Department of Electrical
and
Computer Systems Engineering

Technical Report
MECSE-1-2002

A Simple Pixel-Adaptive Bayesian Approach to Image Denoising Using Wavelet Interscale Dependency

P. Chen and D. Suter

MONASH
UNIVERSITY

A Simple Pixel-Adaptive Bayesian Approach to Image Denoising Using Wavelet Interscale Dependency

Pei Chen and David Suter

ABSTRACT

In this paper, an approach to image denoising is proposed by utilizing the interscale dependency of wavelet coefficients. We first analyze a new property of parent/children-type statistics in the wavelet domain. Then, a Gaussian mixture model (GMM) is employed to fit this statistical property, where the children's variance field is estimated by a linear relation involving their parent. Lastly, MMSE estimates for the noisy wavelet coefficients are obtained. We demonstrate by experimental comparisons that this approach compares favorably with other competing approaches. In particular, though our approach does not outperform other approaches at all noise levels, there are occasions when it does outperform other approaches and where it may fall short of outperforming other approaches, the difference in performance is not large.

Keywords: Wavelet transform, image denoising, interscale statistics.

I. INTRODUCTION

Recently, much effort [1-3, 5, 7-9] has been devoted to signal or image Bayesian denoising in wavelet domain. Wavelet-based image denoising typically consists of three steps: discrete wavelet transform (DWT), denoising processing of the noisy wavelet coefficients, and inverse discrete wavelet transform (IDWT). Irrespective of what wavelet is employed, the first and the last steps are 1-1 mappings, and are of no interest in this paper. So, only the processing of the noisy wavelet coefficients will be studied in the following. Specifically, in a general Bayesian wavelet shrinkage approach, a prior is first

specified for the wavelet coefficients of the unknown image, and then Bayesian estimation is computed.

Consider a signal *x*, which is contaminated with noise *n*, so that one observes $y = x + n$. Assume $x \sim N(0, \sigma_x^2)$, $n \sim N(0, \sigma_n^2)$. The well-known minimal mean squared error (MMSE) estimation is as follows [7]:

$$\hat{x}(y) = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_n^2} y \tag{1}$$

However, the simple Gaussian prior cannot fit well the marginal density of wavelet coefficients, which tend to be dominated by a few large ones. The actual density of wavelet coefficients has a marked peak at zero and heavy tails, and it is strongly non-Gaussian. Moreover, wavelet coefficients cannot be independently treated, although orthogonal wavelet transform can be viewed as an approximate Karhunen-Loeve transform and tends to approximately decorrelate the image. The reason is that other joint statistics, such as clustering, persistency, decay property across scale, and strong persistence at finer scales [9], exist between wavelet coefficients of real-life images.

Statistical properties in the wavelet domain, like interscale dependency [5, 9] and intrascale dependency [7], have been successfully applied in image denoising. Especially, the hidden Markov tree (HMT) was employed by Crowse *et al.* [5] to capture the interscale dependency. A disadvantage of HMT is the computational burden of the training stage. In order to overcome this computational problem, a simplified HMT, named as uHMT [9], was proposed by introducing nine meta-parameters. Although the training stage is reduced in uHMT, its performance is close to that of the more

complicated HMT. The reason is that uHMT captures most interscale dependency of real-life images by specifying the nine meta-parameters.

In this paper, we propose a simple pixel-adaptive approach to image denoising by utilizing these interscale dependencies. The complicated hidden Markov model is not needed in our approach. In section II, we review the marginal non-Gaussian property and the interscale statistical properties of wavelet coefficients. In section III, we propose an approach to image denoising by using interscale dependency. First, we give a true picture of the interscale statistics by studying the conditional density of the children on their parent. Then, a linear formula is used to estimate the variance field. Consequently, a 3-modal GMM is employed to fit the non-Gaussian property of these conditional statistics. Finally, we apply MMSE estimator to get the estimates of noisy coefficients. In section IV, we analyze the performance of our approach and compare it with HMT and uHMT. Simulations show the effectiveness of our approach.

