IMAGE RESTORATION BY MIXTURE MODELLING OF AN OVERCOMPLETE LINEAR REPRESENTATION

L. Mancera[1], S. Derin Babacan[2], R. Molina[1], A. K. Katsaggelos[2]

[1]Departamento de Ciencias
de la Computación e I.A.
Universidad de Granada, 18071 Granada, Spain
mancera@decsai.ugr.es, rms@decsai.ugr.es

[2]Department of Electrical Engineering
and Computer Science
Northwestern University, Evanston, IL 60208, USA
sdb@northwestern.edu, aggk@eecs.northwestern.edu

ABSTRACT

We present a new image restoration method based on modelling the coefficients of an overcomplete wavelet response to natural images with a mixture of two Gaussian distributions, having non-zero and zero mean respectively, and reflecting the assumption that this response is close to be sparse. Including the observation model, the resulting procedure iterates between image reconstruction from the hard-thresholding of the response to the current estimate and a fast blur compensation step. Results indicate that our method compares favorably with current wavelet-based restoration methods.

Index Terms— Image restoration, linear representations, sparsity, hard-thresholding, overcomplete wavelets.

1. INTRODUCTION

Image restoration aims at estimating an image from a degraded observation which is generally modelled as a convolution with a blur matrix plus additive noise. This is an ill-posed problem even if the degradation parameters are known. The choice of a good image prior is basic for obtaining good performance. While methods based on prior modelling in the spatial domain are a very successful approach (e.g., [1]), iterative and global wavelet-based restoration methods (e.g., [2, 3, 4]) and locally adaptive models (e.g., [5, 6]) also produce image estimates of high quality (see [6]).

Sparsity in the synthesis coefficients of a linear generative image model is commonly used as prior for overcomplete wavelet coefficients (e.g., [7, 8, 2, 3]). But assuming *compressibility* (most of the energy concentrated in relatively few coefficients) in the linear analysis representation, has provided better results in different applications (e.g., [9, 10, 11]).

In this paper we follow this *compressibility* approach and model the distribution of the coefficients of an overcomplete wavelet response to natural images as a mixture of two Gaussian distributions, one having a non-zero and the other a zero

mean. They reflect the prior assumption that the response is close to a sparse vector. This prior model, in combination with the observation model, produces an image restoration method which compares favorably, both quantitatively and qualitatively, with current wavelet-based restoration methods.

2. BAYESIAN MODELLING

The image degradation process is represented here by a linear and spatially invariant model as

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n}, \tag{1}$$

where $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{n}$ are, respectively, the $N \times 1$ lexicographically ordered original and observed images and the zero-mean Gaussian noise of variance $\sigma_\mathbf{n}^2$, and $\mathbf{H}$ is a block-circulant matrix. We model the conditional probability $p(\mathbf{y}|\mathbf{x})$ as a Gaussian distribution with mean $\mathbf{Hx}$ and variance $\sigma_\mathbf{n}^2 \mathbf{I}$.

Let $\mathbf{\Phi}^t$ be a full column rank $M \times N$ matrix with $M > N$, satisfying $\mathbf{\Phi}\mathbf{\Phi}^t = \mathbf{I}$. Matrix $\mathbf{\Phi}^t$ represents a wavelet analysis operator, typically producing compressible responses to natural images, so we assume a priori that $\mathbf{\Phi}^t\mathbf{x} = \mathbf{a} + \mathbf{r}$, where $\mathbf{a}$ is a sparse vector and $\mathbf{r}$ a Gaussian correction term. The sparsity of the vector $\mathbf{a}$ is enforced by a random vector $\mathbf{z} = (z_1, \ldots, z_M)^t$ whose components take, independently, the value one with probability $p << 1$. Then, we have the following prior modelling

$$p(\mathbf{z}) = \prod_{i=1}^{M} p^{z_i}(1-p)^{1-z_i}. \tag{2}$$

$$p(\mathbf{x}|\mathbf{z}, \mathbf{a}) = \prod_{i=1}^{M} \left(\frac{1}{\sqrt{2\pi}\sigma_\mathbf{r}} \exp(\frac{-1}{2\sigma_\mathbf{r}^2}(\phi_i^t\mathbf{x} - a_i)^2) \right)^{z_i}$$
$$\cdot \left(\frac{1}{\sqrt{2\pi}\sigma_\mathbf{r}} \exp(\frac{-1}{2\sigma_\mathbf{r}^2}(\phi_i^t\mathbf{x})^2) \right)^{1-z_i}, \tag{3}$$

