Fast Non-Local Means (NLM) Computation with Probabilistic Early Termination

Ramanathan Vignesh†, Byung Tae Oh* and C.-C.Jay Kuo‡
†Dept. of Electrical Engineering, Indian Institute of Technology, Kharagpur - 721302, India
*‡Signal and Image Processing Institute, Ming Hsieh Dept. of Electrical Engineering
University of Southern California, Los Angeles, CA 90089-2564, USA
E-mails: *vigneshram.iitkgp@gmail.com, btoh77@gmail.com and cckuo@sipi.usc.edu*

***Abstract*—A speed up technique for the non-local means (NLM) image denoising algorithm based on probabilistic early termination (PET) is proposed. A significant amount of computation in the NLM scheme is dedicated to the distortion calculation between pixel neighborhoods. The proposed PET scheme adopts a probability model to achieve early termination. Specifically, the distortion computation can be terminated and the corresponding contributing pixel can be rejected earlier, if the expected distortion value is too high to be of significance in weighted averaging. Performance comparative with several fast NLM schemes is provided to demonstrate the effectiveness of the proposed algorithm.**

***Index Terms*—Non-local means (NLM) algorithm, image denoising, probabilistic algorithm, early termination, fast algorithm.**

I. Introduction

Image denoising is a fundamental yet challenging problem that has been studied for decades. A thorough review of state-of-the-art denoising algorithms is given in [1]. One emerging image denoising technique developed within the last five years is the non-local means (NLM) algorithm [1]. Unlike most denoising algorithms that rely on the local regularity assumption, the NLM algorithm estimates an unknown pixel by a weighted average of local and non-local pixels throughout the entire image. The weight decreases exponentially with an increased distortion between the neighborhoods of pixels under consideration. The NLM algorithm has been shown to outperform many contemporary denoising techniques in both PSNR and visual quality improvement.

However, the superior performance of the NLM algorithm is achieved at the cost of higher computational complexity. That is, for a given pixel of interest (POI), the standard NLM algorithm computes the distortion between the block around the pixel and every other block in the image. Consequently, a lot of computation is needed to calculate the distortion between pixel neighborhoods. Various NLM speed-up schemes with limited performance degradation were studied in [2]–[6].

Since the number of pixels with a large weight in computing the restored POI value is limited, one may speed up the NLM algorithm by eliminating dissimilar pixels in the distortion computation. This idea has been adopted in [2]–[5]. The elimination in most previous work was based on a hard decision. For instance, only pixels with their patch means close to that of the POI were considered in [2]. However, it cannot eliminate all insignificant patches thoroughly since patches with the same mean can be quite different in their neighborhoods. In [3], similar patches are grouped into a cluster, and only pixels in the same cluster are considered for averaging. Although the clustering scheme works well for textured images with a strong repeated pattern, its performance is not efficient for real-world images. In this letter, we propose a fast NLM scheme that selects patches using a soft decision, where the elimination criterion varies with the structural difference between patches under consideration. The distortion computation can be terminated and the corresponding block can be rejected earlier, if the expected distortion value is too high to be of significance in weighted averaging. The expected distortion value is estimated at each stage of distortion computation by a probability model based on patch features. Thus, it is called the probabilistic early termination (PET) scheme.

The use of early termination to increase the speed of block matching has been considered in the context of vector quantization [7] and motion estimation [8]. In particular, a probabilistic model for early termination was proposed in [8], which searches the best matching block in terms of the smallest sum of absolute differences (SAD). However, the fast NLM computation problem is different from that in [7], [8]. Here, instead of finding the best match, we have to choose multiple similar blocks and determine their similarity degree that determines the weights in the averaging process to denoise the POI. For this reason, we need a more accurate mechanism to determine which blocks to discard so that the ultimate denoising performance is not severely sacrificed. Besides, our selection criterion is not based on the SAD but the Gaussian weighted sum of the squared distance between the patch around the POI and a non-local block.

The rest of this letter is organized as follows. The basic NLM algorithm for image denoising is briefly reviewed in Sec. II. The proposed fast NLM scheme is introduced in Sec. III. Experimental results are provided to demonstrate the performance of the proposed algorithm in Sec. IV. Finally, concluding remarks are given in Sec. V.

II. Image Denoising with Non Local Means

The objective of image denoising is to restore the unknown original image, X from a noisy image, Y given by

$$Y = X + N\,. \tag{1}$$

The NLM algorithm restores the pixel of interest (POI) with a weighted average of non-local pixels in the image which are in a similar environment. The weight is computed based on the similarity between the neighborhood patches of the POI and contributing pixels.