II. STATISTICS IN WAVELET DOMAIN

Here we review some statistical properties that are utilized in our approach. A good review can be found in [1, 9].

A. Non-Gaussian property

A four-level decomposition of the image Lena is shown in Fig.1, where white pixels denote large magnitude coefficients. A typical normalized histogram for wavelet coefficients of one subband can be seen in Fig. 2, which comes from subband LH1 of Lena. From fig. 2, a salient non-Gaussian property can be seen in the normalized histogram of the wavelet coefficients. An intuitive explanation for this non-Gaussian property is that real-world images typically consist of smooth regions and some

occasional edges. The smooth regions lead to near-zero coefficients and the edges give rise to a few large ones.

GMM [3] and generalized Gaussian distribution (GGD) [8] are the common tools for modeling this non-Gaussian property of wavelet coefficients. Although GGD is more accurate in fitting the actual density of wavelet coefficients than GMM, the latter is preferred, in this paper, for its simple form. In order to model this non-Gaussian property, Chipman *et al.* [3] assume that wavelet coefficients have a Gaussian mixture density:

$$x \mid \gamma \sim \gamma N(0, c^2\tau^2) + (1-\gamma)N(0, \tau^2) \tag{2}$$

where mixture parameter γ has its own prior distribution $P(\gamma = 1) = 1 - P(\gamma = 0) = p$.

Given the mixture density as (2), MMSE estimator, $\hat{x}$, for the wavelet coefficient, y, is:

$$\hat{x}(y) = P(\gamma = 1 \mid y)\frac{(c\tau)^2}{\sigma^2 + (c\tau)^2}y + P(\gamma = 0 \mid y)\frac{\tau^2}{\sigma^2 + \tau^2}y \tag{3}$$

B. Interscale dependency

Interscale dependency can be easily observed from the interscale similarity in quad-tree decomposition of the image, such as fig. 1. Here, we summarize the analysis tool for this type of dependency: HMT and its variation, uHMT. In a HMT for a wavelet transform, each coefficient is modeled as a Gaussian mixture density:

$$f(x) = \sum_{m=0}^{1} p_S(m)N(0, \sigma_m^2) \tag{4}$$

whose components are associated with a hidden state variable S. The hidden state is depicted by $\mathbf{p} = [p_S(0)\ p_S(1)]^T$. Markovian dependency between these hidden variables can be easily modeled due to the natural tree structure of DWT. By tying both within and

across trees, the HMT model is parameterized by the parent/children state transition matrix A, together with two mixture variances, σ_m, for the wavelet coefficients at each scale and two probabilities for the root state variable at the coarsest scale. The transition matrix A_j (j denotes the level) is of the form:

$$A_j = \begin{bmatrix} p_j^{0\to 0} & p_j^{0\to 1} \\ p_j^{1\to 0} & p_j^{1\to 1} \end{bmatrix} \tag{5}$$

The child's state, $\mathbf{p}_c$, can be computed by transition matrix A and its parent's state, $\mathbf{p}_p$:

$$\mathbf{p}_c = A\mathbf{p}_p$$

A drawback with the HMT is the computational cost. For every image, a training step is needed to get the parameters, such as σ_m for all scales, initial $\mathbf{p}$ for the coarsest scale and A for other scales. Therefore, uHMT [9] was proposed to simplify the HMT by introducing nine meta-parameters $\Theta = \left\{\alpha_S, C_{\sigma_S}, \alpha_L, C_{\sigma_L}, \gamma_S, C_{SS}, \gamma_L, C_{LL}, p_{j_0}^L\right\}$. σ_m, $\mathbf{p}$ and A can be obtained from Θ. Both in the HMT and the uHMT, the crux lies in how to get the parameters, $p_S(m)$ and σ_m^2, for each coefficient. With all these parameters known, the estimation, $\hat{x}$, for a wavelet coefficient, y, is of the following form:

$$\hat{x}(y) = \sum_{m=0}^{1} p(S=m \mid y) \frac{\sigma_m^2}{\sigma_m^2 + \sigma_n^2} y \tag{6}$$

III. THE PROPOSED APPROACH

In order to get a truly pixel-adaptive estimator, we take a detailed look at the interscale dependency between children and their parent. An interesting property is that the children's conditional density, on their parent, is also non-Gaussian, with a marked peak at zero and heavy tails. So, given their parent, wavelet coefficients are modeled as a 3-

modal Gaussian mixture distribution in our approach. The crux of our approach lies in obtaining the conditional density, which shows the statistics of the interscale dependency between children and their parent. It consists of two steps: variance estimation and Gaussian mixture modeling.

