



CDDA: color-dominant deep autoencoder for faster and efficient bilateral image filtering

Apurba Das¹ · S. S. Shylaja¹

Received: 11 June 2019 / Revised: 2 November 2020 / Accepted: 30 December 2020 / Published online: 9 March 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd. part of Springer Nature 2021

Abstract

Nonlinear processing of high-dimensional data is quite common in image filtering algorithms. Bilateral, joint bilateral, and non-local means filters are the examples of the same. Real-time implementation of high-dimensional filters has always been a research challenge due to its computational complexity. In this paper, we have proposed a solution utilizing both color sparseness and color dominance in an image which ensures a faster algorithm for generic high-dimensional filtering. The solution speeds up the filtering algorithm further by psycho-visual saliency-based deep encoded dominant color gamut, learned for different subject classes of images. The proposed bilateral filter has been proved to be efficient both in terms of psycho-visual quality and performance for edge-preserving smoothing and denoising of color images. The results demonstrate competitiveness of our proposed solution with the existing fast bilateral algorithms in terms of the CTQ (critical to quality) parameters.

Keywords Range-domain filtering · Deep autoencoder · weighted histogram · fast algorithm · color sparseness

1 Introduction

In low-level image information processing as well as larger domain of computer vision and analytics, image smoothing has been a major and fundamental aspect of research which took a major turn during last couple of decades. The serious impact on edge information on images has been observed for classical image enhancement filters. Range-domain or bilateral filter [1,2] on the contrary ensures edge preserved noise cleaning and smoothing by addressing range and domain of the image in single-stage honoring the photometric dominance/ prominence of color and gray intensity values in one hand and spatial neighborhood in other. Additionally, bilateral filter ensures no phantom color as the side-effect of filtering which is quite common in classical 2D filtering performed in color separations separately [3]. Bilateral filters are used even in enhancing hyper-spectral imaging [4,5]. Various applications of bilateral filter include tone mapping,

denoising, detail manipulation, upsampling, alpha matting, recoloring, stylization, etc. Since the bilateral filter concept was proposed by Tomasi et al. [1], it has been a major area of contribution in low-level image processing community due to its accurate, high-quality denoising and slow performance. Consider a high-dimensional image $f : \mathbb{Z}^d \rightarrow \mathbb{R}^n$ and a *guide* image $p : \mathbb{Z}^d \rightarrow \mathbb{R}^\rho$. Here d is dimension of domain, n and ρ are dimensions of ranges of the input image f and guide p , respectively. The output of the bilateral filter $h : \mathbb{Z}^d \rightarrow \mathbb{R}^n$ is given as:

$$h_i = \frac{1}{k_i} \sum_{j \in W} \omega(j) \phi(p_{i-j} - p_i) f_{i-j}, \quad (1)$$

where

$$k_i = \sum_{j \in W} \omega(j) \phi(p_{i-j} - p_i). \quad (2)$$

The aggregation in Eqs. 1 and 2 are typically performed over a hypercube around the pixel of interest, i.e., $W = [-S, S]^d$, where the integer S is the window size. The terms ω and ϕ are the domain and range kernels, respectively. The complexity of the aforementioned filtering operations is $O(nS^d)$ per pixel. Hence, for large d and n , it is really challenging to

✉ Apurba Das
apurba_das1@hotmail.com
S. S. Shylaja
shylaja.sharath@pes.edu

¹ Department of Computer Science and Engineering, PES University, Bangalore, India

