Image Denoising Using Bandelets and Hidden Markov Tree Models

ZHANG Wenge¹,², WANG Suang²,³, LIU Fang¹,², GAO Xinbo² and JIAO Licheng²,³
(¹School of Computer Science and Technology, Xidian University, Xi’an 710071, China)
(²Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China)
(³Institute of Intelligent Information Processing, Xidian University, Xi’an 710071, China)

Abstract — In this paper, both the marginal and joint statistics of second generation Orthogonal bandelet transform (OBT) coefficients of natural images are firstly studied, and the highly non-Gaussian marginal statistics and strong interscale, interlocation and interdirection dependencies among OBT coefficients are found. Then a Hidden Markov tree (HMT) model in OBT domain which can effectively capture all dependencies across scales, locations and directions is developed. The main contribution of this paper is that it exploits the edge direction information of OBT coefficients, and proposes an image denoising algorithm (B-HMT) based on HMT model in OBT domain. We apply B-HMT to denoise natural images which contaminated by additive Gaussian white noise, and experimental results show that B-HMT outperforms the Wavelet HMT (W-HMT) and Contourlet HMT (C-HMT) in terms of visual effect and objective evaluation criteria.

Key words — Image denoising, Bandelets, HMT model, Image modeling, Statistics models, Multiscale geometry analysis.

I. Introduction

In recent years, statistic information provides useful prior for image processing, which comes from the statistic modeling to image. The aim of statistical modeling is to use a small number of parameters to characterize the image, so, a simple, accurate and tractable model is our pursuit. Recently, the combination of statistic model with wavelet and Multiscale geometry analysis (MGA) tools is popular in image processing. Crouse et al. use HMT to model wavelet coefficients and capture both the marginal and joint statistics and reach the state-of-the-art effect in image denoising, classification and detection in Ref.[1], but the lacking of directions (wavelet only has three directions: horizontal, vertical and diagonal) limits its application. To utilize more directions, Duncan and Minh in Ref.[2] use HMT to model contourlet coefficients and effectively capture all dependencies across scales, locations and directions. Although different directions and flexible number of direction can be used in contourlet transform at each scale, they must be set before transform, and it is difficult to choose the appropriate directional number for each scale. So, a transform which has plenty of selective directions and can adaptively choose best direction is our need.

The best candidate is bandelets[13-17]. It is a new MGA tool which can adaptively capture the geometry of regular image along its edges. It has plenty of selective directions for each edge and finds potential in image denoising. Similar to wavelet and MGA tools such as ridgelets, curvelets, contourlets and directionlets, it has the characteristics of multiscale and time-frequency-localized. Different from them, it is adaptive and both its selective direction number and the nonlinear approximation error decay (O(M^-α), α ≥ 2) outperform them. Another adaptive tool wedgelets is beyond of scope as its nonlinear approximation error decay and the number of selective direction inferior to bandelets.

Motivated by W-HMT and C-HMT, we develop a statistics model by using HMT in OBT domain. OBT can adaptively choose optimal direction, and HMT with Gaussian mixture model can capture the dependencies of scales and locations. Combining them and adding local direction information, we propose B-HMT algorithm and use it to denoise natural image which contaminated by additive Gaussian white noise, intuitively, the denoising results should be improved.

The remainder of this paper is organized as follows. Section II defines three important relationships of OBT coefficients. Section III studies the marginal statistics and joint statistics of OBT coefficients. Section IV models OBT coefficient by applying HMT model to OBT domain. Section V gives the denoising process. Section VI analyses the experimental results. Conclusions are drawn in Section VII.