A. A true picture of the interscale dependency

An obvious property, easily observed from the wavelet quad-tree decomposition in fig. 1, is its persistency between scales. A child x_c is probably large (small), if its parent x_p has a large (small) value. Consequently, the child has a large (small) variance, assuming it is of zero-mean. In order to get a true picture of the interscale dependency, we'll study the conditional density of the children on their parent. This conditional density, $p(x_c \mid x_p)$, can be obtained as the normalized histogram of the children whose parent equals x_p. Fig. 3 is a typical conditional density of the child in vertical band of level 1, whose parent lies between 10 and 11. This conditional density shows a non-Gaussian property, as the normalized histogram of the whole band.

B. *Variance estimation*

An intuitive observation is that the children are of large/small variance if their parent has a large/small value. Fig. 4 shows this kind of dependency of level 1 on level 2 (the data are obtained from 11 images). Other levels share this near-linear property. In our approach, this dependency is utilized to estimate the variance field by the following simple, but practical, linear formula: $\sigma_c = A\left|x_p\right| + B$, where σ_c is the variance of the child x_c and x_p is the value of the corresponding parent. Usually, in the same level, the horizontal band and the vertical band show similar statistics, while the diagonal band has

a smaller variance assuming the value of the parent is same, as it can be seen from fig. 4. So, the true model becomes $\sigma_{c,level,band} = A_{level,band}\left|x_p\right| + B_{level,band}$. For the coarsest level, no parent exists. The parent-on-child dependency is utilized to estimate the variance field of the coarsest level. A similar linear relationship is specified for the parent-on-child dependency, with different *A*s and *B*s. Table 1 lists the experimental parameters for *A* and *B*, which are obtained from 11 real-life images.

C. Gaussian mixture model

Like the marginal density of wavelet coefficients, the conditional densities of $p(x_c \mid x_p)$ and $p(x_p \mid x_c)$ also have the non-Gaussian property, which has a marked peak at zero and heavy tails. As in Chipman *et al.* [3], GMM is used to fit this type of non-Gaussian property. If the variance σ for *x* is known, the following mixture model is specified to fit this non-Gaussian property,

$$x \sim a_1 N(0,\sigma_1^2) + a_2 N(0,\sigma_2^2) + a_3 N(0,\sigma_3^2) \tag{7}$$

where $\sigma_1 = \sigma / n_1$, $\sigma_2 = \sigma$ and $\sigma_3 = n_3\sigma$. To make the variance unchanged, the relation $a_1 / n_1^2 + a_2 + a_3 n_3^2 = 1$ has to hold. We experimentally find that $a_1 = 0.6$, $a_2 = 0.3$, $a_3 = 0.1$ and $n_1 = n_3 = 2.5$ work well for most of the bands and levels. Fig. 3 shows an example and its fitted mixture model. The solid line denotes the conditional density, the dotted one for the fixed mixture model, and the dashed one for the Gaussian model. The marked peak at zero and heavy tails suggest the correctness of the mixture model. Similar to (6), the following MMSE estimator is used to estimate the noisy coefficients

$$\hat{x}(y) = \sum_{m=1}^{3} p(m \mid y) \frac{\sigma_m^2}{\sigma_m^2 + \sigma_n^2} y \tag{8}$$

where $p(m \mid y) = \frac{a_m g(y, \sigma_m)}{\sum_{i=1}^{3} a_i g(y, \sigma_i)}$, $g(y, \sigma_m) = \frac{1}{\sqrt{2\pi}\sigma_m} \exp(-\frac{y^2}{\sigma_m^2})$

D. Noise

For a noisy image, the noise-free wavelet coefficients are unknown and it is impossible to obtain the true variances σ of its children/parent in the model mentioned above. A substitute $\hat{\sigma}$ for σ has to be obtained from the noisy wavelet coefficients. In order to minimize the effect of noise, the following top-down strategy is employed in our approach. First, the detail coefficients of the coarsest level are estimated based on the average of the absolute value of its four noisy children. Then, from coarse to fine, the estimation of the parent $\hat{x}_p$ is used to compute the variances of its children.