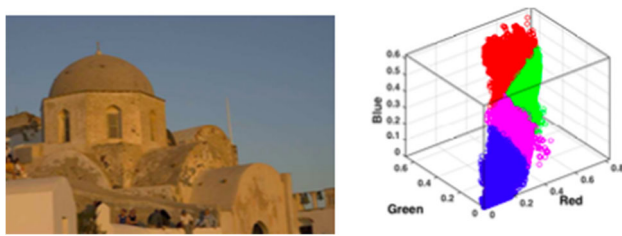


Fig. 1 Distribution of color of the natural image (left side) in the color cube (right side). It is sparse, non-uniform and covers a minimal subset of the entire dynamic range [5]

design real-time high-dimensional filtering. Being motivated by Durand et al. and Yang et al. [6–8], where range space is quantized to approximate the filter using series of fast spatial convolutions, Nair et al. [5] have proposed an algorithm based on clustering of sparse color space. The aforementioned work has observed color sparseness (Fig. 1) in range space and proposed a clustering in high dimensions. The work has approximated p_{i-j} instead of p_i [9] resulting in a completely new algorithm.

Our current work has proposed an algorithm which works in parallel with the existing filtering operations to further improve the performance of high-dimensional bilateral filtering by deriving salient color gamut in an image class of interest by our proposal of CDDA (color-dominant deep autoencoder). This encoded dominant color map (DCM) from CDDA has been designed to be used offline by the principal flow of the filtering operations as prescribed by Nair et al. [5]. Recently proposed convolutional pyramid model [10] and deep bilateral filter [11] for image enhancement have also been shown promising results in range-domain filtering. The algorithm of building DCM from proposed CDDA has been described in Sect. 2 in terms of detailed methodology and pseudo-code. Next, Sect. 3 has described the method of inclusion of DCM evolved out of our algorithm to the principal flow of fast bilateral filtering ensuring further improvement in performance. Section 4 has demonstrated the competitive evaluation of the proposed filter with CDDA in terms of accuracy and performance for different classes of images. Finally, in Sect. 5 we have concluded our findings with a direction to the future research.

2 Color-dominant deep autoencoder (CDDA) leveraging color sparseness and salience

In the current work, we have leveraged the property of dimension reduction of autoencoder [13] to extract dominant color from the larger color gamut present in any image. The idea of sparse color occupancy for any group of images depicting same object or action has been utilized further to create DCM (dominant color map) offline as a table to be referenced in

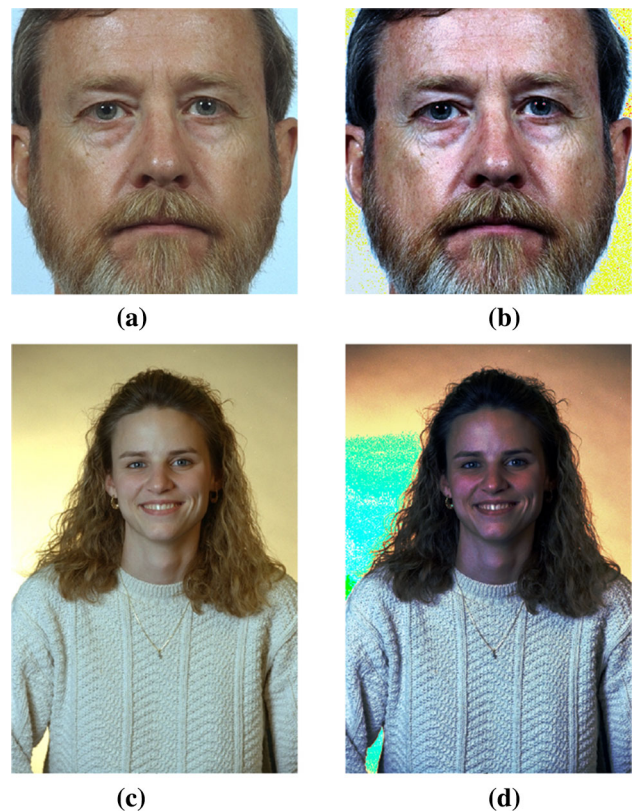


Fig. 2 CDDA maintains face color gamut and compromising on the other colors like background as required: **a, c** original input face image [12], **b, d** recreating face image from the encoded reduced color gamut having 75% compressed DCM

real-time. The offline DCM table has been next used as LUT for real-time processing, ensuring much faster bilateral filtering with respect to the state of art (Fig. 2).

