

自相似数据流的 Hurst 指数小波求解法分析¹

李永利 刘贵忠 王海军 尚赵伟

(西安交通大学电子信息工程学院信通系 西安 710049)

摘 要 该文针对具有自相似性的数据流的特征参数——Hurst 指数的小波提取方法进行了系统的分析和研究,对现有提取 Hurst 指数的小波方法给予了归纳和总结,通过大量的实验和分析,研究了消失矩阶数、小波函数以及分解级数的选择对小波法提取 Hurst 指数的准确性的影响,并得出了一些有意义的结论。

关键词 自相似性, 数据流, 消失矩, Hurst 指数, 小波

中图分类号 TN919.3

1 引 言

在网络和多媒体日益普及的今天,大量的信息都是通过互连网来传输的,对这一类数据的统计分析一直倍受研究学者的关注。1994 年 Leland 等对以太网上的流量统计发现这一类数据具有统计自相似性^[1]。1995 年 Beran 等通过对大量的不同类别的变比特率视频数据的统计分析发现它们都表现出长相关性以及分形的特点^[2]。这一特征不仅使得传统的模型已经无法准确地描述网络流量和变比特率视频数据的特征,同时这一特征对网络的性能,如丢包率、网络拥塞等有着关键性的影响。使得基于这一特征下的相关研究引起了众多学者的广泛兴趣。目前对其确切地描述和建模问题仍旧没有一个令人满意的答案。

自相似数据简单地讲就是具有尺度不变和自相关函数缓慢衰减的特点。Hurst 指数作为表征这一类数据统计特性的一个重要指标,它反映了数据的自相似程度及其二阶统计特性。并且它也是被用来作为网络拥塞控制和接入允许控制的重要指数。对于 Hurst 指数的提取,前人已经提出了许多方法,如 R/S 法、Variance-time 法、以及 Periodogram 法等^[3-5]。小波的出现给这一问题带来了许多新的内容和方法。这是因为利用小波对信号分析时是在尺度和时间域上进行的,尤其它的多尺度特性与自相似过程的尺度不变性有着自然的联系。所以对自相似过程的特征分析中,小波成为了必然的选择。同样用小波求解 Hurst 指数的方法有多种,然而估计的准确性和有效性是我们对任何一种方法取舍的依据,也是对数据流特征把握和对网络性能分析的关键所在。本文将在分析利用小波求解 Hurst 指数的基础上,对这一类方法给予了归纳。通过大量的实验对它的准确性和鲁棒性进行了分析,研究了小波消失矩、小波函数的选择对这一类算法的影响,并给出了由此而得到的一些有意义的结论。为准确地提取 Hurst 指数提供了保证,为后续自相似数据建模等工作提供了可靠的依据。

本文第 2 节介绍了统计自相似过程的有关性质和特点,第 3 节阐述了小波法求解 Hurst 指数的原理,第 4 节在归纳和总结了现有的小波方法基础上针对消失矩阶数和小波的选择问题进行了大量的实验和分析。在文章的最后一节给出了分析结果和结论。

2 统计自相似过程

如果一个随机过程 $x(t)$ 满足 $x(t) = a^{-H}x(at)$, 其中 $=$ 表示在统计意义上的相等, a 为一正实数,那么 $x(t)$ 是一个统计自相似过程,其 Hurst 指数为 H ^[6]。另外在文献 [7] 中也给出了几个等价的定义。

¹ 2001-08-14 收到, 2002-02-15 改回

国家教育部博士点基金 (No.2000069828), 国家教育部骨干教师基金 (2000 年度) 和西安交通大学重点培植项目支持

设 $x(t)$ 为统计自相似过程, 它的方差为 σ_x^2 , 则它具有如下性质^[6,7]: (1) 能量谱 $S_x(\omega)$ 与 $|\omega|^{-\gamma}$ 成正比, ω 为频率, $\gamma = 2H + 1$; (2) 当 $0 < \gamma < 1$, 自相关函数 $R_x(\tau)$ 与 $|\tau|^{\gamma-1}$ 成正比, τ 为时间间隔. 根据统计特征对 H 的求解不外乎利用这些性质, 借助 R/S 法或小波等其它方法来估计 H 的值.

3 小波法求解 Hurst 指数

利用小波求解 Hurst 指数的方法有多种^[3-5,8,9], 我们将其分类如下, 并分析其原理和方法.