IV. RESULTS and CONCLUSION

As in [9], Daubechies' length-8 wavelet D4 [6] is employed to decompose the images into four levels. To evaluate the proposed denoising scheme, we compare it with uHMT and HMT. In table 2-4, the results of PSNR for 11 images* [9] of 256×256 are listed. At noise level of 0.05 (table 2) and 0.1 (table 3), our approach performs better than uHMT, with an average improvement of 0.3~0.4dB; and slightly better than HMT, with an average improvement of 0.1~0.2 dB. From table 4, we can see that these three approaches perform approximately same when the noise is very strong.

A visual display of the image "bridge" can be seen in fig. 5-7, where the PSNR indexes of these three approaches are almost the same. The authors have conducted blind tests over a small number of colleagues and most regard the output from our approach to be slightly, but noticeably, better. For example, the shadow of the bridge and the waterfall in

the images are better preserved by our approach. So, although somewhat subjective, we claim that the visual quality of the output of our approach is slightly superior to that of the other approaches. Of course, there are problems generating from such a small sample and also due to the subjectivity itself.

An improvement of about 1dB, obtained by using the shift-invariant wavelet transform, has been reported in [2, 9]. The extension of our approaches to shift-invariant wavelet transform is on the list of our future work.

Just as with uHMT, the success of our approach can be attributed to the fact that our adaptive model can capture the statistical properties of real-life images. Secondary properties [9], non-Gaussianarity and persistency, are respectively reflected by GMM and the linear estimation of the variance field. Moreover, the tertiary properties [9], decay property across scale and strong persistence at finer scales, coincide with the fact that A and B of coarse scales are larger than those of fine scales in table 1.

REFERENCES

1. R. W. Buccigrossi, and E. P. Simoncelli, 'Image compression via joint statistical characterization in the wavelet domain', IEEE Image Process., Vol. 8, No 12, Dec. 1999, pp. 1688-1701.
2. S. G. Chang, B. Yu, and M. Vetterli: 'Spatially adaptive wavelet thresholding with context modeling for image denoising', IEEE Image Process., Vol. 9, No 9, Sep. 2000, pp. 1522-1531.
3. H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch: 'Adaptive Bayesian wavelet shrinkage', J. Amer. Stat. Assoc., Vol. 92, No 440, Dec. 1997, pp. 1413-1421.

[*] Only Lena in 11 test images is used to obtain the parameters A and B in table 1.

4. T. M. Cover and J. A. Thomas, Elements of Information Theory. New York, NY: John Wiley and Sons, Inc., 1991.
5. M. S. Crouse, R. D. Nowak, and R. G. Baraniuk: 'Wavelet-based statistical signal processing using hidden Markov Models', IEEE Sig. Process., Vol. 46, No 4, Apr. 1998, pp. 886-902.
6. I. Daubechies: 'Ten lectures on wavelet', Philadelphia, SIAM, 1992.
7. M. K. Mihcak, I. K. Kozintsev, K. Ramchandran, and P. Moulin: 'Low-complexity image denoising based on statistical modeling of wavelet coefficients', IEEE Sig. Process. Lett., Vol. 6, No 12, Dec. 1999, pp. 300-303.
8. P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors", IEEE Infor. Theory, Vol. 45, No 3, Apr. 1999, pp. 909-919.
9. J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models", IEEE Image Process., Vol. 10, No 7, Jul. 2001, pp. 1056-1068.

Table captions

Table 1: Parameters for interscale child-on-parent/parent-on-child dependency.