Mathematically, the solution to Eq. (1) obtained by the standard NLM scheme can be written as

$$\begin{aligned}\hat{X}(i) &= \sum_{j\in I} w_{ij} Y(j) \\ &= \frac{1}{Z_i}\sum_{j\in I}\exp\left(-\frac{d(Y(\mathcal{N}_i),Y(\mathcal{N}_j))}{h^2}\right)Y(j)\,, \quad (2)\end{aligned}$$

where w_{ij} is a weight denoting the contribution from $Y(j)$ to $\hat{X}(i)$, d is a distortion measure, $\mathcal{N}_i$ is a squared neighborhood patch of pixel $i = (i_1, i_2)$, h is the parameter of distortion measure to adjust the decay of the weight, I is image space, and Z_i is the normalization factor such that $\sum_j w_{ij} = 1$. One common distortion measure in Eq. (2) is chosen as

$$d(Y(\mathcal{N}_i),Y(\mathcal{N}_j)) = \sum_{k\in K} G_a(k)(Y(i-k)-Y(j-k))^2\,, \quad (3)$$

where K is a local neighborhood centered at the origin and

$$G_a(k) = \frac{1}{\sqrt{2\pi}a}\exp\left(-\frac{k_1^2+k_2^2}{2a^2}\right), k = (k_1, k_2)$$

is the Gaussian kernel with standard deviation $a > 0$.

By including all pixels in the image for the weighted average computation in Eq. (3), which means that K is large enough to cover all pixels in the image of interest, the complexity of the NLM algorithm can go as high as $O(M^4)$ for an image of size $M \times M$. To lower the complexity, only a subset of pixels within a local search window of size $s \times s$ ($s \ll M$) around a POI is used in the computation [1] and, consequently, the complexity is reduced to $O(M^2 s^2)$. We call the latter the *standard* NLM scheme in the experimental section.

III. Probabilistic Early Termination (PET) Scheme

In this section, we propose a fast NLM computation scheme by eliminating dissimilar blocks without full computation of distortion. This is implemented with the help of probabilistic early termination (PET). That is, based on the partial sum of the distortion in Eq. (3), we determine the probability of the distortion value exceeding a pre-determined threshold and terminate the distortion computation if the probability is greater than a fixed value P_τ.

A. Determination of Summation Order

We need to determine a summation order for points in K so that the partial sum may reach the threshold as fast as possible. It is natural to use the radial scanning order, where we start from the inner-most regions and gradually move towards outer regions. This is desirable since the inner position has a larger weight due to the effect of the Gaussian kernel. Mathematically, we decompose set K via

$$K = K_1 \cup \cdots \cup K_n \cup \cdots \cup K_g, \quad (4)$$

where

$$K_n = \{k \in K \mid (n-1)^2 \le \|k\|^2 < n^2\} \quad 1 \le n \le g,$$

corresponds to the ring-shaped region as illustrated in Fig. 1. Then, the distortion sum for each K_n can be written as

$$d_n(i,j) = \sum_{k\in K_n} G_a(k)(Y(i-k)-Y(j-k))^2\,, \quad (5)$$

and the partial sum of the distortion from K_1 to K_n is

$$D_n(i,j) = d_1(i,j) + \cdots + d_n(i,j). \quad (6)$$

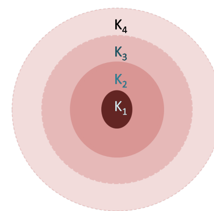


Fig. 1. Neighborhood decomposition into four ring-shaped regions

B. Statistical Characterization of Patch Differences

The distortion in Eq. (3) can be interpreted as the energy of the following weighted patch-difference array:

$$Y_{ij}(k) = \sqrt{G_a(k)}(Y(i-k)-Y(j-k)), \quad k \in K\,. \quad (7)$$

A probability model was used in [8] to describe the SAD distribution over different block. Here, since the value of $D_g(i,j)$ is used to select suitable blocks in weighted averaging, we attempt to estimate the total sum, D_g, from the partial sum, D_n, by modeling Y_{ij} with suitable parameters. Specifically, the mean and variance of Y_{ij} are chosen as the key parameters to model Y_{ij}, since they are the most prominent features in a probabilistic model. The distance between two image patches is modeled as a χ^2 distribution to calculate adaptive weights in [9]. The χ^2 distribution is a direct consequence of the assumption that the terms contributing to the weighted difference between two patches are i.i.d. Gaussian random variables. Following [9], we assume

$$Y_{ij} \sim G(m_{ij}, \sigma_{ij}^2). \quad (8)$$

Then, the sum of Y_{ij}^2 follows the non-central χ^2 distribution.