3.1 小波系数方差法

设有一自相似过程 $x(t)$, $\varphi(t)$ 为一正交小波函数, $\varphi_{jk}(t) = 2^{-j/2}\varphi(2^{-j}t - k)$, $\Psi(\omega)$ 为其傅里叶变换, d_{jk} 为小波变换后的系数: $d_{jk} = \langle x(t), \varphi_{jk}(t) \rangle = 2^{-j/2} \int x(t)\varphi(2^{-j}t - k)dt$. 如果满足 $0 < \gamma < 2R$, 其中 R 为小波函数的原点矩阶数, 则经过小波变换后小波系数间存在如下关系^[6]:

$$E[d_{jk}d_{j'k'}] = \frac{2^{(j+j')/2}}{2\pi} \int_{-\infty}^{+\infty} \frac{\sigma_x^2}{|\omega|^\gamma} \Psi(2^j\omega)\Psi^*(2^{j'}\omega)e^{-i(2^jk-2^{j'}k')\omega}d\omega$$

并且随 $(k - k')$ 的增加, 同一级小波系数之间的相关性减小. 当 $j = j'$, $k = k'$ 时,

$$E[d_{jk}^2] = \text{Var}[d_{jk}] = \frac{2^{j\gamma}}{2\pi} \int_{-\infty}^{+\infty} \frac{\sigma_x^2}{|\omega|^\gamma} |\Psi(\omega)|^2 d\omega = \sigma^2 2^{j\gamma} \quad (1)$$

其中 $\sigma^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\sigma_x^2}{|\omega|^\gamma} |\Psi(\omega)|^2 d\omega$ 为一与 j 无关的常量. 通过对 (1) 式两边的数据取对数后, 在均方误差最小的意义下进行线性拟合得到以 j 为自变量, 以 $\log_2 \text{Var}[d_{jk}]$ 为函数的直线, 斜率即为 γ . 利用 $\gamma = 2H + 1$ 便可以求得 H 的值.

3.2 谱估计法

对于一自相似随机过程 $x(t)$, 它与具有 R 阶原点矩, $R > \gamma/2$ 的正交小波作用后, 在 $2^{-j}\omega_0$ (ω_0 与小波函数的选取有关) 处的能量谱为^[3]

$$S_x(2^{-j}\omega_0) = \frac{1}{n_j} \sum_k |d_{jk}|^2 \quad (2)$$

n_j 为第 j 个子带中的小波系数的个数. 根据自相似过程的性质: $S_x(2^{-j}\omega_0) \propto (2^{-j}\omega_0)^{-\gamma}$, 所以得到

$$\log_2(S_x(2^{-j}\omega_0)) = \log_2\left(\frac{1}{n_j} \sum_k |d_{jk}|^2\right) = \log_2(c(2^{-j}\omega_0)^{-\gamma}) = (2H + 1)j + \hat{c} \quad (3)$$

其中 c , \hat{c} 均为常数, 与小波系数方差法一样采用均方最小意义下的线性拟合得到关于 j 和 $\log_2(S_x(2^{-j}\omega_0))$ 的一条直线, 斜率即为 $(2H + 1)$. 由此我们就可以计算 H .

3.3 能量法

它是利用不同子带之间的能量关系来求取 H 的^[8]. 以 Γ_j 代表第 j 子带的平均能量, $\Gamma_j = \frac{1}{n_j} \sum_k |d_{jk}|^2$.

$$\Gamma_j \propto \int \frac{|\varphi_j(\omega)|^2}{\omega^\gamma} d\omega = 2^{j\gamma} \int \frac{|\varphi_0(\omega)|^2}{\omega^\gamma} d\omega = 2^{j\gamma} \Gamma_0 \quad (4)$$

$\varphi_j(\omega)$ 为第 j 级子带小波系数的傅里叶变换。因此可得

$$\log_2 \Gamma_j = \log_2 \left(\frac{1}{n_j} \sum_k |d_{jk}|^2 \right) = j\gamma + c \quad (5)$$