With the advancement of deep learning, classification inaccuracy has been drastically reduced. Hence, there have been attempts of designing deep bilateral filters [11] in order to improve accuracy of restoration from corrupted images. We know neural network ensures optimum weight for optimum error (i.e., cost function) surface.

2.1 Evolution of DCM through CDDA

The principal idea here is to determine dominant/ salient colors from group of images of homogeneous kind (e.g., faces [12]). The dominant colors might be even interpolated colors of the quantized available color in the group of images. The autoencoder architecture as depicted in Fig. 3 has been employed to determine the salient/ dominant color for different groups of images of homogeneous kind (e.g., faces [12]) at a time with an objective to derive dominant color map in a coded and reduced dimension format, offline. This DCM further would be processed during image filtering. The proposed method of DCM derivation has five following stages:

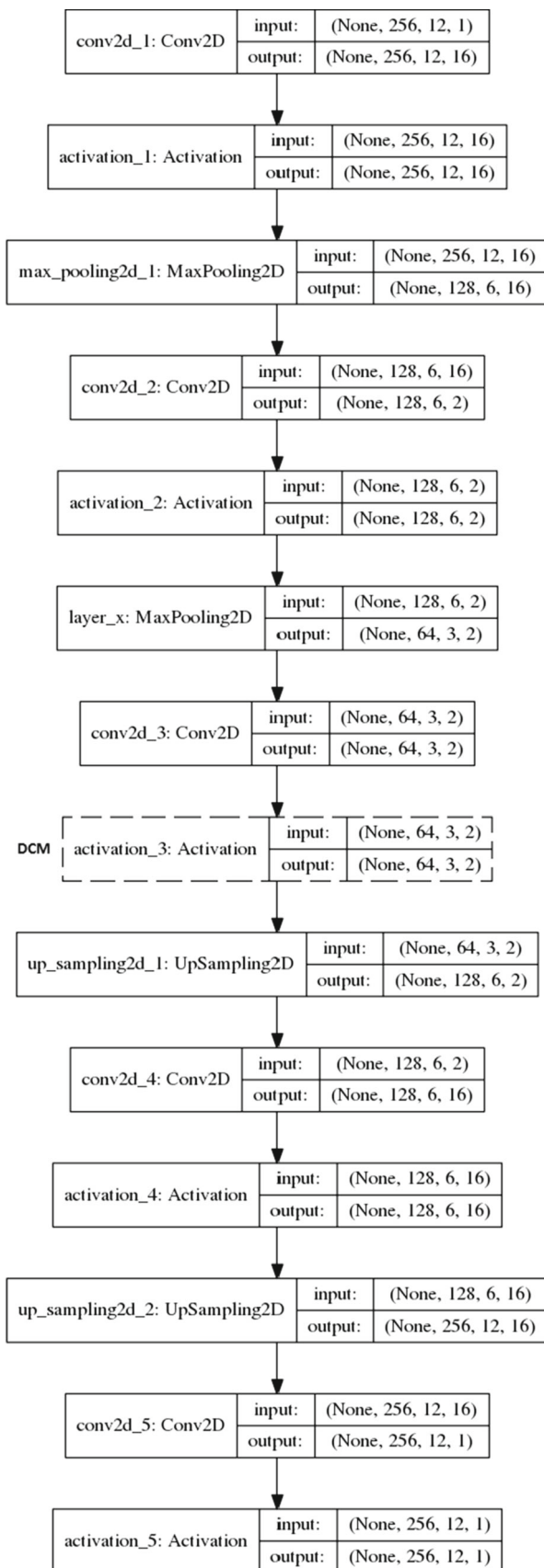


Fig. 3 CDDA: color-dominant deep autoencoder architecture to create the DCM

1. Imagification of weighted histogram as input to the autoencoder (CDDA).
2. Unsupervised learning of dominant color map from large number of images of similar designation for 1000 epochs.
3. Validating the converged DCM for unseen query image of decided arbitrary designation.
4. Hyper-parameter tuning and retraining the CDDA if the result of the previous stage is unsatisfactory.
5. Freezing the CDDA as offline look-up table (LUT) to be referred for primary path of near real-time bilateral image filtering.