The above idea can be used to reject dissimilar patches at an earlier stage than a conventional early termination scheme, which rejects patches when the n^{th} partial sum of distortion exceeds a pre-determined threshold, δ, as

$$D_n(i,j) > \delta. \quad (9)$$

In contrast, the proposed PET scheme estimates the probability distribution of the distortion in the remaining regions given by

$$R_n(i,j) = d_{n+1}(i,j) + \cdots + d_g(i,j)\,,$$

and rejects the patches when

$$P(R_n(i,j) + D_n(i,j) > \tau \big| D_n(i,j)) > P_\tau, \quad (10)$$

where the evaluation of the probability in the left-hand-side will be discussed in Sec. III-C. The criterion in (10) acts as adaptive thresholding, which varies at each ring-shaped region to decide which patch to reject in the NL computation. In the experiments reported in Sec. IV, we choose $\tau = 6h^2$ and $P_\tau = 90\%$.

C. Determination of Mean and Variance

In this subsection, we find ways to determine the mean (m_{ij}) and variance (σ_{ij}^2) for Gaussian random variable Y_{ij} with minimum computation. The mean and variance are computed for each individual block in the image. One simple solution is the use of the sample mean and variance. However, this computation would be high.

The mean of the patch-difference is equal to

$$m_{ij} = m_i - m_j, \tag{11}$$

where m_i and m_j are pre-computed means of the weighted patches centered at pixels i and j given as

$$m_i = \frac{\sum_{k \in K} \sqrt{G_a(k)} Y(i-k)}{|K|}, \tag{12}$$

where $|\cdot|$ denotes the cardinality of a set.

The variance σ_{ij}^2 can be written as

$$\sigma_{ij}^2 = E(Y_{ij}^2) - (E(Y_{ij}))^2$$

where, $E(.)$ represents the expectation of a variable. Due to the i.i.d assumption, $E(Y_{ij}^2)$ can be approximated as the average of $Y_{ij}^2(k)$, where $k \in K_1 \cup \cdots \cup K_n$. Hence, we can approximate σ_{ij}^2 using the following relationship:

$$\sigma_{ij}^2 \approx \frac{D_n(i,j)}{|K_1 \cup \cdots \cup K_n|} - m_{ij}^2. \tag{13}$$

Since $D_n(i,j)$ is already computed, it demands a small increase in complexity to determine σ_{ij}^2. Note that the approximated σ_{ij}^2 varies for each ring-shaped region due to D_n.

The mean and variance for χ^2-distributed R_n is given by

$$E[R_n] = |K_{n+1} \cup \cdots \cup K_g|(m_{ij}^2 + \sigma_{ij}^2), \tag{14}$$

$$VAR[R_n] = 2|K_{n+1} \cup \cdots \cup K_g|(2m_{ij}^2 + \sigma_{ij}^2). \tag{15}$$

Finally, the conditional probability in (10) can be determined accordingly.

IV. Experimental Results

For performance evaluation, we consider several test images corrupted by additive white Gaussian noise with standard deviation $\sigma = 20$. The value of h^2 in Eq. (2) is chosen to be $h^2 = 10\sigma = 200$ as suggested in [1]. We compare the performance of the proposed PET scheme with the standard NLM computation and two fast NLM computational schemes. The two fast schemes are: 1) the pre-selection-by-mean scheme in [2] and 2) the clustering scheme in [3].

For the pre-selection-by-mean scheme, we used the threshold values given in [2]. For the clustering scheme, we adopted clusters without overlapping since overlapping demands higher complexity yet with little PSNR improvement for natural images. For the proposed PET scheme, the threshold on the distortion is set to $\tau = 6h^2$ while the value of P_τ in Eq. (10) is set to 90%. A higher value of P_τ leads to rejection of fewer blocks and the process becomes closer to a deterministic approach as P_τ approaches 100%.