由 (5) 式可见, 它与 (1), (3) 式得到的结果是一致的。只是它们从不同角度入手来得到估计结果的。以上 3 种方法其实都是基于能量的方法去分析自相似信号的。从最终的结果看是一致的。它们是用子带的平均能量谱与子带分解级数的关系进行估计的, 无论是系数方差法中的 $E[d_{jk}^2]$, 还是谱估计法中的 $S_x(2^{-j}\omega_0) = \frac{1}{n_j} \sum_k |d_{jk}|^2$, 或者是能量法中的 $\Gamma_j = \frac{1}{n_j} \sum_k |d_{jk}|^2$, 三者都是平均能量的表达式, 这就是它们共同的本质。许多文献中分别对它们做了研究^[3-5,8,9], 但是缺乏对其进行比较、归纳和分析。这里要指出的是这 3 种方法从本质上讲是一致的、统一的。

4 实验结果和分析

如上所述, 许多前人的工作只是在方法上的分析, 而没有对不同方法进行比较和归纳, 这里我们将从分解级数、消失矩以及小波函数的阶数等方面入手, 对小波方法求解 Hurst 指数的准确性和可行性进行分析和研究。

首先考虑小波分解级数对提取特征指数的影响。由于对 Hurst 指数的提取是在统计意义上的归纳, 所以分解级数越多越具有代表性和广泛性。但是由于数据的有限长使得分解级数不可能取无限多级, 因此下面以最后一级小波分解后所得系数的个数不少于所采用的小波滤波器的长度为约束条件来确定小波分解级数。

消失矩是影响小波分解后系数之间相关性的重要因素。它直接关系到小波方法对 Hurst 特征参数估计的效果。下面通过实验的方法来研究它对特征参数估计的影响。我们选择 Daubechies 小波作为分解小波, 通过改变其消失矩来观察其对特征参数估计的影响。实验中所用的测试序列为 Rose^[10] 所提供的具有自相似性的变比特率视频序列—— Asterix, 详细编码参数可参阅文献 [10]。由于数据长度为 40000 帧, 我们选用分解级数为 11 级。根据统计经验和统计分析的结果^[1,2,8,10] 可知, 一般视频序列的 H 指数值在 0.5~1 之间, 因此在满足 $R > \gamma/2 = (2H + 1)/2$ 的条件下, 我们取 $R \geq 3$ 。

图 1 是利用具有不同消失矩的 Daubechies 小波求解 Hurst 指数的结果, 其中具有较大偏差的低频子带系数(前三级)的结果在拟合的过程中被舍去。从图中可以看出 db3 的拟合结果最好, 误差较小。随着消失矩的增加拟合的误差加大, 这里的误差是指偏差绝对值的平方和。由图 1(d) 的结果看, 此时原始数据的非线性已经很明显了。所以当满足 $R > \gamma/2 = (2H + 1)/2$ 的条件下, 增加消失矩的阶数并不能带来更大的精确度。并且随着阶数的增加, 所求得的 Hurst 指数也逐渐减小, 并呈现出一定的周期性, 如图 2(a) 所示。用 Daubechies 小波求得 Hurst 指数的实验结果列于表 1 中。用 Matlab 工具箱中的 Symlets(sym) 小波进行同样的实验, 结果列于表 2 中, 得到与图 1 相似的结论, 随着消失矩的增加, 求 Hurst 指数过程中的拟合误差呈现周期性增加的趋势, 并且 H 逐渐减小, 同样也呈现出周期性, 如图 2(b) 所示。

另外由以上的实验结果和分析可知, 在消失矩满足约束条件下, 消失矩阶数越小, 拟合误差越小, 所得的指数越精确。在后续分析不同小波对求取 Hurst 指数的影响中将采用 Matlab 工具箱中的 db3、sym3 以及 Coiflets 小波 coif1 和 coif2, 他们对应的消失矩分别为: 3, 3, 2, 4。

图 3 分别为利用上述 4 种小波求取特征参数的曲线和相应的特征参数值。由于无法获得具有三阶消失矩的 Coiflets 小波, 所以用 coif1 和 coif2 来作分析。由前面的实验结果可知随消失矩的增加, H 逐渐减小。这里取与 coif1 和 coif2 相对应的 H 的平均作为具有三阶消失矩的 Coiflets 小波对 H 的估计, $H=0.9902$ 。它与 db3 和 sym3 的实验结果很接近。这说明在满足消失矩约束的条件下, 具有相同消失矩阶数的不同小波具有几乎相同的结果。从表 1 和表 2 也可以得到相似的结论。