Table 2: Comparison of PSNR for different approaches with noise $\sigma_n = 0.05$

Table 3: Comparison of PSNR for different approaches with noise $\sigma_n = 0.1$

Table 4: Comparison of PSNR for different approaches with noise $\sigma_n = 0.2$

Figure captions

Fig. 1 Four-level wavelet quad-tree decomposition of the image Lena.

Fig. 2 Normalized histogram of the horizontal band in level 1. The solid line denotes the normalized histogram, the dotted one for the mixture model, and the dashed one for the Gaussian model.

Fig. 3 The conditional density of the child in the vertical band of level 1, whose parent lies between 10 and 11. The solid line denotes the conditional density, the dotted one for the fixed mixture model, and the dashed one for the Gaussian model.

Fig. 4 Variance field estimation in level 1. The solid line denotes the horizontal band of level 1, the dotted one for the vertical band and the dashed one for the diagonal band.

Fig. 5 Image bridge with its noisy and denoised copies, noise $\sigma_n = 0.05$.

Fig. 6 Image bridge with its noisy and denoised copies, noise $\sigma_n = 0.1$.

Fig. 7 Image bridge with its noisy and denoised copies, noise $\sigma_n = 0.2$.

Table 1: Parameters for interscale child-on-parent/parent-on-child dependency.

Level	LH		HL		HH	
	A	B	A	B	A	B
1	3.5	0.26	3.7	0.26	2.3	0.15
2	8.5	0.38	10	0.41	6.5	0.29
3	24	0.35	30	0.35	13.5	0.6
4	60	1.9	62	2.4	37	1.1

In page 7

Table 2: Comparison of PSNR for different approaches with $\sigma_n = 0.05$

image	proposed	uHMT	HMT
baby	32.0	32.4	**32.6**
birthday	**30.3**	28.9	29.1
boats	**31.0**	30.4	30.6
bridge	28.1	28.1	**28.3**
buck	**32.8**	32.5	**32.8**
building	29.7	29.7	**30.0**
camera	30.2	30.3	**30.5**
clown	30.7	30.6	**30.9**
fruit	32.4	32.2	**32.6**
kgirl	**31.8**	31.6	**31.8**
Lena	30.4	30.4	**30.5**
average	***30.66***	*30.43*	*30.65*

In page 8

Table 3: Comparison of PSNR for different approaches with $\sigma_n = 0.1$

Image	proposed	uHMT	HMT
baby	28.8	28.9	**29.2**
birthday	**27.4**	25.8	25.8
boats	**27.3**	26.4	26.5
bridge	24.8	24.6	**25.0**
buck	**28.8**	28.4	28.6
building	**26.3**	25.9	**26.3**
camera	**26.5**	26.2	26.4
clown	**27.0**	26.8	26.8
fruit	**28.8**	28.5	28.6
kgirl	**28.8**	28.3	28.3
Lena	**26.9**	26.7	26.7
average	***27.22***	*26.76*	*26.92*

