A LARGEST MATCHING AREA APPROACH TO IMAGE DENOISING

Jack Gaston, Ji Ming, Danny Crookes

Institute of Electronics, Communications and Information Technology
Queen's University Belfast, Belfast BT3 9DT, UK

ABSTRACT

Given the success of patch-based approaches to image denoising, this paper addresses the ill-posed problem of patch size selection. Large patch sizes improve noise robustness in the presence of good matches, but can also lead to artefacts in textured regions due to the *rare patch* effect; smaller patch sizes reconstruct details more accurately but risk over-fitting to the noise in uniform regions. We propose to jointly optimize each matching patch's identity and size for grayscale image denoising, and present several implementations. The new approach effectively selects the *largest matching* areas, subject to the constraints of the available data and noise level, to improve noise robustness. Experiments on standard test images demonstrate our approach's ability to improve on fixed-size reconstruction, particularly at high noise levels, on smoother image regions.

***Index Terms*—** Image denoising, variable patch size, largest matching area, example-based denoising, internal denoising

1. INTRODUCTION

Approaches to image restoration fall into two main categories: those that construct a model (e.g., a Sparse Representation (SR) dictionary or a deep neural network (DNN)) from a priori knowledge such as training data (*external denoising*), and those that learn the characteristics of the underlying signal from the given data itself (e.g., non-local means (NLM) or BM3D) (*internal denoising*). In [1] it is shown that small image patches where signal dominates (eg. strong edges) are more accurately denoised by exploiting clean *external* patches. However, uniform regions which are dominated by noise have a preference for *internal* patches.

SR assumes that an image can be sparsely represented by a linear combination of atoms from a dictionary. The problems are then to learn appropriate dictionaries from training data [2] and to sparsely reconstruct with respect to a fidelity constraint based on the known noise level [3–5]. DNNs learn a model of the relation between a noisy patch and its clean counterpart [6, 7]. Internal denoising methods exploit natural redundancy in images to maintain the clean signal and attenuate noise. NLM [8–11] reconstructs by a weighted average of similar patches in a search window. This was further extended in BM3D [12], which identifies and groups similar patches in 3D arrays to reduce noise in a collaborative filtering approach.

The choice of patch-size in both categories is ill-posed, as the diversity of image content may not be represented well at large sizes, while small patches risk reconstructing the noise [1]. Therefore, we propose optimizing the matching patches' identity and size jointly. We hope to increase the affect of the clean signal on patch-matching in uniform regions by including more discriminative data, and improve the reconstructed accuracy of textured regions by finding their largest similar patches, subject to the constraints of the training data.

To confirm our intuition, we conducted fixed-size example-based denoising experiments at four noise levels (algorithm and data will be detailed later). Fig. 1 shows the average PSNR and SSIM of the reconstructed images as a function of the fixed patch size k. For each noise level, as k increases, the PSNR and SSIM also increase to a peak size. Higher noise levels introduce greater ambiguity, rapidly degrading performance with small patches. Larger patches are more invariant to noise, but close matches may not exist for very large patches, decreasing the reconstructed accuracy. Ideally, we would like to use the *largest* matching patches at each location, subject to the constraints of the training dataset and the noise level. We call the new approach largest matching area (LMA) based denoising.

1.1. Related Work

Internal denoising algorithms use larger patches and search-windows at higher noise levels [13–16], improving reconstructed quality. Extensions to the NLM algorithm proposed adaptive spatial support for superior results, by classifying patches as textured or smooth by edge-detection and morphological operations [17], or by clustering the SVD (singular-value decomposition) of the blocks' gradient fields [18], allowing spatial adaptation for each block. An adaptive neighbourhood is selected for each pixel in [19, 20], by iteratively increasing the search-window and carrying out NLM denoising, until the L2-risk with respect to the search-window size is minimized - minimizing the variance subject to a fidelity constraint.

In [21] locally varying the NLM filter decay rate to make the choice of patch-size less critical is suggested, to alleviate the *rare patch* effect. The later work in [17] handles these artefacts with locally adapted patch and search-window sizes. Dabov *et al.* proposed shape adaptive patches for use in the BM3D algorithm [22], where a locally adaptive region of the matching patches is used for filtering, to maintain structures (eg. edges) and reduce distortion (eg. halo artefacts). A shape adaptive approach to the NLM algorithm [23] also produced improved results with reduced artefacts.