II. Bandelet Coefficient Relationships

Definition

For each OBT coefficient B, we define three important relationships to study its statistics. Parent-children relationship (PB): the coefficient in the same spatial location in the immediately coarser scale as its parent; Neighbor relationship (NB): eight adjacent coefficients in the same subband as its neighbors; colleague relationship...
(CB): coefficients in the same directional subsquare which share the common direction as its colleagues. Wavelet in a subband has only one direction, contourlet in a subband has the pre-designated directions, whereas bandelet segments a subband to many subsquares, and each subsquare has many directions to choose. For example, a subsquare with size of $8 \times 6$ has 72 selective directions, and $16 \times 16$ has 288. Using directional information of each subsquare efficiently, B-HMT should have a better application than W-HMT and C-HMT. The directional information produced in the process of quadtree segmentation is used to exhibit the colleague relationship clearly. Fig.1 shows the three relationships. Fig.1(a) shows that each child has one parent and each parent has four children. Fig.1(c) is part geometry of a subband of natural image “Barbara”, in which the black denotes the considering node, and each different gray subsquare denotes an edge (Inf denotes the smooth area, that is, there is no edge in that subsquare), and values in it denote its direction. Fig.1(d) is its corresponding bandelet coefficients. Assuming one coefficient in a subsquare of OBT is $B$, others in this subsquare are its colleagues, $B$ and its colleagues share one common direction. From above we find that OBT have many direction subsquares and each subsquare has many directions to choose which outperform wavelet and contourlet.

III. Bandelet Statistics

1. Marginal statistics

To investigate the marginal statistics of OBT coefficients of natural images, both the histograms of three finest subbands HH, HL, LH and their $16 \times 16$ subsquares of the image “Barbara” in Fig.2 are exhibited. The distributions of three finest subbands show high non-Gaussian statistics: a sharp peak at zero and heavy tails at both sides. They also prove that OBT has sparsity property: the majority coefficients are close to zero while the minority coefficients contain the main energy. Their kurtosises are $30.0327$, $27.034$ and $68.0335$, respectively. Kurtosis: $(a) 30.0327$; $(b) 27.034$; $(c) 68.0335$; $(d) 4.4308$; $(e) 4.2368$; $(f) 25.376$

2. Joint statistics

Since the marginal statistics only describe the individual behaviors of OBT coefficients and do not consider their dependencies, so, we study their joint statistics by using natural image “Barbara” to exploit their dependencies from three aspects: interscale, interlocation and interdirection. Fig.3 exhibits the conditional distribution of bandelet coefficients conditioned on their parents (PB), neighbors (NB), and colleagues (CB). We find that each conditional distribution reflects one important characteristic of OBT coefficients:

Persistency: the dependencies of parent-children in Fig.3(a) show that large/small values of OBT coefficients tend to propagate through the scales of the quadtree, the children are always large if their parent is large.

Clustering: the dependency between the current coefficient $B$ and its neighbors $NB$ exhibits that the OBT coefficients have the clustering, that is, if the neighbors of a OBT coefficient are large/small, it is always large/small, as Fig.3(b) shows.

Similarity: the dependencies between the current coefficient $B$ and its colleagues $CB$ reveal that the OBT coefficients in a directional subsquare show high similarity. In a non-edge subsquare, coefficients are homogeneous and have the similar magnitude; in an edge subsquare, coefficients that construct the edge always have similar large values, and coefficients that do not construct the edge always have similar small values, as Fig.3(c) shows.

From the joint statistics of OBT coefficients, we conclude that there exist strong interscale, interlocation, and interdirection dependencies among the OBT coefficients. How to model these dependencies is our task in the subsequent sections.

IV. Image Modeling

Our objective is to develop a probability models to capture both the non-Gaussian statistics and complex dependencies of OBT coefficients while easy to process. Conditions mentioned above can be achieved by using HMT model which introduced by Crouse et al. in Ref.[1]. The HMT model is an HMM that uses a quadtree structure. HMM has been successfully used in the wavelet domain and contourlet domain[1-2]. An n-state HMM links each coefficient with a hidden state variable, so each coefficient can be characterized by an n-dimensional state probabilities vector $p$ and an n-dimensional standard deviation vector $\sigma$