In order to make the bilateral image filtering, it is important to identify dominant color for defined number of color clusters from the sparse color occupancy in the entire color gamut. The process of histogram imagification and extracting DCM are the activities to achieve the aforementioned target. First, the histogram of Red, Green, and Blue color separations are calculated and normalized between 0 and 255 to be represented as image as shown in Fig. 4. The representation has been depicted in Eq. 3.

$$red = count\ Of(I(:, :, 1)) \tag{3a}$$

$$green = count\ Of(I(:, :, 2)) \tag{3b}$$

$$blue = count\ Of(I(:, :, 3)) \tag{3c}$$

$$Imghist(:, 1 : 4) = red \times 256 \frac{red(:, :) - min(red)}{max(red) - min(red)} ones(4, 256) \tag{3d}$$

$$Imghist(:, 5 : 8) = green \times 256 \frac{green(:, :) - min(green)}{max(green) - min(green)} ones(4, 256) \tag{3e}$$

$$Imghist(:, 9 : 12) = blue \times 256 \frac{blue(:, :) - min(blue)}{max(blue) - min(blue)} ones(4, 256) \tag{3f}$$

As described in Eq. 3, the color histogram has been imagified and repeated 4 times as 4 columns of $Imghist(:, :)$ to enable the imagified weighted histogram to be consumed by the autoencoder (Fig. 3). The autoencoder just expects the group of homogeneous kind of images as input and attempts to reconstruct the same. The input image shouldn't have mixture of different kinds of images. In Fig. 4, we have shown it for face images/ videos and the same is applicable to even medical images (e.g., pathology, laparoscopy, etc.). We have used the same DCM for cleaning endoscopic images and the results are shown in Fig. 5. Only one example face image and its imagified histogram is shown in Fig. 4. For training the autoencoder to extract dominant color map (DCM) for faces, 10,000 face images have been used for training samples. As depicted in the autoencoder architecture, the encoded form (DCM) has dimension 64×3 which is the reduced dimen-

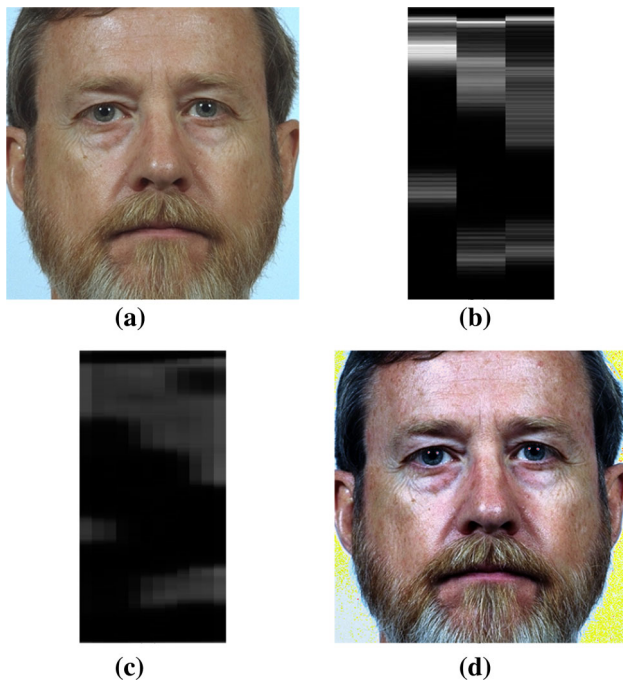


Fig. 4 Validating the converged DCM for unseen query image of decided “face” designation: **a** query face image [12], **b** imagified weighted histogram of query face, **c** reconstructed imagified histogram of the query face, **d** reconstructed face through CDF linearization of imagified histogram to validate DCM (not though bilateral filtering)