In external denoising patch-size optimization remains largely an open problem. Multi-scale SR [24] used dictionaries at two to three sizes and showed that large structures can be more accurately coded by larger atoms, obtaining state-of-the-art results in denoising and inpainting, especially at high noise levels, learning from large *global* databases. DNN based approaches perform well if the sample patch-size is large enough to contain sufficient information to recover a noise-free version, requiring a large quantity of training data.

1.2. Contributions

In this paper, we investigate image denoising by optimizing both the identity and size of matching patches. We introduce a simple, novel framework for fixed-size example-based denoising, where a noisy image is reconstructed by selecting the best fixed-size match for each pixel, in an *external* gallery of clean images, and averaging these over-lapping patches. Then, we extend this approach to example-based LMA denoising, to select the best variable-sized gallery patch.

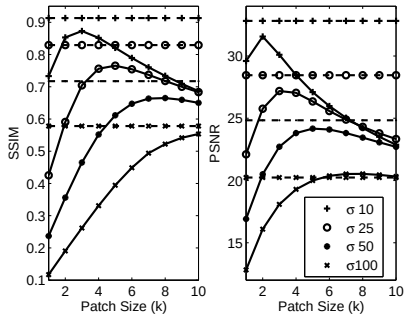


Fig. 1: Average reconstructed PSNR and SSIM for 4 test images over 2 datasets at 4 noise levels (solid curves), across fixed-size patches of $(2k+1)\times(2k+1)$, and the LMA based results (dotted lines)

In the following, we introduce the LMA algorithm for an *example-based* system, giving the detailed maximum *a posteriori* (MAP) implementation algorithm. Then, we propose SR and BM3D extensions incorporating the LMA approach. Experimentally, we show that the new LMA approach outperforms the fixed patch-size based approach, and our SR and BM3D extensions produce superior or competitive results across a range of noise levels.

2. LARGEST MATCHING AREA (LMA) BASED DENOISING

2.1. Example-based denoising

Before discussing the LMA algorithm for estimating the optimal matching patches, we briefly describe our approach to example-based denoising. Assume we have a set of clean training images (i.e. examples). Given a noisy image, we search for the most likely fixed-size patches in training (by (1)) and use them for reconstruction. This is easily extended to the search for variable-sized matches. We describe the LMA algorithm for this example-based denoising system before extending the algorithm to other denoising systems, including SR, which requires dictionary learning for each patch size.

For simplicity, we consider square matching patches. Let $\mathbf{y}$ represent a noisy image with $\mathbf{y}_{k,i,j}$ being a square patch centered at (i,j) with size $(2k+1)\times(2k+1)$, where $k\geq 1$. Similarly, let $\mathbf{x}^m$ represent the mth clean training image and $\mathbf{x}^m_{k,u,v}$ denote a square patch in $\mathbf{x}^m$ centered at (u,v) of the same size. In our example-based system, the problem of image denoising can be stated as identifying for each noisy image patch $\mathbf{y}_{k,i,j}$ a matching training image patch $\mathbf{x}^m_{k,u,v}$, such that their combination can reconstruct the underlying target image. Next, we identify the matching patches $\mathbf{y}_{k,i,j}$ and $\mathbf{x}^m_{k,u,v}$ by optimizing both the size k and the location of $\mathbf{x}^m_{k,u,v}$.