where $\sigma_\mathbf{r}^2$ is the variance of each $r_i = a_i - \phi_i^t\mathbf{x}$, which is assumed to be constant for all i in this work. Notice that inte-

grating out $\mathbf{z}$ we have

$$
\begin{aligned}
\log p(\mathbf{x}|\mathbf{a}) &= \sum_{i=1}^{M} \left(p\frac{1}{\sqrt{2\pi}\sigma_{\mathbf{r}}} \exp(\frac{-1}{2\sigma_{\mathbf{r}}^2}(\phi_i^t\mathbf{x} - a_i)^2) \right. \\
&+ \left. (1-p)\frac{1}{\sqrt{2\pi}\sigma_{\mathbf{r}}} \exp(\frac{-1}{2\sigma_{\mathbf{r}}^2}(\phi_i^t\mathbf{x})^2) \right). \quad (4)
\end{aligned}
$$

Let us discuss briefly the proposed model. Firstly note that we are defining a prior model not on the image but on a linear transformation of it, $\mathbf{\Phi}^t\mathbf{x}$. Also, we do not assume a prior distribution on $\mathbf{a}$. Furthermore, the second term in the above mixture representation is expected to model those coefficients $\phi_i^t\mathbf{x}$ which are close to zero, while the first term corresponds to high amplitude coefficients. In the next section we perform Bayesian inference and apply the Iterated Conditional Mode (ICM) procedure [12] to estimate $\mathbf{x}$, $\mathbf{a}$ and $\mathbf{z}$.

3. SOLVING THE OPTIMIZATION PROBLEM

A local maximum of $p(\mathbf{x}, \mathbf{a}, \mathbf{z}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\mathbf{z}, \mathbf{a})p(\mathbf{z})$ is obtained by maximizing sequentially in $\mathbf{a}$, $\mathbf{z}$ and $\mathbf{x}$ given the other two vectors. Let $\mathbf{x}(n)$ and $\mathbf{z}(n)$ be the estimates of $\mathbf{x}$ and $\mathbf{z}$ at iteration n. Then, to find $\mathbf{a}(n)$ by maximizing $p(\mathbf{x}(n), \mathbf{a}, \mathbf{z}(n)|\mathbf{y})$ we have to maximize $p(\mathbf{x}(n)|\mathbf{z}(n), \mathbf{a})$ with respect to $\mathbf{a}$ (see Eq. (3)), which produces to

$$
\mathbf{a}(n) = \mathbf{\Phi}^t\mathbf{x}(n). \quad (5)
$$

Now for the second step, given $\mathbf{x}(n)$ and $\mathbf{a}(n)$ we solve

$$
\mathbf{z}(n+1) = \arg\max_{\mathbf{z}} p(\mathbf{x}(n)|\mathbf{z}, \mathbf{a}(n))p(\mathbf{z}). \quad (6)
$$

¿From Eqs. (2), (3) and (5) we have

$$
z_i(n+1) = \begin{cases} 1 & \text{if } \frac{p}{1-p} \geq \exp\left(\frac{-1}{2\sigma_{\mathbf{r}}^2}(\phi_i^t\mathbf{x})^2\right) \\ 0 & \text{otherwise.} \end{cases} \quad (7)
$$

Since $p < 1 - p$, we can write $e^{-\lambda} = \frac{p}{1-p}$, where λ is a positive number. Then the condition for $z_i(n+1) = 1$ in Eq. (7) becomes $|\phi_i^t\mathbf{x}| \geq \theta$, where $\theta = \sigma_{\mathbf{r}}\sqrt{2\lambda}$.

