# TRAINED BILATERAL FILTERS AND APPLICATIONS TO CODING ARTIFACTS REDUCTION

*H. Hu*, and G. de Haan*†*

*Eindhoven University of Technology, Den Dolech 2,
5600 MB Eindhoven, the Netherlands
†Philips Research Laboratories, High Tech Campus 36,
5656 AE Eindhoven, the Netherlands

## ABSTRACT

Bilateral filtering is a simple and non-linear technique to remove the image noise while preserving edges. However, it is difficult to optimize a bilateral filter to obtain desired effect by supervised training. In this paper, we propose a new type of trained bilateral filter, which possesses the essential characteristics of the original bilateral filter and can be optimized offline by Least Mean Square optimization. In applications of JPEG and H.264/MEPG4 AVC deblocking, we compared the proposed filter with the original bilateral filter and other state-of-the-art methods. Experimental results show that the proposed method has a better performance at artifacts reduction and edge preserving.

***Index Terms*—** Nonlinear filters, Adaptive filters, Image processing, Image restoration, Image enhancement.

## 1. INTRODUCTION

As a non-iterative and simple filtering technique, bilateral filtering [1] has received considerable attention in areas of image processing and computer vision. Unlike the conventional linear filters of which the coefficients are pre-determined, the bilateral filter adjusts its coefficients to the geometric closeness and photometric similarity of the pixels. Due to this adaptivity, it has shown good performance at edge-preserving smoothing for image restoration applications, such as noise reduction and digital coding artifacts reduction [2].

In applications of the bilateral filter, a Gaussian function has been typically used to relate coefficients to the geometric closeness and photometric similarity of the pixels, which seems somewhat arbitrary. For a linear filter, its coefficients can be adjusted to achieve desirable effects by supervised training and Least Mean Square optimization. However, this is not trivial for the bilateral filter. In order to solve that problem, we proposed in the paper a new type of *trained bilateral filters*. The proposed trained bilateral filter adopts a linear combination of spatially ordered and rank ordered pixel samples, which has proposed in a hybrid filter [3]. Different from the hybrid filter where the similarity had been heavily quantized, the rank ordered pixel samples in the proposed method are further transformed to reflect the photometric similarity of the pixels. Consequently, the trained bilateral filter possesses the essential characteristics of the original bilateral filter. On the other hand, the design of the proposed bilateral filter makes it feasible to optimize the filter coefficients. That is, the optimal coefficients for the combined pixel samples can be obtained by Least Mean Square optimization as the linear filters.

In the paper, to illustrate the effectiveness of the new concept, we also show the applications of the trained bilateral filter to reduce coding artifacts for JPEG and H.264/ MEPG4-AVC [5] standards. In these applications, the local image structure information is utilized for classifying coding artifacts and true image edges. Then a trained bilateral filter, optimized for every class, is used to remove coding artifacts and preserve true image edges. The optimal filter coefficients are obtained by training the filters on original images and their compressed versions. To evaluate the proposed bilateral filter, the comparison with the original bilateral filter and other state-of-the-art methods are presented.

The paper is organized as follows. Section 2 gives the definition of the trained bilateral filter. In Section 3, we show its application to coding artifacts reduction. The experimental results of the evaluation for the application are shown in Section 4. Finally, in Section 5, we draw our conclusions.

## 2. TRAINED BILATERAL FILTER

Suppose $X = (x_1, x_2, ..., x_N)^T$ is a vector containing $n$ pixels arranged by the spatial order within a filter aperture. The output $y_b$ of a bilateral filter is defined by:

$$y_b = \frac{\sum_{i=1}^{N} x_i \cdot c(x_i, x_c) \cdot s(x_i, x_c)}{\sum_{i=1}^{N} c(x_i, x_c) \cdot s(x_i, x_c)}$$

$$\begin{aligned} s(x_i, x_c) &= exp[-(x_c - x_i)^2 / 2\sigma_s^2] \\ c(x_i, x_c) &= exp[-d(x_c, x_i)^2 / 2\sigma_c^2] \\ i &= 1, 2, ..., N. \end{aligned} \tag{1}$$

where $d(x_c, x_i)$ is the Euclidean distance between the pixel position of $x_i$ and $x_c$.