2.2. Example-based LMA - a MAP algorithm

Consider a statistical approach to measure how well a noisy image patch $\mathbf{y}_{k,i,j}$ is matched by a clean training image patch $\mathbf{x}^m_{k,u,v}$. Assume stationary noise in $\mathbf{y}_{k,i,j}$ and assume that the likelihood function of $\mathbf{y}_{k,i,j}$ associated with a matching $\mathbf{x}^m_{k,u,v}$ can be written as

$$p(\mathbf{y}_{k,i,j}|\mathbf{x}^m_{k,u,v}) = a\exp(-\frac{\|\mathbf{y}_{k,i,j}-\mathbf{x}^m_{k,u,v}\|^2}{h^2}) = a\prod_n \exp\big(-\frac{(y_{k,i,j}(n)-x^m_{k,u,v}(n))^2}{h^2}\big) \quad (1)$$

where $y_{k,i,j}(n)$ denotes the nth pixel in $\mathbf{y}_{k,i,j}$ (likewise for $x^m_{k,u,v}(n)$), and the product is taken over all pixels within the patches. Like the vectorial NLM likelihood [25], this likelihood adapts euclidean distance using the parameter h to characterise the uncertainty caused by noise; a is a normalization constant. To reduce the uncertainty, at each location (i,j), we aim to find, by extending k, the largest (and hence most recognizable) noisy patch $\mathbf{y}_{k,i,j}$ with a matching training patch. This can be formulated as a MAP problem by maximizing the likelihood (1).

Given a noisy image patch $\mathbf{y}_{k,i,j}$, assume an equal prior probability P for all possible clean image patches $\mathbf{x}_k$. We define the posterior probability of the match of a clean training patch $\mathbf{x}^m_{k,u,v}$ as

$$P(\mathbf{x}^m_{k,u,v}|\mathbf{y}_{k,i,j}) = \frac{p(\mathbf{y}_{k,i,j}|\mathbf{x}^m_{k,u,v})P}{\sum_{\text{All }\mathbf{x}_k} p(\mathbf{y}_{k,i,j}|\mathbf{x}_k)P} \simeq \frac{p(\mathbf{y}_{k,i,j}|\mathbf{x}^m_{k,u,v})}{\sum_{m'}\sum_{u',v'} p(\mathbf{y}_{k,i,j}|\mathbf{x}^{m'}_{k,u',v'}) + p(\mathbf{y}_{k,i,j}|\phi_k)} \quad (2)$$

where $p(\mathbf{y}_{k,i,j}|\mathbf{x}^m_{k,u,v})$ is the likelihood defined in (1). The denominator, the average likelihood of $\mathbf{y}_{k,i,j}$, is approximated by a sum of two terms. The first is the average likelihood of $\mathbf{y}_{k,i,j}$ over all training patches, assuming that $\mathbf{y}_{k,i,j}$ will be matched by at least one of these training patches. The second term, $p(\mathbf{y}_{k,i,j}|\phi_k)$, represents the average likelihood of $\mathbf{y}_{k,i,j}$ when a close match cannot be found in the training dataset. This can happen, for example, when $\mathbf{y}_{k,i,j}$ is too large. This unseen data likelihood can be suitably represented using a mixture model. We use the expression

$$p(\mathbf{y}_{k,i,j}|\phi_k) = a\prod_n\Big[\sum_{l=0}^{L-1} P(l)\exp\big(-\frac{(y_{k,i,j}(n)-l)^2}{h^2}\big)\Big] \quad (3)$$

where L possible values are assumed for each image pixel ($L=256$ for grayscale images) and $P(l)$ is a prior probability for the underlying pixel in $y_{k,i,j}(n)$ taking value l. Assuming a uniform $P(l)$, (3) can effectively model any image patch in $\mathbf{y}_{k,i,j}$ by giving a non-zero likelihood, representing image patches unseen in the training data.