Finally, for the third step we have that $\mathbf{x}(n+1)$ satisfies

$$
\begin{aligned}
\mathbf{x}(n+1) &= \arg\min_{\mathbf{x}} \left\{ \frac{1}{2\sigma_{\mathbf{n}}^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 \right. \\
&+ \frac{1}{2\sigma_{\mathbf{r}}^2}\sum_{i=1}^{M} z_i(n+1)(\phi_i^t\mathbf{x} - a_i(n))^2 \\
&+ \left. \frac{1}{2\sigma_{\mathbf{r}}^2}\sum_{i=1}^{M}(1 - z_i(n+1)(\phi_i^t\mathbf{x})^2 \right\}. \quad (8)
\end{aligned}
$$

Let $\mathbf{S}(n+1)$ be the $M \times M$ diagonal matrix with diagonal entries $S_{i,i}(n+1) = z_i(n+1)$. We can rewrite Eq. (8) as

$$
\begin{aligned}
\mathbf{x}(n+1) &= \arg\min_{\mathbf{x}} \left\{ \frac{1}{2\sigma_{\mathbf{n}}^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 \right. \\
&+ \left. \frac{1}{2\sigma_{\mathbf{r}}^2}\|\mathbf{\Phi}^t\mathbf{x} - \mathbf{S}(n+1)\mathbf{a}(n)\|_2^2 \right\}, \quad (9)
\end{aligned}
$$

and so we obtain

$$
\mathbf{x}(n+1) = (\frac{1}{\sigma_{\mathbf{r}}^2}\mathbf{I} + \frac{1}{\sigma_{\mathbf{n}}^2}\mathbf{H}^t\mathbf{H})^{-1}\frac{1}{\sigma_{\mathbf{r}}^2}\mathbf{\Phi}\mathbf{S}(n+1)\mathbf{a}(n) + \frac{1}{\sigma_{\mathbf{n}}^2}\mathbf{H}^t\mathbf{y}.
$$

4. IMPLEMENTATION

Representation used. For highly-textured images, we observed a good comparative performance by using 6-scale Candes' Curvelets (redundancy factor, r.f. ≈ 7.2). For little-textured images, it is better to combine the 9-scale Translation Invariant Haar Pyramids [6] (TIHP, r.f. ≈ 7) with the 8-scale Kingsbury's Dual-Tree Complex Wavelets (DT-CWT, r.f. $= 4$). This is done by concatenating the coefficients from both representations and dividing by $\sqrt{0.5}$ in order to assure that $\mathbf{\Phi}\mathbf{\Phi}^t = \mathbf{I}$. Real and imaginary part of each complex coefficient are treated as independent coefficients. Combined representations often provides a higher potential of finding good sparse approximations to natural images (e.g., [10]).

Algorithmic parameters. The selection of the algorithmic parameters λ, $\sigma_{\mathbf{r}}^2$ and $\sigma_{\mathbf{n}}^2$ is crucial for the performance of the algorithm. We have assumed known or previously estimated (e.g., [1]) the observation noise variance $\sigma_{\mathbf{n}}^2$. The variance $\sigma_{\mathbf{r}}^2$ could be estimated by the variance of $\{\phi_i^t\mathbf{y}\}_{i=1,...,M}$, but this is probably a poor estimate of the true value of $\sigma_{\mathbf{r}}^2$.

We will now examine how λ and $\sigma_{\mathbf{r}}^2$ are updated in the iterative procedure. We have analyzed iterative schemes for updating both parameters. By initializing the threshold θ to half the maximum amplitude of $\mathbf{\Phi}^t\mathbf{y}$, $\theta(0)$, and using a large initial value of λ, $\lambda(0)$, we obtain the initial value of $\sigma_{\mathbf{r}}^2$, $\sigma_{\mathbf{r}}^2(0)$, from the equation $\theta(0) = \sigma_{\mathbf{r}}(0)\sqrt{2\lambda(0)}$. The large value of λ provides a very poor restoration since very few coefficients are taken into account. Notice that, from $e^{-\lambda(0)} = \frac{p(0)}{1-p(0)}$, the probability $p(0)$ is very small. To add more coefficients we can decrease the values of λ and $\sigma_{\mathbf{r}}$ at each iteration by an exponential factor. Notice that this implies that at each iteration $p(n)$ increases and the sequence of thresholds $\theta(n) = \sigma_{\mathbf{r}}(n)\sqrt{2\lambda(n)}$ is a decreasing one. Similar dynamic thresholding strategies have been successfully used in several applications (e.g. [10, 11]). Instead of using two decreasing factors, we have experienced better results by decreasing $\sigma_{\mathbf{r}}^2$ by a factor of 0.9 and fixing λ for the whole iterative process. Notice that fixing λ amounts to providing an estimate of the number of active coefficients.