The output $y_l$ of a linear filter is:

$$y_l = W^T X. \tag{2}$$

where $W$ is an $N \times 1$ vector of weights. Consequently, the linear filter only takes consideration of the geometric closeness of the pixel samples.

To include the rank order information, the hybrid filter has been introduced [3]. It orders the samples in the vector $X$ according to their rankings and obtains the vector $X_r = (x_{(1)}, x_{(2)}, ..., x_{(N)})^T$. Then the concatenation of $X$ and $X_r$ gives a vector $X_h$ and the output of the hybrid filter $y_h$ is:

$$y_h = W_h^T X_h. \tag{3}$$

where $W_h$ is an $2N \times 1$ vector of weights.

However, the rank ordering in the hybrid filter only gives some indications of the pixel similarity. In order to incorporate the complete similarity information as the original bilateral filter does, we obtain the vector $X_s = (x_{[1]}, x_{[2]}, ..., x_{[N]})^T$ by sorting the pixels according to their pixel value distance to the spatially central pixel $x_c$ in the aperture. The ordering is defined by:

$$|x_{[i+1]} - x_c| \geq |x_{[i]} - x_c|, i = 1, 2, ..., N. \quad (4)$$

Then we transform the vector $X_s$ into $X_s\prime = (x_{[1]}\prime, x_{[2]}\prime, ..., x_{[N]}\prime)^T$. The transform is defined as:

$$x_{[i]}\prime = \mu(x_c, x_{[i]}) \cdot x_c + (1 - \mu(x_c, x_{[i]})) \cdot x_{[i]}, i = 1, 2, ..., N. \quad (5)$$

where $\mu(x_c, x_{[i]})$ is a membership function between $x_{[i]}$ and $x_c$. The membership function is defined as:

$$\mu(x_c, x_{[i]}) = MIN(\frac{|x_{[i]} - x_c|}{K}, 1). \quad (6)$$

where $K$ is a pre-set constant. Other membership functions such as a Gaussian function are also possible. The vector $X_{tb}$ is obtained by concatenating the vectors $X$ and $X_r\prime$:

$$X_{tb} = (x_1, x_2, ..., x_n, x_{[1]}\prime, x_{[2]}\prime, ..., x_{[N]}\prime)^T. \quad (7)$$

Similar to the linear filter, we define the output of the proposed trained bilateral filter is:

$$y_{tb} = W_h^T X_{tb}. \quad (8)$$

where $W_{tb}$ is an $2N \times 1$ vector of weights. The advantage of the trained bilateral filter is that, in the presence of true image structures, the weights of the transformed samples that are similar to the center sample value are increased to better preserve edges. On the other hand, when low-pass filtering is needed to suppress the blocking artifacts, the weight of the linear part can be adjusted accordingly. Essentially the trained bilateral filter behaves as the original bilateral filter whose coefficients are continuously dependent on the spatial and intensity difference of pixels.

The optimization of the trained bilateral filter can be accomplished in a similar fashion as for the linear filter. Suppose the output of the trained bilateral filter $y_{tb}(t) = W_{tb}^T X_{tb}(t)$ is used to estimate the desired signal $d(t)$. The optimal filter coefficients are obtained when the mean square error between the output and desired signal is minimized. The mean square error $MSE$ is:

$$MSE = E[(y_{tb}(t) - d(t))^2] = E[(W_{tb}^T X_{tb}(t) - d(t))^2]. \quad (9)$$

Taking the first derivative with respect to the weights and setting it to zero, we obtain [3]:

$$W_{tb}^T = E[X_{tb} X_{tb}^T]^{-1} E[X_{tb} d]. \quad (10)$$

## 3. APPLICATIONS TO CODING ARTIFACTS REDUCTION

In the applications to coding artifacts reduction, we use the local image structure information to classify coding artifacts and true image edges as [4]. The diagram of the proposed scheme is shown in Fig. 1. A diamond shaped filter aperture centered at the output pixel is used. The local image structure within the filter aperture is first classified by Adaptive Dynamic Range Coding (ADRC)[6] and local contrast information. The optimal linear filter is used to calculate the output pixel with filter coefficients obtained from the look-up-table (LUT). The filter aperture slides pixel by pixel over the entire image.