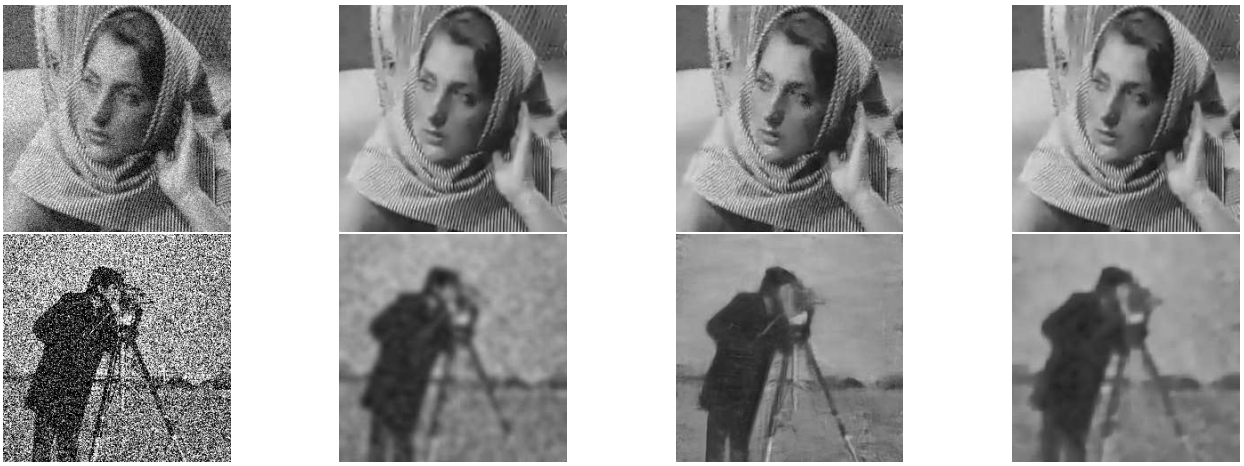
Noisy patches containing unseen images are likely to produce low Equation (1) likelihoods but not necessarily low Equation (3) likelihoods. The presence of the unseen-data likelihood (3) helps to reduce the posterior probability (2) for a mismatch. Conversely, for the noisy patch $\mathbf{y}_{k,i,j}$ with a good match $\mathbf{x}^m_{k,u,v}$, the likelihood $p(\mathbf{y}_{k,i,j}|\mathbf{x}^m_{k,u,v})$ is greater than $p(\mathbf{y}_{k,i,j}|\phi_k)$. This is because

$$p(\mathbf{y}_{k,i,j}|\phi_k) \simeq a\prod_n P(x^m_{k,u,v}(n))\exp\big(-\frac{(y_{k,i,j}(n)-x^m_{k,u,v}(n))^2}{h^2}\big) \quad (4)$$

and this is smaller than (1) as $P(x^m_{k,u,v}(n))<1$. In (4), we assume that the closely matched, and highly likely, training patch dominates the mixture in (3). Therefore, a good match will dominate (2) and produce a large posterior probability.

Importantly, we can further show that the posterior probability increases as the size of the matching patches increases, therefore it can be used to identify the largest matching patches. To show this, assume $\mathbf{y}_{k,i,j}$ and $\mathbf{x}^m_{k,u,v}$ are a pair of matching patches in terms of the greatest likelihood, i.e., $p(\mathbf{y}_{k,i,j}|\mathbf{x}^m_{k,u,v}) \geq p(\mathbf{y}_{k,i,j}|\mathbf{x}^{m'}_{k,u',v'})$ for any $\mathbf{x}^{m'}_{k,u',v'} \neq \mathbf{x}^m_{k,u,v}$, and $p(\mathbf{y}_{k,i,j}|\mathbf{x}^m_{k,u,v}) \geq p(\mathbf{y}_{k,i,j}|\phi_k)$. Further, assume $\mathbf{y}_{\eta,i,j}$ is a sub-patch in $\mathbf{y}_{k,i,j}$ with the same origin but $\eta\leq k$, and $\mathbf{x}^m_{\eta,u,v}$ is a corresponding sub-patch in $\mathbf{x}^m_{k,u,v}$. We can have the following likelihood ratio inequality

$$\frac{p(\mathbf{y}_{k,i,j}|\mathbf{x}^m_{k,u,v})}{p(\mathbf{y}_{k,i,j}|\mathbf{x}^{m'}_{k,u',v'})} = \frac{p(\mathbf{y}_{\eta,i,j}|\mathbf{x}^m_{\eta,u,v})p(\tilde{\mathbf{y}}_{\eta,i,j}|\tilde{\mathbf{x}}^m_{\eta,u,v})}{p(\mathbf{y}_{\eta,i,j}|\mathbf{x}^{m'}_{\eta,u',v'})p(\tilde{\mathbf{y}}_{\eta,i,j}|\tilde{\mathbf{x}}^{m'}_{\eta,u',v'})} \geq \frac{p(\mathbf{y}_{\eta,i,j}|\mathbf{x}^m_{\eta,u,v})}{p(\mathbf{y}_{\eta,i,j}|\mathbf{x}^{m'}_{\eta,u',v'})} \quad (5)$$