sion from original dimension of 256×3 . This dimensionality reduction is exactly 75% and the same could be reconstructed from the DCM as depicted in Fig. 4c which could even reconstruct the face image by CDF linearization idea described by Das et al. [3]. In this case, the number 64 could be treated as number of clusters having dominant encoded color of the selected class (e.g., face) images. The same can be interpolated to any other number of clusters. Figure 4 also shows that the compromise of color is at the background of the scene, not at the face region. The reconstructed histogram (Fig. 4c) has similar relative pattern as input imagified histogram (Fig. 4b). This is only in the verification stage through CDF linearization method. The objective here is to show how the offline look-up table is formed based on dominant color or salient color. Figure 4 is not showing the bilateral filtered output. This shows, even from large number of image groups, it could extract the salient sparse color and face color is still intact. Decoding the converged histogram in turn has been validated through CDF linearization where it is clearly shown that the salient color (face color) is not deteriorated although maintaining same background color is non-salient with respect to the objective of creating dominant color map. In bilateral filters, even the background would not affect the same.

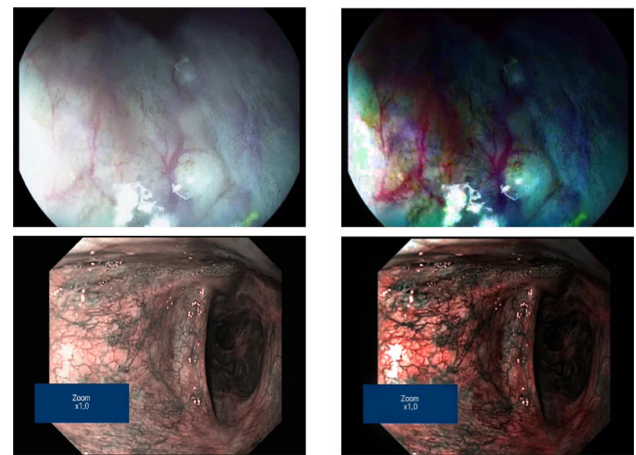


Fig. 5 Dynamic bilateral filtering on endoscopic video [14]: left column: 2 sample frames as inputs; right columns: Corresponding outputs

3 Inclusion of DCM into principal flow of bilateral filtering

Based on clustering of sparse color space, Durand et al. [6] and Yang et al. [7,8] have proposed to quantize the range space to approximate the filter using series of fast spatial convolutions. Motivated by the aforementioned work, Nair et al. [5] has proposed an algorithm based on clustering of sparse color space. There, the idea is to perform high-dimensional filtering on a cluster-by-cluster basis. For K number of clusters, where $1 \leq k \leq K$,

$$h_k(i) = \sum_{j \in W} \omega(j) \phi(p_{i-j} - \mu_k) f_{i-j} \quad (4)$$

$$\alpha_k(i) = \sum_{j \in W} \omega(j) \phi(p_{i-j} - \mu_k) \quad (5)$$

Here, Eqs. 4 and 5 are representing numerator and denominator of Eqs. 1 and 2 replacing p_i by the cluster centroids μ_k . The scheme of hybridizing online and offline processing is depicted in Fig. 6. As the algorithm to construct color LUT is working offline and the principal flow of filtering is operated real-time, the performance has been improved significantly as presented in Sect. 4.

4 Experimental results

In the current section, we would demonstrate the effectiveness of our novel approach of faster bilateral filtering of color images. For the bilateral filter, f and p are identical, and the spatial and range kernels are Gaussian (Eqs. 6, 7):

$$\omega(x) = \exp\left(-\frac{\|x\|^2}{2\sigma_s^2}\right) \quad (6)$$

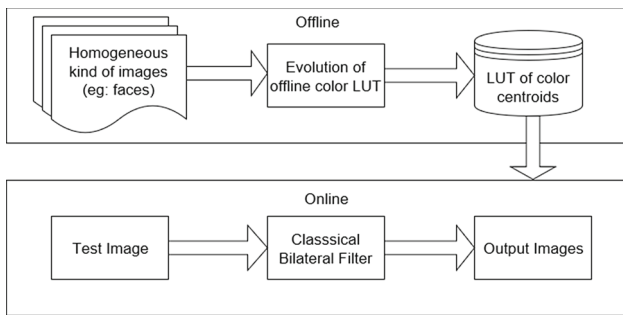


Fig. 6 Scheme of hybridization between offline and online processing in proposed fast bilateral filtering: the LUT has been derived from DCM through decoding the same in the reverse path of the autoencoder