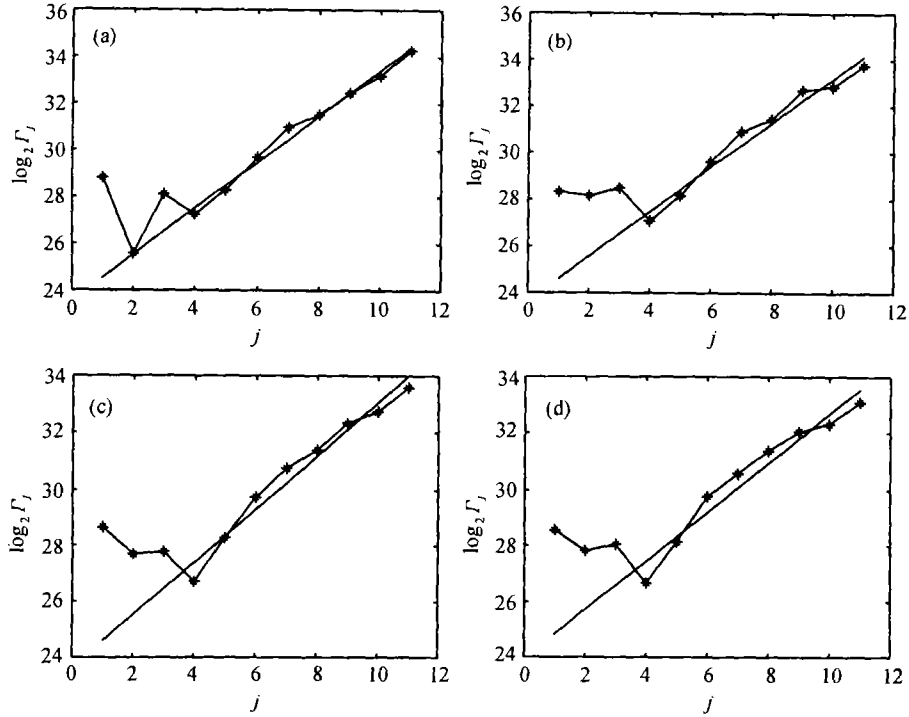


图 1 利用具有不同消失矩的 Daubechies 小波求解 Hurst 指数
 (a) db3, $H=0.9882$, (b) db6, $H=0.9754$, (c) db11, $H=0.9681$, (d) db20, $H=0.9353$

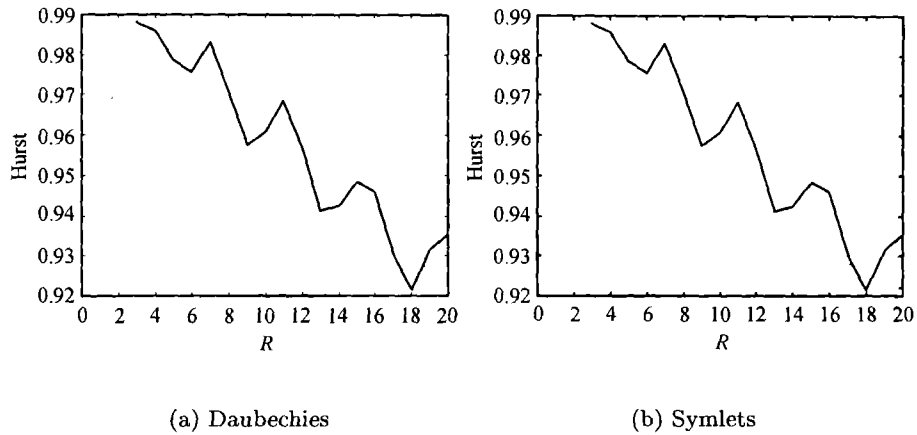


图 2 Hurst 参数随消失矩的变化曲线

表 1 不同消失矩的 Daubechies(db) 小波求解 Hurst 指数的结果

| 小波 | db3 | db4 | db5 | db6 | db7 | db8 | db9 | db10 | db11 |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| H | 0.9882 | 0.9860 | 0.9788 | 0.9754 | 0.9832 | 0.9707 | 0.9574 | 0.9606 | 0.9681 |
| 误差 | 0.4921 | 0.7506 | 1.2735 | 1.1004 | 0.9363 | 1.3188 | 1.9472 | 1.6735 | 1.3107 |
| 小波 | db12 | db13 | db14 | db15 | db16 | db17 | db18 | db19 | db20 |
| H | 0.9568 | 0.9412 | 0.9424 | 0.9484 | 0.9459 | 0.9300 | 0.9216 | 0.9316 | 0.9353 |
| 误差 | 1.4930 | 2.2540 | 2.2688 | 1.4893 | 1.6628 | 2.3871 | 2.5105 | 1.9333 | 1.8506 |