The iterations are stopped when a maximum number of iterations is reached (e.g., 200 is used in the experiments) or the estimated error is above the known variance of the noise, that is, $\|\mathbf{y} - \mathbf{H}\mathbf{x}(n)\|_2^2 \geq \sigma_{\mathbf{n}}^2$.

5. EXPERIMENTAL RESULTS

We have used one 512×512 highly textured test image, namely *Barbara* and two 256×256 less textured ones, *Cameraman* and *House*. Restoration quality is measured

Blur	PSF 1		PSF 2	PSF 3	PSF 4	PSF 5
$\sigma_\mathbf{n} \Rightarrow$	$\sqrt{2}$	$\sqrt{8}$	$\sqrt{0.308}$	7	2	8
	HOUSE					
PSNR	25.62	25.47	24.11	28.10	27.83	30.01
[2]	8.47	6.63	*10.71*	4.22	*4.49*	4.76
[6]	*8.64*	*7.03*	9.04	*4.30*	4.11	*6.02*
[4]	**9.05**	**7.64**	*10.71*	**5.10**	4.03	**6.13**
Proposed	8.13	6.32	**10.80**	4.22	**4.74**	3.89
	CAMERAMAN					
PSNR	22.23	22.16	20.77	24.63	23.26	29.83
[8]	6.93	4.88	7.59	**2.94**	-/-	-/-
[2]	*7.46*	5.24	8.16	*2.84*	3.18	3.65
[6]	7.45	*5.55*	7.33	2.73	**3.25**	*4.19*
[4]	**8.25**	**6.34**	*8.57*	2.56	3.05	**5.15**
Proposed	7.20	5.31	**8.89**	2.51	*3.22*	2.12
	BARBARA					
PSNR	23.33	23.26	22.49	24.21	23.77	29.77
[2]	3.76	1.99	3.98	0.9	0.92	2.55
[6]	**6.85**	*3.80*	*5.07*	**1.94**	**1.36**	**5.27**
[4]	5.73	3.01	4.88	1.58	0.91	*4.04*
Proposed	*6.24*	**3.83**	**5.08**	*1.22*	*1.03*	3.72

Table 1. Performance comparison in terms of increase of the Signal-to-Noise ratio, in dB. First row shows the blurring kernel used, and the second the noise standard deviation. Best results are shown in bold, and second best in italic.

numerically using the Peak Signal-to-Noise Ratio (PSNR), defined as $10 \cdot log_{10}(255^2/MSE)$ with MSE the mean square error. We compare the performance of the proposed method with the algorithms proposed in [2], [8], [6] and [4], which represent the current state-of-the-art in wavelet-based image restoration. Table 1 shows the increase in PSNR values obtained by all algorithms for a number of noise levels and with different PSFs. PSF1 is $h_{i,j} = (1 + i^2 + j^2)^{-1}$, for $i, j = -7, ..., 7$; PSF2 is a 9×9 uniform kernel; PSF3 is a 5×5 separable kernel with coefficients $[1, 4, 6, 4, 1]/16$; PSF4 and PSF5 are Gaussian kernels with standard deviations 1.6 and 0.4. The values for the noise standard deviation are shown in the second row. Finally, the first row for each image shows the PSNR values of the observed images.

From Table 1, note that our method outperforms simultaneously the four methods we have compared with in 5 out of the 18 tests. Also, it is the best in most cases among global ones, which minimize non-convex norms (≈ 0.7) of synthesis coefficients ([8, 2]). When using a more textured image, such as *Barbara*, the performance of the proposed method is comparably better, mainly because our method is very well fitted to preserve sharp edges.