and

$$\phi(z) = \exp\left(-\frac{\|z\|^2}{2\sigma_r^2}\right) \tag{7}$$

where $x \in \mathbb{R}^2$ and $z \in \mathbb{R}^\rho$. The window size is set as $S = 3\sigma_s$. For color images, $\rho = 3$. The approximation error has been qualified through root mean square error (RMSE) as follows:

$$RMSE^2 = \frac{1}{|\Omega|} \sum_{i \in \Omega} \|\hat{g}(i) - g(i)\|^2 \tag{8}$$

where g is the exact bilateral filter in Eq. 1 and \hat{g} is the approximation from the respective algorithms. It is evident that that the RMSE for our proposed algorithm consistently decreases with increase in K (number of clusters), whereas that of the adaptive manifolds oscillates with increase in K (number of manifolds) [2].

The experiments have been performed in an Intel Core i7-7500U CPU @ 2.70 GHz PC having NVIDIA 940 Mx GPU with 8GB RAM. Table 1 depicts the comparison between fast high-dimensional filter [5] and our proposed filtering approach based on CDDA in terms of RMSE and performance (ms). The DCM has been derived over a large number of homogeneous kind of images through CDDA. We have used *CelebA* dataset for extracting dominant colors for face images [15]. Next, noise of 12 dB has been introduced to the image (Fig. 7) to be treated as unknown input for filtering. The RMSE has been calculated based on the clean version of the aforementioned image and is shown in Fig. 8 .

The relationship depicted in Table 1 is illustrated in Fig. 8 where the following observations are made:

1. With increasing cluster number, RMSE has been reduced and performance has been degraded for both the methods as expected.
2. Impact of increasing cluster number in LUT is higher.
3. With higher number of clustering, both the RMSEs are converging.

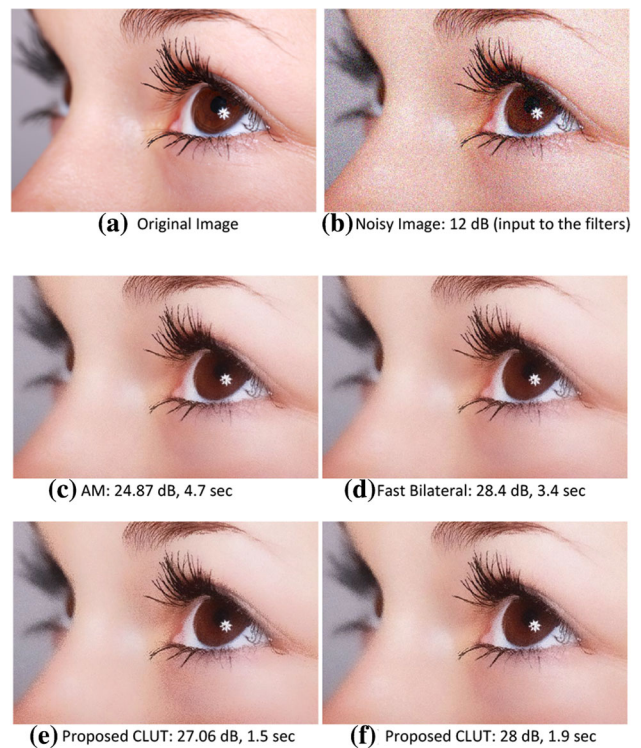


Fig. 7 Edge preserved denoising for natural RGB image (Size: 640 × 427): comparisons of CDDA with adaptive manifold [4] and fast high-dimensional filtering [5]