(a) Noisy (b) SR (c) LMA (d) SR-LMA

Fig. 2: Reconstructed results, from the external approaches, of Barbara and Cameraman, at noise levels $\sigma = 25$ and $\sigma = 100$ respectively, over the TD2 dataset. The noisy image is presented in the first column, followed by results from SR, and our LMA and SR-LMA approaches.

where each sub-patch $\tilde{\mathbf{z}}_\eta$ denotes the complement of the patch $\mathbf{z}_\eta$ against the full patch $\mathbf{z}_k$, and the inequality is due to the assumption that $\tilde{\mathbf{y}}_{\eta,i,j}$ and $\tilde{\mathbf{x}}^m_{\eta,u,v}$ match and hence $p(\tilde{\mathbf{y}}_{\eta,i,j}|\tilde{\mathbf{x}}^m_{\eta,u,v}) \geq p(\tilde{\mathbf{y}}_{\eta,i,j}|\tilde{\mathbf{x}}^{m'}_{\eta,u',v'})$. Based on (4), we show a similar inequality concerning the likelihood ratios of the unseen patches. Rewriting (2) as a function of the appropriate likelihood ratios, and applying the above inequalities, we can obtain the posterior inequality

$$P(\mathbf{x}^m_{\eta,u,v}|\mathbf{y}_{\eta,i,j}) \leq P(\mathbf{x}^m_{k,u,v}|\mathbf{y}_{k,i,j}) \tag{6}$$

i.e., (2) increases along with the matching patch size.

Based on (6), by maximizing the posterior probability at each location (i, j) we can obtain an estimate of the largest noisy patch with a matching training patch, i.e.,

$$\begin{aligned} \mathbf{x}^m_{k,u,v} = \arg\max_{\eta} \max_{m',u',v'} P(\mathbf{x}^{m'}_{\eta,u',v'}|\mathbf{y}_{\eta,i,j}) \\ \text{s.t. } \nabla_\eta P(\mathbf{x}^m_{\eta,u,v}|\mathbf{y}_{\eta,i,j}) \geq 0 \text{ for all } \eta \leq k \end{aligned} \tag{7}$$

We impose a monotonicity constraint on the posterior to prevent the selection of partially matching patches. The derivative of the posterior $P(\mathbf{x}_k|\mathbf{y}_k)$ with respect to size k can be calculated by using a regression formula

$$\nabla_k P(\mathbf{x}_k|\mathbf{y}_k) = \frac{\sum_{\eta=1}^{\theta} \eta[P(\mathbf{x}_{k+\eta}|\mathbf{y}_{k+\eta}) - P(\mathbf{x}_{k-\eta}|\mathbf{y}_{k-\eta})]}{2\sum_{\eta=1}^{\theta} \eta^2} \tag{8}$$

In our experiments $\theta = 2$. As indicated, the optimal $\mathbf{x}^m_{k,u,v}$ is obtained by iteratively finding the most probable fixed-size patch and increasing the patch size (i.e., k), until the maximum posterior probability is found, subject to the monotonicity constraint. Obtaining the matching training patch at each location, we finally estimate the target image by averaging the overlapping matching patches.

2.3. Extension to other denoising approaches

In order to combine SR's relative invariance to the training data and the potential for improved noise robustness of LMA, we investigate a combined SR-LMA approach. We learn dictionaries at various patch-sizes from training images. Given a noisy image, we sparsely reconstruct the target image at each patch size, and select the patch size with the highest posterior probability, at each location. That is, in the likelihood function (1) we replace the example patches $\mathbf{x}^m_{k,u,v}$ with the sparsely reconstructed patches for $\mathbf{y}_{k,i,j}$.

Next, we consider the use of the LMA approach for internal image denoising. BM3D groups similar patches in a noisy image to perform 3D collaborative filtering. We suggest the use of the LMA approach to help identify the similar patches by finding the largest matching patches at each location. Then we filter using the standard BM3D algorithm on fixed-size patches at the co-ordinates of the selected patches. We call this combined approach BM3D-LMA.