Tables I and II compare the computational complexity, PSNR and the mean SSIM index [10] results of four schemes with three search window sizes (23×23, 43×43 or 63×63) and two neighborhood patch sizes (7×7 or 11×11) for Pepper and Lena images of size 256×256. The computational complexity is represented as the percentage of the total amount of computations required by the standard NLM scheme. We see that the proposed PET scheme achieves the lowest complexity and the highest PSNR in most cases. Denoised Lena images of the four schemes are shown in Fig. 2 for visual comparison. The denoised image obtained by the proposed PET scheme as shown in Fig. 2-(d) is perceptually similar to that obtained by the standard NLM computation as shown in Fig. 2-(a). In contrast, the clustering scheme has unpleasant visual artifacts.

V. Conclusion

A probabilistic early termination (PET) scheme for fast NLM computation was presented in this work. The PET scheme terminates the computation of the distortion term, when the partial sum of the distortion term exceeds a predefined threshold. The threshold value is decided by a probability model based on patch differences of the POI and the pixel under consideration. It was shown experimentally that the proposed PET scheme outperforms two benchmark schemes by considering several performance metrics together, including complexity reduction, PSNR quality improvement and visual perception enhancement.

Acknowledgements

This research is supported in part by a gift grant from Xerox.

References

[1] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *SIAM Interdisciplanary J.*, vol. 4, no. 2, pp. 490–530, 2005.

[2] M. Mahmoudi and G. Sapiro, "Fast image and video denoising via nonlocal means of similar neighborhoods," *IEEE Signal Process. Lett.*, vol. 12, no. 12, pp. 839–842, 2008.

[3] T. Brox, O. Kleinschmidt, and D. Cremers, "Efficient nonlocal means for denoising of textural patterns," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1083–1092, Jul. 2008.

[4] G. Gilboa and S. Osher, "Nonlocal linear image regularization and supervised segmentation," CAM-06-47 Dept. Math., Univ. California, Los Angeles, Tech. Rep., 2006.

[5] P. Coupe, P. Yger, C. Barillot, R. Larsen, M. Nielsen, and J. Sporring, "Fast nonlocal means denoising for 3D MR images," in *Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention*, vol. 4191, 2006, pp. 33–40.

[6] J. Wang, Y. Guo, Y. Ying, Y. Liu, and Q. Peng, "Fast non-local algorithm for image denoising," in *Proc. of IEEE International Conference on Image Processing*, 2006.

[7] C.-D. Bei and R. M. Gray, "An improvement of the minimum distortion encoding algorithm for vector quantization," *IEEE Trans. Commun.*, vol. 33, no. 10, pp. 1121–1133, Oct. 1985.

[8] K. Lengwehasatit and A. Ortega, "Probabilistic partial distance fast matching algorithms for motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 2, pp. 139–152, Feb. 2001.

TABLE I

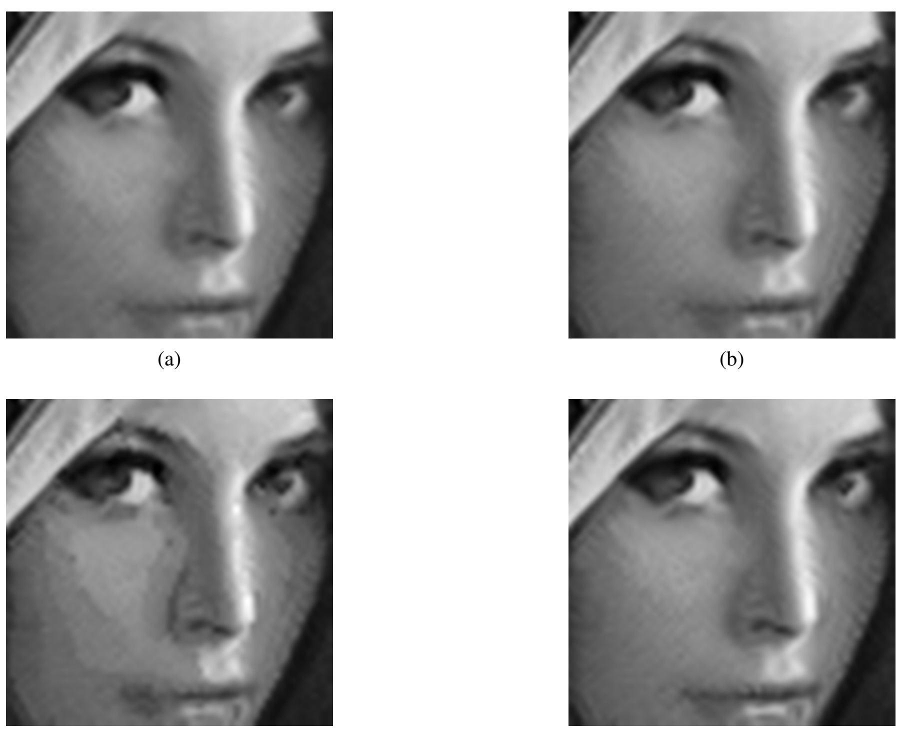
COMPARISON OF COMPUTATIONAL SPEED AND DENOISED IMAGE QUALITY OF SEVERAL NLM SCHEMES FOR THE PEPPERS TEST IMAGE. THE COMPUTATION OF THE FASTEST SCHEME, THE BEST PSNR AND MEAN SSIM INDEX AMONG THE FAST NLM SCHEMES ARE HIGHLIGHTED.

