining Eq. 3 and Eq. 4 into Eq. 1 we obtain,
\[
\hat{f}(x) = \eta^{-1}(x) \sum_{y \in \Omega} \theta(|r(x) - r(x-y)|) \varphi(|y|) f(x-y) \\
= \eta^{-1}(x) \sum_{y \in \Omega} \theta \left( |r(x) - \sum_{k \in K} m_k (x-y) \mu_k| \right) \cdot \varphi(|y|) \sum_{k \in K} m_k (x-y) f(x-y)
\] (24)
Note that the following holds:
\[
m_k (x-y) \theta \left( |r(x) - \sum_{k \in K} m_k (x-y) \mu_k| \right) = m_k (x-y) \theta \left( |r(x) - \mu_k| \right)
\] (25)
which is based on the observation that for all pixels for which \(m_k (x-y) = 1\) it holds that
\[
\theta \left( |r(x) - \sum_{k \in K} m_k (x-y) \mu_k| \right) = \theta \left( |r(x) - \mu_k| \right)
\] (26)
and for the pixels for which \(m_k (x-y) = 0\) holds, Eq. 25 is evidently true. Applying this to Eq. 24 we obtain the equation for the bilateral filter approximation \(\hat{f}_a\):
\[
\hat{f}_a(x) = \eta^{-1}(x) \sum_{y \in \Omega} \theta \left(|r(x) - \mu_k|\right) \varphi(|y|) \hat{f}_k(x-y) \\
= \eta^{-1}(x) \sum_{y \in \Omega} \theta \left(|r(x) - \mu_k|\right) \hat{f}_k(x)
\] (27)
and we obtain Eq. 5. In the second step of Eq. 27 we use the fact that \(\hat{f}_k(x) = \sum_{y \in \Omega} \varphi(|y|) f_k(x-y)\). A similar derivation yields the equation for \(\eta(x)\) in Eq. 5.

**REFERENCES**

Fig. 7. Illustration of the filtering results for the Dragon test image (courtesy of S. Paris and F. Durand) with the parameters $\sigma_s = 10$ and $\sigma_r = 10\%$.

(left) the input image; (middle) the output of the bilateral filter with the exhaustive implementation; (right) the output of the bilateral filter $\hat{f}_{sp}$ with our implementation $K = 4$.

Fig. 8. Illustration of the filtering results for the Mammoth test image (968x648) with the parameters $\sigma_s = 10$ and $\sigma_r = 15\%$.

(left) the input image; (middle) the output of the bilateral filter with the exhaustive implementation; (right) the output of the bilateral filter $\hat{f}_{sp}$ with our implementation $K = 8$.


Fig. 9. Example of color detail enhancement with the parameters $\sigma_s = 15$ and $\sigma_r = 10\%$. The detail layer is boosted by a factor of 5. (left) True bilateral filter enhances image features; (middle) Our approximation based on $\hat{f}^{sp}$ with $K = 3$; (right) Our approximation based on $\hat{f}^p$ does not suffer from gradient reversal artifacts.


Mikhail G. Mozerov is a Senior Scientist at the Computer Vision Center Barcelona. He received his MS degree in physics from the Moscow State University, in 1982, and his PhD degree in digital image processing from the Institute of Information Transmission Problems, Moscow, in 1995. From 2006 to 2010, he was a Ramon y Cajal Fellow at the Universidad Autonoma de Barcelona. His research interests include signal and image processing, stereo and optical flow, and pattern recognition.

Joost van de Weijer is a Senior Scientist at the Computer Vision Center Barcelona and member of the LAMP team. He received the Ph. D. degree in 2005 from the University of Amsterdam. From 2005 to 2007, he was a Marie Curie Intra-European Fellow in the LEAR Team, INRIA Rhone-Alpes, France. From 2008 to 2012, he was a Ramon y Cajal Fellow at the Universidad Autonoma de Barcelona. His main research is on the usage of color information in computer vision application.