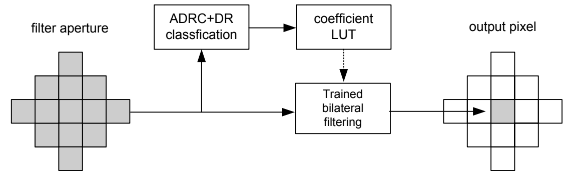



**Fig. 1**. The block diagram of the proposed scheme: the local image structure is classified using pattern classification and the filter coefficients are obtained from the LUT.

### 3.1. Local structure classification

ADRC is a simple and efficient way to classify local image structures. The 1-bit ADRC code of every pixel is defined by:

$$ADRC(x_i) = \begin{cases} 0, & if\ x_i < \frac{x_{max} + x_{min}}{2} \\ 1, & otherwise \end{cases} \quad (11)$$

where $x_i$ is the value of pixels in the filter aperture and $x_{max}$, $x_{min}$ are the maximum and minimum pixel value in the filter aperture. We find that ADRC classification is not enough to distinguish between coding artifacts and real image structures. For example, a real image edge and a boundary of blocking artifact can have an identical ADRC pattern. Clearly, image areas with regular patterns and high contrast are more likely to be the real image edges while image areas with regular patterns and low contrast suggest possible coding artifacts. Therefore, we add one extra bit, $DR$, to the ADRC code. The extra bit includes the contrast information in the aperture. The extra bit is defined as:

$$DR = \begin{cases} 0, & if\ x_{max} - x_{min} < Tr \\ 1, & otherwise \end{cases} \quad (12)$$

where $Tr$ is the pre-set threshold value related to the coding quantization. The concatenation of $ADRC(x_i)$ of all pixels in the filter aperture and the extra bit $DR$ gives the class code, used to address the coefficient look-up table. The number of classes given by ADRC code can be reduce by bit inversion [6] to $2^{N-1}$. Together with the extra bit $DR$, the number of classes becomes $2^N$.

### 3.2. The training procedure

The optimization procedure of the proposed method is shown in Fig. 2. To obtain the training set, we use original images as the reference output image. Furthermore, we compress the original images with the expected compression ratio. These corrupted versions of the original images are our simulated input images. Before training, the simulated input and the reference output pairs are collected pixel by pixel from the training material and are classified using the mentioned classification method on the input. The pairs that belong to one specific class are used for the corresponding training, resulting in optimal coefficients for this class.

## 4. EXPERIMENTS AND RESULTS

In this section, we will evaluate the proposed method (referred as *Pro-tb*) in the applications of JPEG de-blocking and MPEG4-AVC/ H.264 de-blocking. For JPEG de-blocking, we choose Nosratinia's method [7] (referred as *Nos*) as the comparison. For MPEG4-AVC/

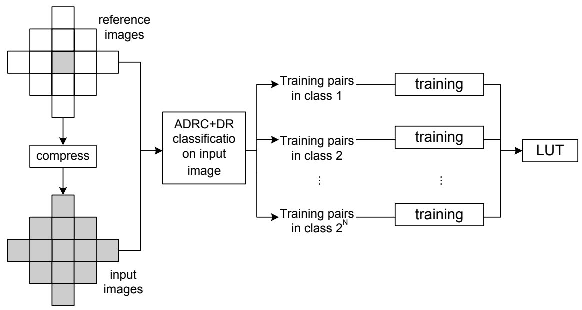



**Fig. 2**. The training procedure of the proposed algorithm. The input and output pairs are collected from the training material and are classified by the mentioned classification method. The filter coefficients are optimized for specific classes.