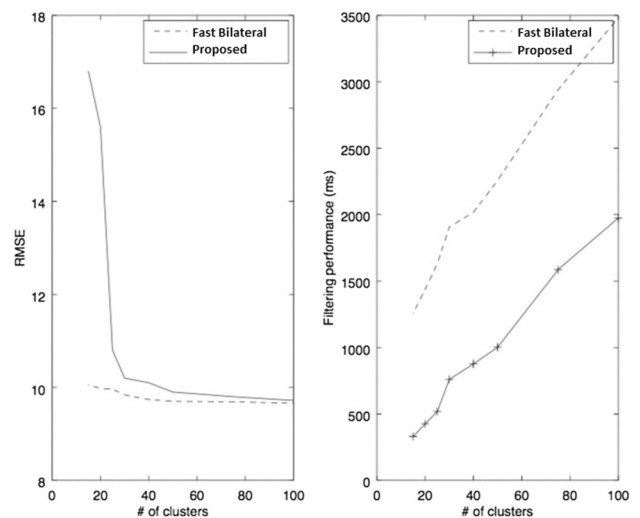


Fig. 8 High-performance CDDA: performance and RMSE

4. As expected, performance of our approach is much higher than its equivalent algorithm [5].
5. If we observe both RMSE and performance graph, it is clear that fast high-dimensional [5] has taken equal time for 25 clusters of that of 75 clusters of our proposed approach. This signifies that our approach can achieve same psycho-visual quality of reconstruction much faster than fast high-dimensional [5].

Table 1 Comparison between fast range-domain filtering [5] and proposed CDDA in terms of performance, RMSE and PSNR

No. of cluster, K	15	20	25	30	40	50	75	100
Performance of [5] (ms)	1257	1440	1628	1906	2017	2257	2939	3482
Performance of CDDA (ms)	331	426	518	761	877	1003	1585	1974
RMSE of [5]	10.05	9.9	9.96	9.84	9.74	9.7	9.69	9.66
RMSE of CDDA	16.8	15.6	10.8	10.2	10.1	9.9	9.8	9.72
PSNR of [5] (dB)	28.07	28.14	28.16	28.26	28.35	28.38	28.4	28.42
PSNR of CDDA (dB)	22.6	23.21	25.91	26.38	26.71	26.06	27.06	27.952

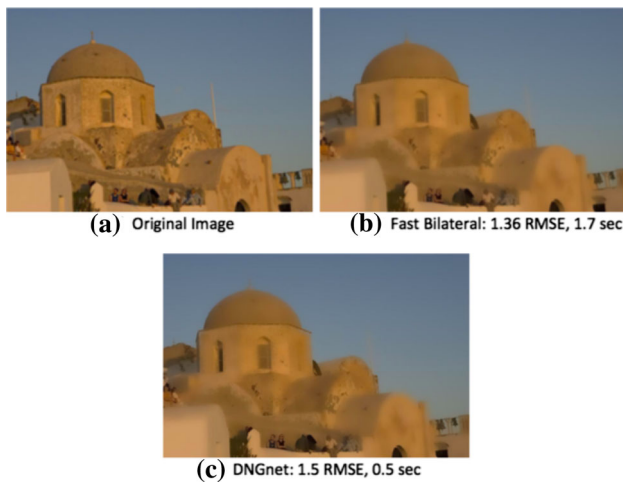
**Fig. 9** Bilateral filtering of natural heterogeneous RGB image (Size: 876×584): comparisons of RMSE and time between CDDA and fast high-dimensional filtering [5]

Figure 7 is evident that the color LUT outperformed all state-of-art fast filters of color sparseness category [2]. Next, just to test heterogeneity in kind of images to be filtered, we have tested completely different genres of image as depicted in Fig. 9. Here, the problem statement is to restore image of a historical monument. The bilateral filtering is again the answer to the problem as it enhances images by ensuring edge preservation. As the said figure depicts, the CDDA-based bilateral filtering has shown similarly promising result even in image restoration. The proposed bilateral filtering based on CDDA has shown improved time performance (more than one thrice) keeping similar RMSE with respect to fast bilateral filter [5].