3. EXPERIMENTAL SETTINGS AND RESULTS

To validate our LMA algorithm we performed tests on four standard test images - Barbara, Cameraman, Parrots and Boats [26], degraded by Gaussian white noise. For external approaches we used two generic training sets TD1 and TD2 from [26]. Each set contains five natural images with varying contents. For comparison, we used the K-SVD Matlab Toolbox provided in [27] to generate SR results using dictionaries of 256 8×8 atoms from each training set. To implement the SR-LMA approach we learned dictionaries at each size from 7×7 to 21×21, and adapted the SR denoising code to reconstruct with the largest matching sparse reconstructions.

Considering four noise levels, $\sigma = 10, 25, 50, 100$, we generated results for each image over three realisations of the noise. For the example-based approach described in Section 2.2, as in SR, we include the weighted noisy image in reconstruction. We tuned the optimal upper and lower limits of k to be searched. In general, low search limits give better reconstruction for lower-level noise as this avoids over-smoothing, while higher-level noise requires higher search limits to avoid local maxima. In all experiments, h took a value approximately equalling the noise σ.

Example reconstructions are presented in Fig. 2, with a summary of the results by averaging over all images and training datasets in Table 1. SR-LMA is shown to outperform SR in terms of PSNR and SSIM in Table 1. In Fig. 2(d) SR-LMA produces reduced artefacts and clearer structure (e.g., the tripod). Table 1 shows that example-based LMA outperforms SR at high noise levels (e.g., $\sigma = 50, 100$), in which the correlation between neighbouring pixels is heavily altered, requiring larger patches to identify the underlying structure. The reduced blotchy-artefacts in Fig. 2(c) illustrate that LMA also outperforms SR-LMA at very high noise (e.g., $\sigma = 100$). This sug-

gests that the diversity of image contents, at the large patch-sizes required to identify the underlying signal, is less well-represented in the learned dictionaries of 256 atoms than in the example images. The SR-based Cameraman examples in Fig. 2 introduce blur, while example-based LMA displays sharp edges. This artefact suggests LMA's potential in the application of image deblurring.

LMA performance is dependent on image contents. Fig. 2(c) shows loss of detail in the Barbara image - the pattern of the scarf is degraded by noisy artefacts. The absence of similar structures in training forces small matches, resulting in a noisy reconstruction. This problem is effectively solved in the SR-LMA approach, showing improved accuracy over both LMA and the original SR approach. Smooth images (e.g., Cameraman) avoid the rare patch problem and can be more accurately reconstructed on generic datasets. The added benefit of our example-based approach is that, as no dictionary learning is needed, targeted training data can be easily employed to improve performance [28].

Table 1: PSNR and SSIM of reconstructions by LMA, SR-LMA and SR at four noise levels, averaged over our four test images and two datasets.

σ	SR		LMA		SR-LMA	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
10	33.81	0.9284	32.82	0.9132	33.95	0.9280
25	28.83	0.8345	28.46	0.8295	28.93	0.8382
50	24.30	0.6961	24.83	0.7174	24.76	0.7162
100	19.76	0.5472	20.24	0.5781	20.05	0.5668

Table 2: PSNR and SSIM of reconstructions by BM3D-LMA, NLM and BM3D at three noise levels, averaged over our three test images and two datasets.

σ	NLM		BM3D		BM3D-LMA	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
10	33.37	0.9074	33.99	0.9370	34.60	0.9370
25	28.85	0.8099	29.72	0.8642	29.90	0.8612
50	24.30	0.7038	26.02	0.7575	25.68	0.7578

To compare internal approaches, we obtained NLM results from the source code and optimal parameters in [29] and BM3D results from the C++ implementation in [14], also adapting this code to our BM3D-LMA approach. We generated results averaged over three instances of the noise, at three noise levels, $\sigma = 10, 25, 50$. Table 2 includes a summary of the results by averaging across all test data at each noise level and Fig. 3 presents example images. Despite the BM3D algorithm approaching the theoretical limits of reconstruction, in terms of PSNR [30, 31], BM3D-LMA achieves minor improvements at low noise. The reconstructions of Barbara in Fig. 3 show little visible difference in quality. Small image patches which are dominated by noise, eg. smooth regions, can be better reconstructed with internal patches, while patches which are dominated by signal are better reconstructed with clean external data [1]. The LMA approach attempts to find a patch-size which is dominated by the clean signal for matching. Our extension may be improved by calculating the similarity criterion in the frequency domain as in BM3D, or by exploiting variable-sized patches in the filtering step. However, we deduce that the benefits of the LMA algorithm are more applicable to the case of external denoising.