Figure 1 shows a visual comparison of the results for the *House* image among the methods in Table 1 for the severely blurred case (PSF2). It can be observed that, compared to the other methods, the image restored by the proposed method exhibits almost no artifacts, even around the edges, thus pro-
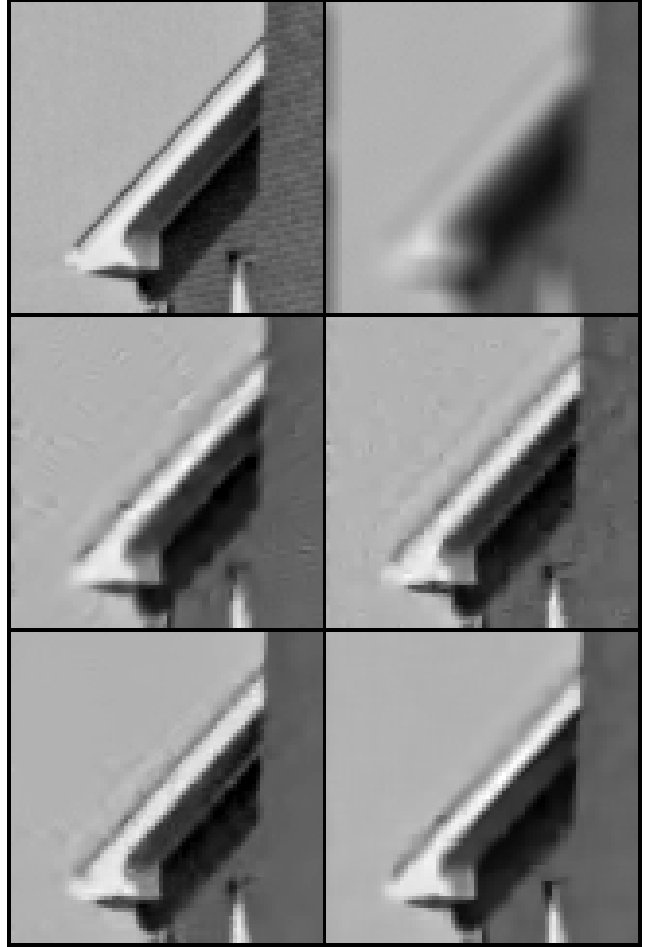
Fig. 1. Visual comparison of the various methods. From top to bottom and from left to right: 80×80 crop of *House* starting at (78,1); degraded with PSF2 and $\sigma_\mathbf{n} = \sqrt{0.308}$ (PSNR = 24.11 dB); result by [6] (PSNR = 33.15 dB); result by [2] (PSNR = 34.82 dB); result by [4] (PSNR = 34.82 dB); proposed method (PSNR = 34.91 dB).

viding a very high visual quality. It is also important to point out that we have seen this absence of visual artifacts in the results of our method in all the experiments we have performed, and that it is in part due to the combined use of two wavelet representations. Another example is shown in Figure 2. Note that the proposed method achieves a better reconstruction of the stripes than the rest of the methods, including [6], which provides a slightly higher PSNR due to local adaptivity.

6. CONCLUSIONS

We have proposed an image restoration method based on a MAP approach with a prior on the compressibility of the linear response of overcomplete wavelets to natural images. We have derived an efficient method providing local optima to the resulting non-convex cost function. Our method performs

comparably to state-of-the-art wavelet-based methods, both visually and in terms of PSNR. Particularly, it clearly outperforms other iterative global methods. Future work includes estimation of the optimal algorithmic parameters and incorporation of local adaptivity for better texture recovery.

7. REFERENCES

[1] S.D. Babacan, R. Molina, and A. Katsaggelos, "Parameter estimation in TV image restoration using variational distribution approximation," *IEEE Trans. on Im. Proc.*, vol. 17, no. 3, pp. 326–339, Mar. 2008.

[2] M.A.T. Figueiredo and R.D. Nowak, "A Bound Optimization Approach to wavelet-based image deconvolution," in *IEEE Int. Conf. on Im. Proc.*, vol. 2, pp. 782–785.

[3] J.M. Bioucas-Dias and M.A.T. Figueiredo, "A new TwIST: Two Step Iterative Shrinkage/Thresholding algorithms for image restoration," *IEEE Trans. on Im. Proc.*, vol. 16, no. 12, pp. 2992–3004, Dec. 2007.

[4] A. Foi, K. Dabov, V. Katkovnik, and K. Egiazarian, "Shape-Adaptive DCT for denoising and image reconstruction," in *SPIE Electronic Imaging 2006, Im. Proc.: Alg. and Syst. V*, 2006, pp. 6064A–18.