**Table 1**. MSE scores for JPEG deblocking

| | Mean Square Error | | | | | |
|---|---|---|---|---|---|---|
| Image | Input | *Nos* | *Bil* | *Pro-l* | *Pro-h* | *Pro-tb* |
| Bicycle | 63.5 | 51.0 | 50.7 | 46.4 | 44.6 | 41.3 |
| Birds | 13.8 | 11.1 | 11.7 | 11.1 | 10.9 | 10.5 |
| Boat | 72.3 | 63.7 | 62.1 | 61.9 | 60.7 | 58.6 |
| Motor | 108.9 | 93.9 | 92.9 | 90.2 | 86.9 | 82.8 |
| Lena | 36.9 | 30.4 | 31.5 | 29.2 | 29.3 | 28.6 |
| **Average** | 59.1 | 50.0 | 49.8 | 47.8 | 46.5 | 44.4 |

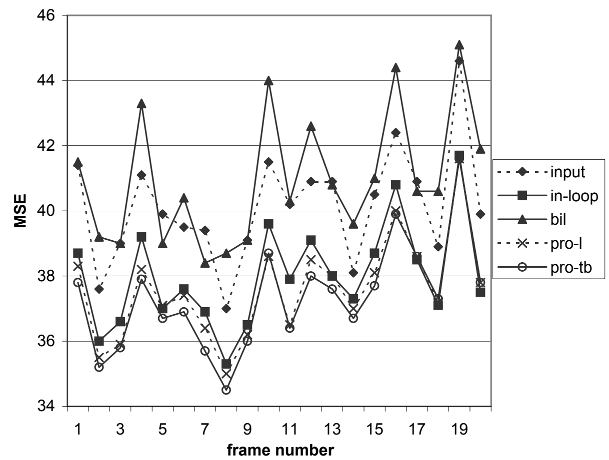
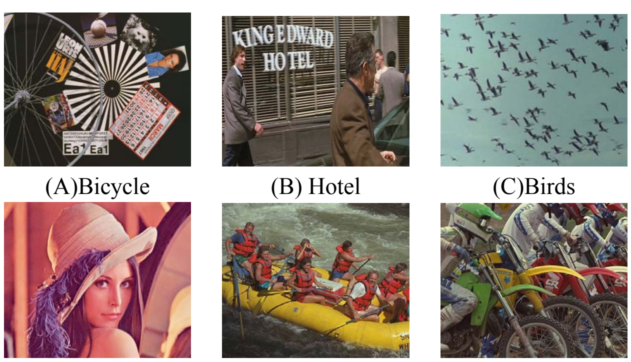



**Fig. 4**. MSE results of H.264 de-blocking for Hotel sequence

H.264, we compared our proposed method with the in-loop filter used in the standard. In order to provide a more complete evaluation, we also include the original bilateral filter (referred as *Bil*) and alternative methods which uses the same classification as the proposed scheme but with linear filters (referred as *Pro-l*) and hybrid filters (referred as *Pro-h*).

### 4.1. Test Material

In order to enable a quantitative comparison, we first compress the original uncompressed images using the same setting as in the training procedure but with different images. Then we use the images as the input test images. The Mean Square Error (MSE) can be calculate from the original uncompressed images and the processed images. The test images are shown in Fig. 3.




(A)Bicycle (B) Hotel (C)Birds

(D)Lena (E)Boat (F)Motor

**Fig. 3**. The testing material used for evaluation.

### 4.2. JPEG de-blocking

We apply JPEG compression at quality factor 20 (quality factor 100 is the best). The free baseline JPEG software from the Independent JPEG Group website (http://www.ijg.org/files/jpegscrc.v6b.tar.gz) has been used for JPEG encoding and decoding. Table 1 shows the MSE comparison of the evaluated methods. In term of MSE score, one can see that *Pro-l*, *Pro-h* and *Pro-tb* all outperform *Nos*, especially in the sequence Bicycle , which contains various image structures. Overall *Pro-tb* gives the best result, which is about 25 percent improvement over the input on the average.