Fig. 4 shows the average percentage of patches selected at each size (k) by LMA reconstruction of our test images over TD2, at four noise levels. The preference for larger patch-sizes at high noise levels is clear, but at each level the selected patch-sizes vary. Eg., at

(a) BM3D

(b) BM3D-LMA

Fig. 3: Reconstructions of Barbara, at noise level $\sigma = 25$

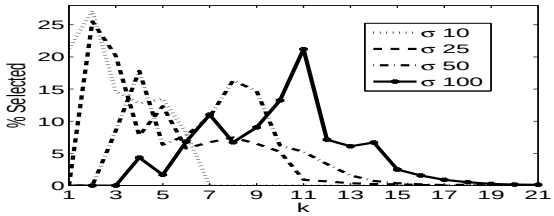


Fig. 4: Average percentage of patches selected at each size $(2k + 1) \times (2k + 1)$ by LMA, on the TD2 dataset over four images

$\sigma = 50$ we can see two peaks, where structured areas with no large matches in the dataset are reconstructed with small patches, while smooth regions can be accurately reconstructed with large patches. The patch-sizes are not focused on the optimum fixed patch-size in Fig. 1, but vary across locations, allowing for textured and smooth regions to be denoised appropriately. To illustrate this, Fig. 5 shows the patch sizes selected at each location in an example image. In Fig. 1, alongside the fixed-size results we include the LMA based results, indicating the LMA capability of automatically adapting to different noise conditions. At $\sigma = 100$, very large fixed-size patches are shown to achieve slightly higher PSNR, suggesting that the high ambiguity is directing the LMA approach to local maxima in smooth regions. However, the higher SSIM, a more accurate measure of visual quality, indicates that fine details are still better maintained by the LMA approach.

4. CONCLUSIONS

We have presented a method for jointly optimizing the patch size and location for improved noise robustness in image denoising. With generic datasets and single noisy inputs as training, we showed some performance improvements in our extensions to existing methods. Using largest matching areas makes the method more dependent on the training data. Our example-based approach improved on SR at higher noise levels, particularly in smoother regions, with generic datasets. Our internal denoising extension produced competitive results, to which we have suggested improvements, but the algorithm is more suited to external denoising.

Fig. 5: Patch-sizes selected for each pixel in the cameraman image reconstructed at $\sigma = 25$. Dark pixels indicate small patches, and brighter pixels larger patches (Left). The clean image (Right).

5. REFERENCES

[1] I. Mosseri, M. Zontak, and M. Irani, "Combining the power of internal and external denoising," *IEEE International Conference on Computational Photography*, pp. 1–9, 2013.

[2] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing.*, vol. 54, no. 11, pp. 4311-4322, 2006.

[3] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing.*, vol. 20, no. 1, pp.33 -61, 1998.

[4] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Transactions on Image Processing.*, vol. 17, no. 1, pp. 53-69, 2008.

[5] M. Elad, M. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," *Proceedings of the IEEE.*, vol. 98, no.6, pp. 972-982, 2010.

[6] H. Burger, C. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?," *2012 IEEE Conference on Computer Vision and Pattern Recognition.*, vol. 1, pp.2392 - 2399, 2012.

[7] Y. Wang and J-M. Morel, "Can a single image denoising neural network handle all levels of gaussian noise?," *IEEE Signal Processing Letters.*, vol. 21, no. 9, pp.1150-1153, 2014.

[8] A. Buades, B. Coll, and J-M. Morel, "Image denoising methods. a new nonlocal principle," *SIAM Review.*, vol. 52 no. 1, pp. 113-147, 2010.