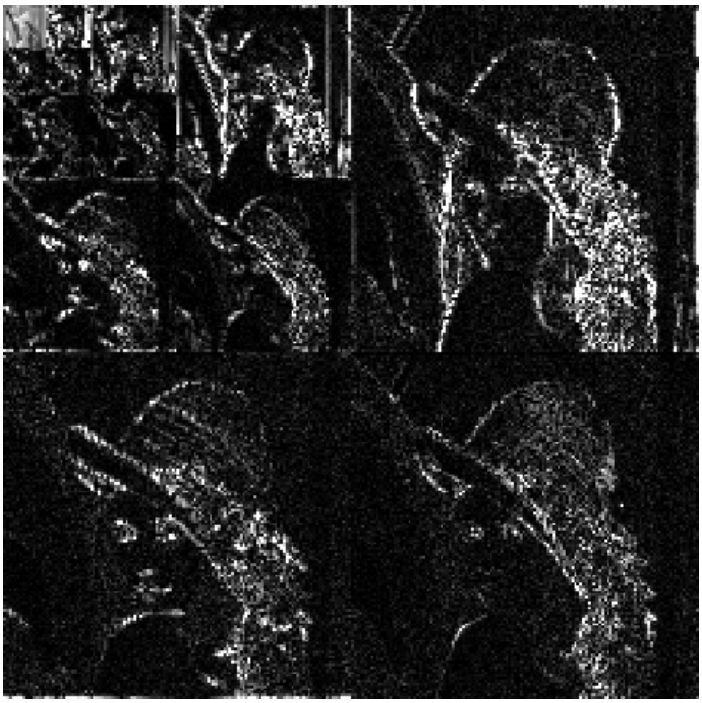
In page 8

Table 4: Comparison of PSNR for different approaches with $\sigma_n = 0.2$

image	proposed	uHMT	HMT
baby	**25.9**	25.8	25.4
birthday	**24.9**	23.1	23.0
boats	**24.1**	23.3	23.3
bridge	22.0	22.0	**22.2**
buck	**25.2**	24.7	24.5
building	**23.0**	22.8	**23.0**
camera	**23.3**	23.1	23.2
clown	23.6	**23.7**	23.6
fruit	**25.5**	25.3	25.0
kgirl	**25.4**	**25.4**	25.3
Lena	**24.1**	23.8	23.8
average	***24.11***	*23.76*	*23.73*