$$p = (p_1, p_2, \cdots, p_n)^T \quad \quad (1)$$

$$\sigma = (\sigma_1, \sigma_2, \cdots, \sigma_n)^T \quad \quad (2)$$

where $1, 2, \cdots, n$ denote the states.
We choose HMT to model the statistics of OBT coefficients lies in four aspects. First of all, by establishing links between the hidden state of each OBT coefficient and its four children, the HMT directly models the interscale dependencies of parent-children of OBT coefficients; second, by linking neighbors to a common ancestor, the HMT indirectly model the interlocation dependencies of neighboring OBT coefficients; third, by making a subsquare to share a σ, the HMT also model the interdirection dependencies of colleagues of OBT coefficients, which is different from W-HMT with a subband share a common σ, and this makes our model use more local directional information than W-HMT, and has a better pertinence and should produce a better result, intuitively; last but not the least is that, by using the EM algorithm\[8\], the HMT model in OBT domain can be trained efficiently, and its two-states mixture models imply that two states corresponding to the edges and smooth areas lacks of direction information. In C-HMT model, the directional coefficients in each subband and considering the directional information of each subsquare, the HMT directly share a common state variable \( p_{ij,b}^{(1)} \). Second, by linking neighbors to a common ancestor, the HMT also model the interdirection dependencies of colleagues and this makes our model use more local directional information. In C-HMT model, the directional coefficients in each subband and considering the directional information of each subsquare, the HMT directly share a common state variable \( p_{ij,b}^{(1)} \).

1. Capturing non-Gaussian: Gaussian mixture models

To capture the non-Gaussian statistics of OBT coefficients, the B-HMT models the marginal probability density function (pdf) of each OBT coefficient as a Gaussian mixture density by using a hidden state to designate the large/small of a coefficient. In practice, a two-state, zero-mean Gaussian mixture model is often used to characterize each OBT coefficient. Fig.4 is an example of two-state, zero-mean Gaussian mixture model for a random variable B.

Where \( p_{ij,b}^{(1)}(1) \) and \( 1 - p_{ij,b}^{(1)}(1) \) are the probability mass function of the state variable \( S \), the variances of the Gaussian pdf corresponding to each state.

\[
\begin{align*}
\mathcal{f}_B(b) &= \sum_{m=1,2} p_{ij,b}^{(m)} \mathcal{f}_{B|m}(b|S=m) \\
&= \mathcal{f}_{B|S=1}(b) \cdot p_{ij,b}^{(1)}(1) + \mathcal{f}_{B|S=2}(b) \cdot (1 - p_{ij,b}^{(1)}(1))
\end{align*}
\]

Fig. 4. An example of two-state, zero-mean Gaussian mixture model for a random variable B

The pdf of an OBT coefficient is

\[
\mathcal{f}_B(b) = \sum_{m=1,2} p_{ij,b}^{(m)} \mathcal{f}_{B|m}(b|S=m)
\]

where \( \sum_{m=1,2} p_{ij,b}^{(m)} = 1 \), \( S \) is the hidden state variable which controls the amplitude of OBT coefficient and it is invisible. A coefficient is small when \( m = 1 \), otherwise it is large. We model each OBT coefficient in this way and can fit real OBT coefficients closely.

2. Capturing dependencies: HMT modeling

To capture the dependencies of interscale, interlocation and interdirection among OBT coefficients, the B-HMT uses a probabilistic tree to model Markovian dependencies between the hidden states.

To an OBT decomposition of \( J \) scales, \( K \) subband and \( L \) directional subsquare, a B-HMT model contains the following parameters.

- \( P_{ij,k} \) \((j=1,2,3;k=1,2,3)\): the root state probability vector of \( k \) subband at the coarsest scale.
- \( A_{ij,k} \) \((j=2,\cdots,J;k=1,2,3)\): the state transition probability matrix of the subband \( k \) from scale \( j-1 \) to scale \( j \).
- \( \sigma_{ij,k,l} \) \((j=2,\cdots,J;k=1,2,3;l=1,\cdots,L)\): the Gaussian standard deviation vector of the directional subsquare \( l \) at subband \( k \) of scale \( j \).