[9] L. Shao, R. Yan, X. Li, and Y. Liu, "From heuristic optimization to dictionary learning: A review and comprehensive comparison of image denoising algorithms," *IEEE Transactions on Cybernetics.*, vol. 44, no. 7, pp.1001-1013, 2014.

[10] C. Sutour, C. Deladelle, and J. Aujol, "Adaptive regularization of the nl-means: Application to image and video denoising," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3506-3520, 2014.

[11] V. Duval, J-F. Aujol, and Y. Gousseau, "On the parameter choice for the non-local means," *Tech. Rep. HAL-00468856, HAL*, 2010.

[12] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing.*, vol. 16, no. 8, pp. 2080-2095, 2007.

[13] A. Buades, B. Coll, and J-M. Morel, "A non local algorithm for image denoising," *IEEE Computer Vision and Pattern Recognition*, vol. 2, pp. 60–65, 2005.

[14] M. Lebrun, "An analysis and implementation of the BM3D image denoising method," *Image Processing On Line*, vol. 2, pp. 175-213, 2012.

[15] V. Bhateja, K. Rastogi, A. Verma, and C. Malhotra, "A non-iterative adaptive median filter for image denoising," *2014 International Conference on Signal Processing and Integrated Networks.*, vol. 1, pp. 113-118, 2014.

[16] M. Zontak, I. Mosseri, and M. Irani, "Separating signal from noise using patch recurrence across scales," *2013 IEEE Conference on Computer Vision and Pattern recognition.*, vol. 1, pp. 1195-1202, 2013.

[17] Z. Jing-Juan, Z. Zuo-Feng, C. Jian-Zhong, and Z. Hui, "Doubly adaptive nonlocal means image denoising algorithm based on mathematical morphology," *Fifth International Conference on Intelligent Human-Machine Systems and Cybernetics*, vol. 1, pp. 301-304, 2013.

[18] T. Thaipanich, O. Byung Tae, W. Ping-Hao, X. Daru, and C. C. J. Kuo, "Improved image denoising with adaptive nonlocal means (anl-means) algorithm," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2623-2630, 2010.

[19] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 2866-2878, 2006.

[20] C. Kervrann and J. Boulanger, "Local adaptivity to variable smoothness for exemplar-based image regularization and representation," *International Journal of Computer Vision.*, vol. 79, no.1, pp.45-69, 2008.

[21] J-F.Aujol V. Duval and Y. Gousseau, "A bias-variance approach for the nonlocal means," *SIAM J. Imaging Sci.*, vol. 4, no. 2, pp. 760–788, 2011.

[22] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "BM3D image denoising with shape-adaptive principal component analysis," *Proc. Workshop on Signal Processing with Adaptive Sparse Structured Representations*, vol. 1, 2009.

[23] V. Duval C-A. Deledalle and J. Salmon, "Non-local methods with shape-adaptive patches (NLM-SAP)," *Journal of Mathematical Imaging and Vision*, vol. 43, no. 2, pp. 103–120, 2011.

[24] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *Multiscale Modelling and Simulation.*, vol. 7, no. 1, pp. 214-241, 2008.

[25] J. Boulanger C. Kervrann and P. Coupe, "Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal," *Proc. Conf. Scale-Space and Variational Methods,*, pp. 520–532, 2007.

[26] W. Dong, D. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Transactions on Image Processing.*, vol. 20, no. 7, pp. 1838-1857, 2011.

[27] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing.*, vol. 15, no. 12, pp. 3736-3746, 2006.

[28] E. Luo, S. H. Chan, and T. Q. Nguyen, "Image denoising by targeted external databases," *IEEE International Conference on Acoustic, Speech and Signal Processing*, vol. 1, pp. 2450-2454, 2014.

[29] A. Buades, B. Coll, and J-M. Morell, "Non-local means denoising," *Image Processing On Line*, vol. 1, 2011.

[30] C. Knaus and M. Zwicker, "Progressive image denoising," *IEEE Transactions on Image Processing*, vol. 23, no. 7, pp. 3114-3125, 2014.

[31] A. Levin and B. Nadler, "Natural image denoising: Optimality and inherent bounds," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2833–2840, 2011.