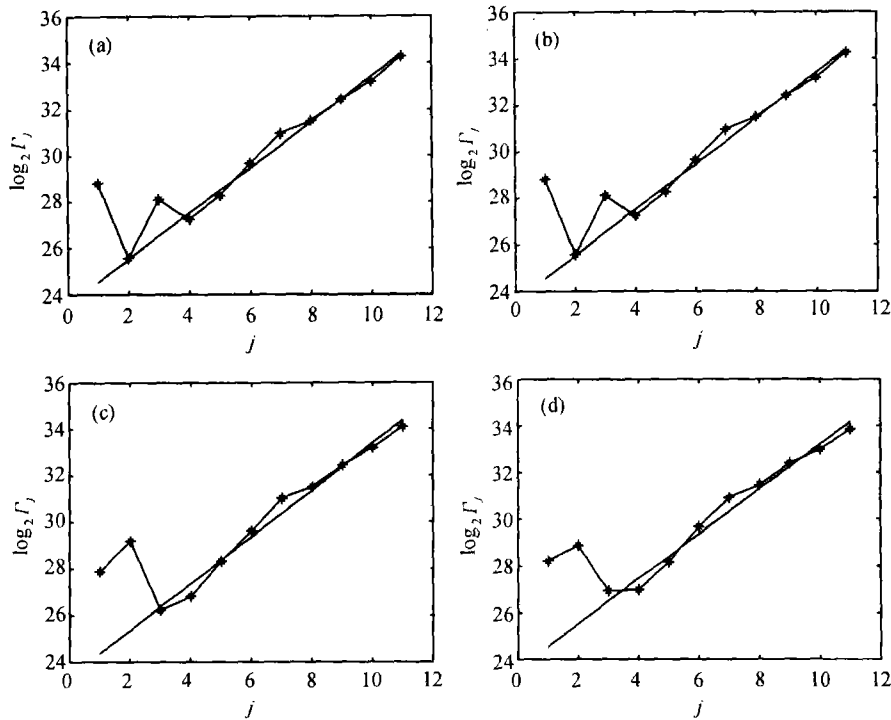


图3 不同小波求取 Hurst 参数的结果比较

(a) db3, $H=0.9882$, (b) Sym3, $H=0.9882$, (c) Coif1, $H=1.0008$, (d) Coif2, $H=0.9706$

表2 不同消失矩的 Symlets(sym) 小波求解 Hurst 指数的结果

| | | | | | | | | | |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|---------|
| 小波 | sym3 | sym4 | sym5 | sym6 | sym7 | sym8 | sym9 | sym10 | sym11 |
| H | 0.9882 | 0.9925 | 0.9904 | 0.9769 | 0.9661 | 0.9736 | 0.9738 | 0.9650 | 0.95051 |
| 误差 | 0.4921 | 0.8808 | 1.0413 | 0.8616 | 1.3414 | 1.0681 | 1.5291 | 1.1450 | 2.1559 |
| 小波 | sym12 | sym13 | sym14 | sym15 | sym16 | sym17 | sym18 | sym19 | sym20 |
| H | 0.9603 | 0.9425 | 0.9544 | 0.9333 | 0.9469 | 0.9326 | 0.9345 | 0.9203 | 0.9339 |
| 误差 | 1.3388 | 2.2563 | 1.4457 | 2.4125 | 1.5608 | 2.2655 | 1.8743 | 2.4175 | 1.7823 |

5 结 论

由于现实中的许多现象和数据都表现出自相似的特征, 为准确把握这些特性和对其进一步的分析研究, 提取自相似特征参数 H 是一个重要而不可缺少的环节。本文通过大量的实验和分析, 得到以下很有意义的结论:

(1) 现有的利用小波求取 H 指数的方法——方差法、谱估计法、能量法, 三者殊途同归。其本质是相同的、一致的。

(2) 在满足消失矩约束的条件下, 增加消失矩阶数并不能提高所得拟合结果的精度, 反而有所下降。同时随消失矩阶数的增加, H 呈现下降趋势, 并具有一定的周期性。

(3) 在满足消失矩约束的同时, 利用具有相同消失矩阶数的正交小波求解所得的 H 值相差很小, 可以认为是一致的。

通过以上的实验分析可见, 利用现有小波方法求解 Hurst 指数时, 在满足正交性、消失矩约束 $R > \gamma/2 = (2H + 1)/2$ 的条件下, 消失矩阶数越小所得特征参数的值越精确。

参 考 文 献

- [1] W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson, On the self-similar nature of Ethernet traffic (extended version), *IEEE/ACM Trans. on Networking*, 1994, 2(1), 1-15.
- [2] J. Beran, R. Sherman, M. S. Taqqu, W. Willinger, Long range dependence in variable bit rate video traffic, *IEEE Trans. on Communication*, 1995, 43(2/3/4), 1566-1579.
- [3] P. Abry, D. Veitch, Wavelet analysis of long-range dependent traffic, *IEEE Trans. on Information Theory*, 1998, 44(1), 2-15.
- [4] Z. Sahinoglu, S. Tekinay, Multiresolution decomposition and burstiness analysis of traffic traces, *Wireless Communications and Networking Conference, WCNC. New Orleans, LA, USA, IEEE, Vol.2, 1999, 560-563.*
- [5] P. Abry, P. Goncalvés, P. Flandrin, Wavelet-based spectral analysis of $1/f$ process, *IEEE International Conference on Acoustics, Speech, and Signal Processing, Minneapolis, MN, USA, Vol.3, 1993, 237-240.*
- [6] G. Wornell, *Signal Processing with Fractal: A Wavelet Based Approach*, Prentice Hall, Znc. NJ, 1995, 30-57.
- [7] B. Tsybakov, N. D. Georganas, On self-similar traffic in ATM queues: Definitions, overflow probability bound, and cell delay distribution, *IEEE/ACM Trans. on Networking*, 1997, 5(3), 397-409.
- [8] S. Giordano, S. Miduri, M. Pagano, F. Russo, S. Tartarelli, A wavelet-based approach to the estimation of the Hurst parameter for self-similar data, *International Conference on Digital Signal Processing, DSP 97, Santorini, Greece, Vol.2, 1997, 479-482.*
- [9] Z. Fan, P. Mars, Self-similar traffic generation and parameter estimation using wavelet transform, *IEEE Global Telecommunications Conference, Phoenix, AZ, USA, Vol.3, 1997, 1419-1423.*
- [10] Oliver Rose, Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems. *Proc. of the 20th Annual Conference on Local Computer Networks, Minneapolis, MN, 1995, 397-406.*

ON WAVELET-BASED METHODS FOR HURST INDEX
ESTIMATION OF SELF-SIMILAR TRAFFIC

Li Yongli Liu Guizhong Wang Haijun Shang Zhaowei

(School of Electronic & Information Eng., Xi'an Jiaotong Univ., Xi'an 710049, China)

Abstract Existing wavelet methods for the estimation of the Hurst parameter of self-similar traffic are systematically analyzed and examined. The effects of wavelet functions, vanishing moments and wavelet decomposition levels on the results of wavelet methods for acquiring the Hurst index are investigated via numerical experiments. Some useful conclusions are drawn on the relationship between the accuracy of the methods and the selection of the order of vanishing moments and the selection of wavelet functions.

Key words Self-similarity, Traffic data, Vanishing moments, Hurst index, Wavelet

李永利: 男, 1975 年生, 博士生, 主要研究兴趣包括图像处理、视频压缩与传输。

刘贵忠: 男, 1962 年生, 教授, 博士生导师, 1989 年 6 月荷兰 Eindhoven 大学博士毕业, 主要从事的研究工作为通信, 多媒体, 探测, 测控及预测中的信号与信息处理; 非平稳信号分析与自适应信号处理。

王海军: 男, 1971 年生, 博士生, 主要从事信号处理、模式识别等方面的研究。

尚赵伟: 男, 1968 年生, 讲师, 博士生, 主要研究兴趣包括数据库、图像处理、视频处理等。