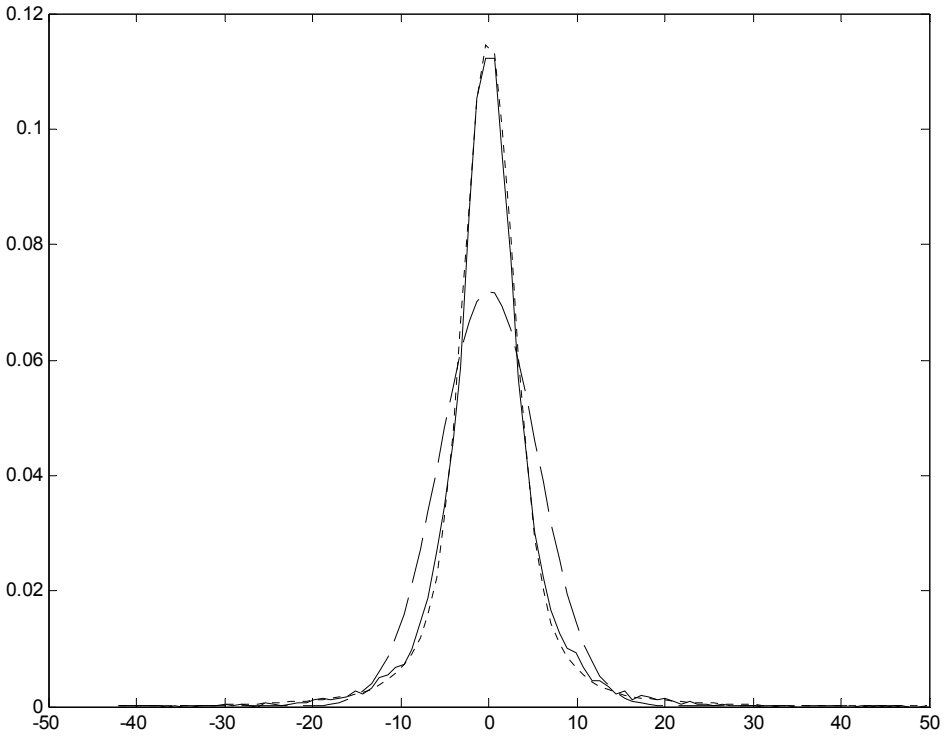
In page 8

Fig. 1



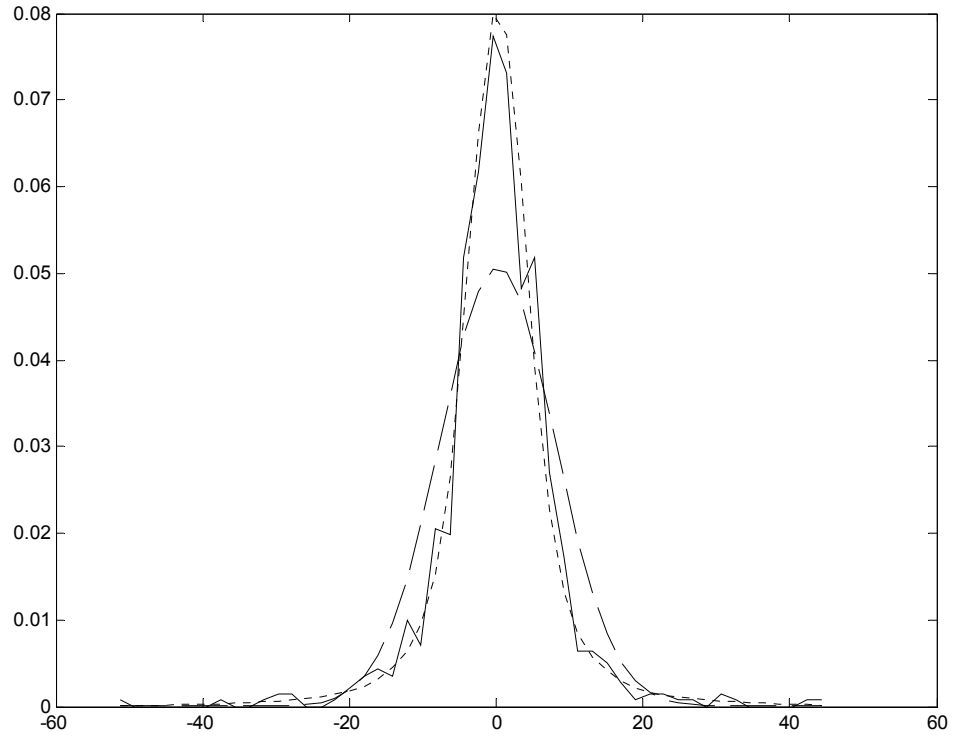
In page 3

Fig. 2



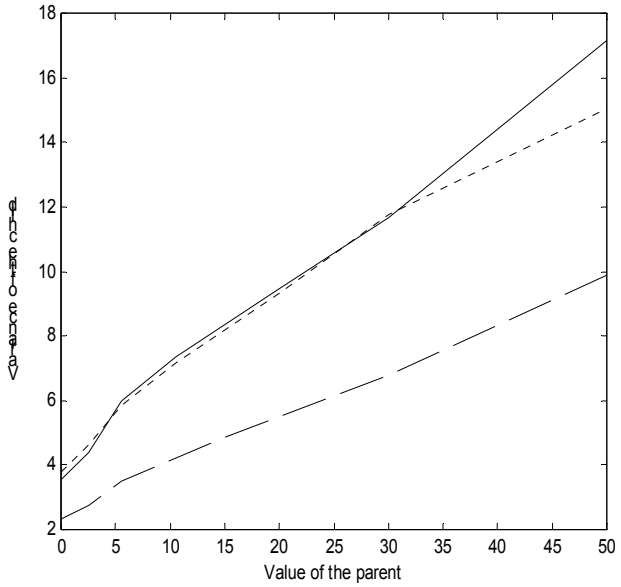
In page 3

Fig. 3



In page 6

Fig. 4



In page 6

Fig. 5

Original bridge

Noised bridge,
with noise=0.05

Denoised bridge,
by proposed approach
PSNR=28.1

Denoised bridge,
by uHMT
PSNR=28.1

Denoised bridge,
by HMT
PSNR=28.3

In page 8

Fig. 6

Original bridge

Noised bridge,
with noise=0.1

Denoised bridge,
by proposed approach
PSNR=24.7

Denoised bridge,
by uHMT
PSNR=24.6

Denoised bridge,
by HMT
PSNR=25.0

In page 8

Fig. 7

Original bridge

Noised bridge,
with noise=0.2

Denoised bridge,
by proposed approach
PSNR=22.0

Denoised bridge,
by uHMT
PSNR=22.0

Denoised bridge,
by HMT
PSNR=22.2

In page 8