5 Conclusion

This current work has targeted to make bilateral filtering significantly efficient in terms of performance, not compromising the psycho-visual quality. The work has utilized color sparseness in images and proposed a dominant color map (DCM)-based approach for accelerating filtering opera-

tion. The proposal ensured high accuracy and performance of reconstruction as well. The principal idea behind the proposal is twofold. The DCM construction algorithm determines the salient colors in homogeneous image kind exploiting the color sparseness property of images. The experimental results have shown promise to the new areas of research of hybrid filtering approach combining classical low-level image processing. The outcome of our research has not only shown the effectiveness of the proposed solution, but also has shown another direction of applying unchanged DCM to heterogeneous image restoration. The possible future work definitely would open a new door of generic image bilateral filtering in real-time through the CDDA (Color-Dominant Deep Autoencoder) framework proposed in this work [16–18].

Acknowledgements The authors are thankful to Pallavi Saha for helping us in data collection, annotation, and simulation.

References

- Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: International Conference on Computer Vision, pp. 839–846 (1998)
- Das, A., Nair, P., Shylaja, S.S., Chaudhury, K.N.: A concise review of fast bilateral filtering. In: 2017 Fourth International Conference on Image Information Processing (ICIIP), 1–6 (2017)
- Das, A.: Guide to Signals and Patterns in Image Processing. Springer, Berlin (2015)
- Gastal, E.S.L., Oliveira, M.M.: Adaptive manifolds for real-time high-dimensional filtering. SIGGRAPH (ACM TOG) **31**(4), 1–13 (2012)
- Nair, P., Chaudhury, K.N.: Fast high-dimensional filtering using clustering. In: International Conference of Image Processing (2017)
- Durand, F., Dorsey, J.: Fast bilateral filtering for the display of high-dynamic-range images. ACM Trans. Graph. **21**, 257–266 (2002)
- Yang, Q., Tan, K. H., Ahuja, N.: Real-time $O(1)$ bilateral filtering. In: CVPR 2009. IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 557–564. IEEE (2009)
- Yang, Q., Tan, K.H., Ahuja, N.: Constant time median and bilateral filtering. Int. J. Comput. Vision **112**, 307–318 (2015)
- Mozerov, M.G., van de Weijer, J.: Global color sparseness and a local statistics prior for fast bilateral filtering. IEEE Trans. Image Process. **24**, 5842–5853 (2015)

10. Shen, X., Chen, Y.C., Tao, X., Jia, J.: Convolutional neural pyramid for image processing. *ArXiv e-prints* (2017)
11. Gharbi, M., Chen, J., Barron, J.T., Hasinoff, S.W., Durand, F.: Deep bilateral learning for real-time image enhancement. *ACM Trans. Gr (TOG)* **36**(4), 118 (2017)
12. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The feret evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(10), 1090–1104 (2000)
13. Guo, X., Liu, X., Zhu, E., Yin, J.: Deep clustering with convolutional autoencoders. In: *International Conference on Neural Information Processing* (2017)
14. Ye, M., Giannarou, S., Meining, A., Yang, G.Z.: Online tracking and retargeting with applications to optical biopsy in gastrointestinal endoscopic examinations. *Med. Image Anal.* (2015)
15. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. Presented at the (2015)
16. Das, A., Ajithkumar, N.: Engineering the perception of recognition through interactive raw primal sketch by HNFGS and CNN-MRF. In: *Computer Vision and Image Processing (CVIP)*. Springer, Berlin (2017)
17. Konar, A.: *Computational Intelligence: Principles, Techniques and Applications*. Springer, Berlin (2005)
18. Das, A.: *Bacterial foraging optimization for digital filter synthesis: a computational intelligence approach to dsp and image processing*. Lambert Academic Publishing, Cambridge (2013)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.