Schemes	Search Window = 23×23						Search Window = 43×43						Search Window = 63×63					
	NPS = 7×7			NPS = 11×11			NPS = 7×7			NPS = 11×11			NPS = 7×7			NPS = 11×11		
	Comp. (%)	PSNR	SSIM	Comp. (%)	PSNR	SSIM	Comp. (%)	PSNR	SSIM	Comp. (%)	PSNR	SSIM	Comp. (%)	PSNR	SSIM	Comp. (%)	PSNR	SSIM
Standard NLM	100	27.83	0.829	100	28.25	0.856	100	27.64	0.828	100	28.15	0.844	100	28.11	0.848	100	27.91	0.864
Fast NLM [2]	46.89	**27.78**	**0.819**	49.31	28.11	0.848	37.55	27.37	**0.828**	38.68	27.83	**0.845**	35.08	**28.25**	**0.839**	35.80	27.84	0.852
Fast NLM [3]	37.32	27.51	0.748	38.13	**28.25**	0.785	19.14	27.21	0.761	20.22	28.01	0.773	12.84	28.04	0.755	14.74	27.78	0.781
Proposed PET	**22.97**	27.60	0.816	**19.05**	28.22	**0.853**	**10.13**	**27.47**	0.822	**6.87**	**28.13**	**0.845**	**6.63**	28.15	0.836	**3.16**	**28.05**	**0.865**

TABLE II

COMPARISON OF COMPUTATIONAL SPEED AND DENOISED IMAGE QUALITY OF SEVERAL NLM SCHEMES FOR THE LENA TEST IMAGE. THE COMPUTATION OF THE FASTEST SCHEME, THE BEST PSNR AND MEAN SSIM INDEX AMONG THE FAST NLM SCHEMES ARE HIGHLIGHTED .

Schemes	Search Window = 23×23						Search Window = 43×43						Search Window = 63×63					
	NPS = 7×7			NPS = 11×11			NPS = 7×7			NPS = 11×11			NPS = 7×7			NPS = 11×11		
	Comp. (%)	PSNR	SSIM	Comp. (%)	PSNR	SSIM	Comp. (%)	PSNR	SSIM	Comp. (%)	PSNR	SSIM	Comp. (%)	PSNR	SSIM	Comp. (%)	PSNR	SSIM
Standard NLM	100	28.68	0.821	100	29.06	0.838	100	28.79	0.812	100	28.83	0.831	100	28.61	0.810	100	28.72	0.827
Fast NLM [2]	40.71	28.44	**0.816**	41.32	28.93	**0.834**	30.32	28.62	**0.815**	31.98	28.70	**0.824**	30.11	28.42	**0.802**	29.62	28.36	0.817
Fast NLM [3]	37.46	28.32	0.757	38.51	**29.05**	0.788	20.14	28.58	0.735	22.23	28.66	0.760	16.84	28.53	0.732	15.19	**28.82**	0.749
Proposed PET	**20.12**	**28.54**	0.814	**17.36**	29.03	0.831	**12.79**	**28.74**	0.804	**8.40**	**28.71**	0.822	**5.16**	**28.63**	0.789	**2.80**	28.68	**0.824**



(a) (b) (c) (d)

Fig. 2. Denoised Lena Image results: (a) standard NLM (PSNR=28.68dB, SSIM=0.821), (b) the pre-selection-by-mean scheme in [3] (PSNR=28.44dB, SSIM=0.816), (c) the clustering scheme in [4] (PSNR=28.32dB, SSIM=0.757) and (d) the proposed PET scheme (PSNR=28.54dB, SSIM=0.814).

[9] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2866–2878, Oct. 2006.

[10] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.