The major advantage of HMT in OBT domain is that it uses the direction dependency within the directional subsquare while the W-HMT model does not. In the W-HMT model, each subband can only share one direction of vertical, horizontal and diagonal, so it lacks of direction information. In C-HMT model, the directional number in a subband is designated before the transform, and it is difficult to designate the appropriate directional number for all subbands at the very start. In contrast, B-HMT segments coefficients of a subband to many subsquares and each subsquare has many directions to choose, and it can use the local directional information better than W-HMT and C-HMT, intuitively, and the application effect of B-HMT may outperform W-HMT and C-HMT.

V. Denoising Application

1. Obtain the noisy coefficients: OBT

We apply the OBT to the noisy image and obtain the noisy coefficients. Before the transform, it is necessary for us to standardize the noisy image to a square one, and normalize the image, then set the transform parameters. We decompose the noisy image to the second layer, and set the maximum size of the decomposed subsquare to \( 16 \times 16 \), the minimum ones to \( 4 \times 4 \), the threshold \( T = \sigma \sqrt{2 \log N} \), where \( N \) is the size of image, and calculate the geometry which records the directional information of each subsquare. In OBT domain, the relationship among noisy coefficients \( y \), clean coefficients \( x \) and noise \( e \) with zero-mean additive white Gaussian can be described as

\[
y = x + e
\]

So the denoising problem is to estimate \( x \) given \( y \).

2. Obtain model parameters: HMT modeling and EM training

We use HMT model to match the noisy OBT coefficients and obtain a parameter set \( \theta_y \). We first use a two-state, zero-mean Gaussian mixture model to characterize each OBT coefficient, and use \( y \) to initialize the HMT model and use the variance in each subsquare which represents the directional information of the subsquare to update the initial variance of corresponding location. Then, we use the Markov tree to capture the dependencies of interscale, interlocation and interdirection, and use EM algorithm to train and obtain a model parameter set \( \theta_y \). By subtracting the noise variances from the variances in \( \theta_y \), we obtain a model of the clean coefficients \( \theta_x \):

\[
\sigma_{ij,k,n}^{(e)} = \frac{\text{Median}(|Y_{ij}|)}{0.6745}, \quad Y_{ij} \in HH_1
\]

3. Obtain the clean coefficients: Bayesian estimation

With the HMT model \( \theta_x \), we can estimate \( E[x_{i,j,k,n}|y_{i,j,k,n},\theta_x] \) for each OBT coefficient \( b_{i,j,k,n} \) by using Bayesian estimation. Because of the state \( S_{ij,k,n} \) of each OBT coefficient is given, we can estimate \( E[x_{i,j,k,n}|y_{i,j,k,n},\theta_x] \) by using the following formula which proposed in Ref.[1]

\[
E[x_{i,j,k,n}|y_{i,j,k,n},\theta_x] = \sum_m p(S_{ij,k,n} = m|y_{i,j,k,n},\theta_x) \cdot E[x_{i,j,k,n}|S_{ij,k,n} = m,\theta_x] = \frac{\mathcal{N}_1}{\mathcal{N}_2 + \mathcal{N}_3} y_{i,j,k,n}
\]
image denoising using different algorithms in terms of PSNR

<table>
<thead>
<tr>
<th>Images</th>
<th>σ</th>
<th>Noisy</th>
<th>B-HT</th>
<th>W-HMT</th>
<th>C-HMT</th>
<th>B-HMT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20</td>
<td>23.211</td>
<td>25.984</td>
<td>27.982</td>
<td>27.875</td>
<td>28.169</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>18.227</td>
<td>23.024</td>
<td>23.497</td>
<td>23.349</td>
<td>23.756</td>
</tr>
<tr>
<td>Barbara</td>
<td>10</td>
<td>23.478</td>
<td>25.984</td>
<td>27.982</td>
<td>27.875</td>
<td>28.169</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>18.227</td>
<td>23.024</td>
<td>23.497</td>
<td>23.349</td>
<td>23.756</td>
</tr>
</tbody>
</table>

The denoising effects of B-HMT are also better than W-HMT and C-HMT. This is because that we model all the dependencies of interscale, interlocation and interdirection at the same time. Especially, modeling each directional subsquare can effectively use the local directional information and has a strong pertinence.