To enable a subjective comparison, image fragments from the Motor sequence processed by these methods are shown in Fig. 5. As one can see, all the methods can reduce the coding artifact significantly. However, *Pro-l*, *Pro-h* and *Pro-tb* repair the distorted edges much better than other methods because they adapt image structure information. *Pro-tb* shows better edge and details preservation than *Pro-l* and *Pro-h* and it produces the processed image closest to the original.

### 4.3. H.264/MPEG4 AVC deblocking

The H.264/MPEG4 AVC compression we used for evaluation is the reference software implementation, JM 11.0 system (http://iphome.hhi.de/suehring/tml/download/). The in-loop filter is included in JM 11.0 system. The compression quality parameter $QP$, which has a range from 0 to 51, is set to 35. The MSE results for the sequence Hotel are shown in Fig. 4. Both *Pro-l* and *Pro-tb* give a better performance than the in-loop filter in every frame. And *Pro-tb* has a slightly lower MSE score, compared with *Pro-l*. The image fragments in Fig. 6 show that although the in-loop filter can reduce most of the blocking artifacts significantly in the flat area due to its in-loop advantage, it does not repair the corrupted edge structures well which can be nicely reconstructed with the post-processing methods *Pro-l* and *Pro-tb*.


**Fig. 5**. Image fragments from Motor image: (A) original uncompressed (B) JPEG compressed (C) Bilateral filter (D) Nosratinia's method (E) *Pro-l* (F) *Pro-h* (G) *Pro-tb*

## 5. CONCLUSION

In this paper, we have introduced a new type of trained bilateral filter. The proposed filter incorporates pixel similarity and closeness information to adapt the coefficients as the original bilateral filter, and it can be easily optimized by Least Mean Square optimization. In the application to digital coding artifacts reduction, the trained bilateral filter significantly increases the perceptual quality of the processed image by removing the artifacts and well preserving the edges. It has shown superior performance and appears promising in different applications.

## 6. REFERENCES

[1] C. Tomasi and R. Manduchi,"Bilateral Filtering for Gray and Color Image", *Proceedings of the 1998 IEEE International Conference on Computer Vision*, Bombay, India.


**Fig. 6**. Image fragments from Hotel sequence: (A) original uncompressed (B) H.264 compressed (C) *Bil* (D) In-loop filter (E) *Pro-l* (F) *Pro-tb*

[2] S. Wan, M. Mrak and E. Izquierdo, "Perceptually Adaptive Joint Deringing Deblocking Filtering for Scalable Video Coding", *2nd International Mobile Multimedia Communications Conference*, Alghero, Sardinia, Italy, Sept. 2006.

[3] H. Hu and G. de Haan,"Classification-based hybrid filters for image processing", *Proceedings of SPIE, Visual Communications and Image Processing*, San Jose, USA, Jan. 2006.

[4] H. Hu and G. de Haan,"Simultaneous coding artifacts reduction and sharpness enhancement", *IEEE International Conference on Consumer Electronics*, Las Vegas, USA, Jan. 2007.

[5] ITU-T Recommendation and International Standard of Joint Video Specification (ITU-T Rec. H.264/ISO/IEC 14 496-10: 2005), Dec. 2005.

[6] T. Kondo, Y. Fujimori, S. Ghosal and J. J. Carrig, " Method and apparatus for adaptive filter tap selection according to a class", US-Patent: US 6,192,161 B1, Feb. 20, 2001.

[7] A.Nosratinia,"Denoising of JPEG images by re-application of JPEG," *Journal of VLSI Signal Processing*, vol. 27, no. 1, pp. 6979, 2001.