[5] J. Portilla and E.P. Simoncelli, "Image restoration using Gaussian scale mixtures in the wavelet domain," in *IEEE Int. Conf. on Im. Proc.*, Barcelona, Spain, Sep. 2003, vol. 2, pp. 965–968.

[6] J.A. Guerrero-Colón, L. Mancera, and J. Portilla, "Image restoration using space-variant Gaussian Scale Mixtures in overcomplete pyramids," *IEEE Trans. on Im. Proc.*, vol. 17, no. 1, pp. 27–41, Jan. 2008.

[7] P. Moulin and J. Liu, "Analysis of multiresolution image de-noising schemes using generalized-Gaussian and complexity priors," *IEEE Transactions on Information Theory, Special Issue on Multiscale Analysis*, vol. 45, no. 3, pp. 909–919, Apr. 1999.

[8] M.A.T. Figueiredo and R. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans, on Im. Proc.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.

[9] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inv. Prob.*, vol. 23, no. 3, pp. 947–968, Jun. 2007.

[10] J. Bobin, J.L. Starck, J. Fadili, Y. Moudden, and D.L. Donoho, "Morphological Component Analysis: an adaptive thresholding strategy," *IEEE Trans. on Im. Proc.*, vol. 16, no. 11, pp. 2675–2681, Nov. 2007.

[11] L. Mancera and J. Portilla, "Non-convex sparse optimization through deterministic anneling and applications," in *IEEE Int. Conf. on Im. Proc.*, San Diego, CA, 12-15, Oct. 2008, pp. 917–920.

[12] J. Besag, "On the statistical analysis of dirty pictures (with discussions)," *J. of the Royal Stat. Soc., Series B*, pp. 259–302, 1986.
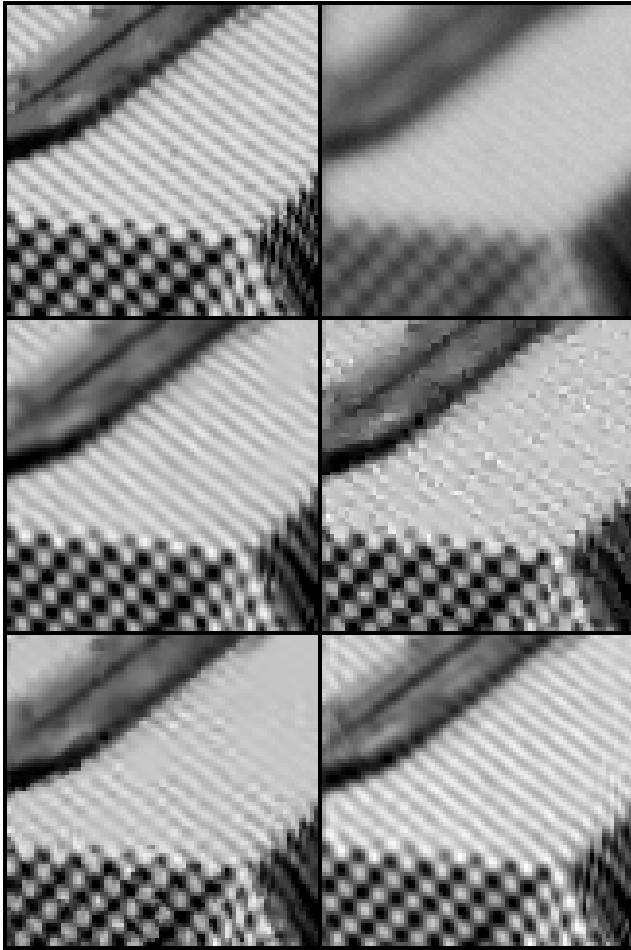
Fig. 2. Visual comparison of the various methods. From top to bottom and from left to right: 80×80 crop of *Barbara* starting at (221,1); degraded with PSF1 and $\sigma_{\mathbf{n}} = \sqrt{2}$ (PSNR = 23.33 dB); result by [6] (PSNR = 30.18 dB); result by [2] (PSNR = 27.09 dB); result by [4] (PSNR = 29.06 dB); proposed method (PSNR = 29.57 dB).