The good denoising performance of our method is also exhibited in terms of PSNR. From Table 1, we find that, the PSNR of our method outperforms other methods in most cases. This is because that our method models all the dependencies of parent-children, neighbors and colleagues, so it can capture all dependencies of scales, locations and directions. The local directional information makes B-HMT obtain a better denoising result than W-HMT and C-HMT.

In the lower noise level (σ = 10), PSNR of B-HMT for Lena is lower than that of W-HMT. This is probably because W-HMT is a classical statistical model and reaches the state-of-the-art denoising results, and our goal is to model a HMT in OBT domain, so we only simply use the universal threshold $T = \sigma \sqrt{2 \log N}$ as the initial threshold and do not select other initial parameters appropriately in B-HMT, so PSNR of B-HMT is lower than W-HMT sometimes.

### VII. Conclusion

By using HMT to model OBT coefficients, this paper proposed an image denoising algorithm B-HMT which can make full use of the subsquare direction information of OBT and capture the all dependencies across scales, locations and directions. The denoising results of B-HMT for additive Gaussian white noise outperform that of the W-HMT and C-HMT in terms of visual effect and objective evaluation criteria. To find the best initial threshold for geometry calculation and optimize the initial parameters in order to obtain a better visual effect and a higher performance index are the further
research.

References


ZHANG Wenge was born in Shaanxi Province, China in 1969. He received the B.S. degree from the Electronic Engineering Institute of PLA, Hefei, China in 1990 and the M.S. degree in computer network from the Chool of Computer Science and Technology of Xidian University, Xi’an, China, in 2006. Currently, he is a Ph.D. candidate at Xidian University. His research interests include the representation and application of multiscale geometry analysis, image denoising and SAR image processing.

WANG Shuang was born in Shaanxi Province, China in 1978. She received the B.S. degree in electrical engineering in 2000, the M.S. degree in computer applications and technology and Ph.D. degree in circuit and system from Xi-dian University in 2003 and 2007, respectively. Currently, she is an associate professor at Xidian University. Her research interests are multiscale geometry analysis, image processing and high-resolution SAR image processing. (Email: shwang@mail.xidian.edu.cn)

LIU Fang was born in Shaanxi Province, China in 1978. She received the B.S. degree in electrical engineering in 2000, the M.S. degree in computer applications and technology and Ph.D. degree in circuit and system from Xi-dian University in 2003 and 2007, respectively. Currently, she is an associate professor at Xidian University. Her research interests include machine learning, optimization problems, image processing and high-resolution SAR image processing. (Email: shwang@mail.xidian.edu.cn)

ZHANG Wenge was born in Shaanxi Province, China in 1969. He received the B.S. degree from the Electronic Engineering Institute of PLA, Hefei, China in 1990 and the M.S. degree in computer network from the Chool of Computer Science and Technology of Xidian University, Xi’an, China, in 2006. Currently, he is a Ph.D. candidate at Xidian University. His research interests include the representation and application of multiscale geometry analysis, image denoising and SAR image processing.

GAO Xinbo was born in Shandong Province, China in 1972. He received the B.S., M.S. and Ph.D. degrees in signal and information processing from Xidian University, China, in 1994, 1997 and 1999 respectively. Currently, he is a professor at Xidian University. His research interests include machine learning and multimedia information processing and analysis.

JIAO Licheng was born in Shaanxi Province, China in 1959. He received the B.S. degree from Shanghai Jiaotong University, Shanghai, China in 1982 and the M.S. and Ph.D. degrees from Xi’an Jiaotong University, Xi’an, China in 1984 and 1990, respectively. His current research interests include signal and image processing, nonlinear circuit and systems theory, learning theory and algorithms, optimization problems, wavelet theory, and data mining.