

# 极低码率下运动图象的压缩

陈康力 王绍霖  
(同济大学电气系, 上海 200092)

**摘要** 首先根据当前运动图象压缩的国际标准, 讨论了极低码率下的编码问题, 然后介绍了 Matching Pursuit 在该领域中的应用, 其具体过程是利用二维 Gabor 函数字典对帧间预测差值进行 Matching Pursuit<sup>[1]</sup>, 这样能获得低码率下更高的压缩比和更好的效果, 而且由于函数字典不是基于块的, 因此消除了低码率下 DCT 变换的方块效应. 此外, 在进行帧间预测差值编码时, 是从能量最集中的区域开始, 逐渐分配比特, 从而使得解码后的图象能体现运动变化显著的区域.

**关键词** H. 261 MPEG-I H. 263 Matching Pursuit 低码率

中图法分类号: TN919. 81 文献标识码: A 文章编号: 1006-8961(2000) 07-0597-06

## The Application of Matching Pursuit on Video Compression at Very Low Bit Rate

CHENG Kang-li, WANG Sao-ling

(Department of Electrical Engineering, Tongji University, Shanghai 200092)

**Abstract** After reviewing some international recommendations for video compression, the application of Matching Pursuit on video compression at very low bit rate is introduced. This algorithm decomposes the signal into linear expansions of waveforms using an overcomplete basis of Gabor functions. With adapting Matching Pursuit to the motion residual, a substantial improvement in both PSNR and perceived visual quality can be gotten. The dictionary which is not block-based can reduce block distortions compared with the low bit rate DCT systems. When coding motion residual, bits are assigned progressively to high energy areas in the motion residual. Thus the decoded image can show the areas where motion changes quickly. Another advantage is the flexible choice of dictionary.

**Keywords** H. 261, MPEG-I, H. 263, Matching Pursuit, Very low bit rate

## 0 引言

国际上一般把低于 64kbit/s 的码率称为低码率. 现有的普通电话交换网 GSTN 上的调制解调器的速率通常为 33. 6kbit/s, 而数字无线通信的信道传输速率仅仅为几 kbit/s, 它们都可认为是极低码率.

由表 1 可以看到, 在 10kbit/s 时, 运动图象压缩比要求很高. 由于原有的 MPEG-1、MPEG-2、H. 261 等国际标准并没有针对极低码率的视频编码, 所以它们所采用的传统 DCT 混合编码则暴露了其不足之处. DCT 能量集中特性仅次于最佳的 K-L 变换, 目

表 1 码率为 10kbps 时各图象格式下压缩率比较

帧率(帧/s)	CCIR601 (720× 576)	CIF (352× 288)	QCIF (176× 144)
7. 5	4 979: 1	915: 1	229: 1
10	6 637: 1	1 216: 1	304: 1
15	9 952: 1	1 824: 1	456: 1

前对它的研究较深入, 它具有简单实用、算法快速等优点. 但由于分块 DCT 把图象分为 8× 8 的子块进行变换, 且各块独立量化, 因此在要求低码率、高压缩比时, 其量化粗糙, 且有明显的方块效应, 并且噪声容易像飞蚊一样, 从目标越过互不相关的块, 迅速向解码的图象蔓延, 而引起“蚊式噪声”.

虽然模型编码有相当高的压缩比,其压缩比甚至达到  $10^4$  或  $10^5$  倍.但由于其中物体基模型必须用每帧 2kbit 左右的码字来进行与模型一致区域的编码,并用每帧 4kbit 左右的码字来进行模型失误区域<sup>[2]</sup>的编码,故其码率很难达到 20kbit/s;而语义基模型又必须事先要知道模型的基本框架,因此它现在还只应用于头肩像的编码.另外,小波变换用于低码率下的运动图象编码现有很多人正在研究.

最近的 H. 263、MPEG-4 编码标准就是适应于低码率的编码标准.其中, H. 263 采用了双线性内插的方法来实现半像素精度的运动估计.它具有非限制运动矢量方式、基于句法的算术编码、高级预测方式、PB 帧方式等 4 种协商附加的编码方式,从而进一步提高了编码效率.但它依旧保留了传统进行帧间编码和帧间预测编码的 DCT 变换方式,因此在极低码率时仍有块的痕迹.据称, MPEG-4 将公布其正式文献,有望对低码率编码有更大的促进.

本文介绍的基于 Matching Pursuit 的运动图象编码,是当前提出的一种比较新的方法,也是一个可供选择的极低码率下图象编码方法.本文实验过程中,帧内编码和 H. 263 相同,但对帧间差值用 Matching Pursuit 处理. Mallat 和 Zhang 首先提出把 Matching Pursuit 用于信号分析<sup>[3]</sup>,即信号通过函数字典进行多阶段的扩张,而每一阶段,都要进行大量的计算,以便从字典中选取合适的函数来表示信号,再从信号中减去相应的能量,以便进行下一阶段的匹配过程.由于 Matching Pursuit 使用很大的函数字典,并且不是基于块的,因此能有效地消除低码率时 DCT 变换产生的块边界的失真.

### 1 Matching Pursuit 的基本原理<sup>[1]</sup>

Matching Pursuit 利用一个冗余的 Gabor 函数字典把信号分解成相应函数的线性组合.它首先定义一个一维的高斯函数

$$g(t) = \sqrt[4]{2} e^{-\pi^2 t^2}$$

Gabor 函数集合定义为

$$g_Y(t) = \frac{1}{\sqrt{s}} g\left|g \frac{t-u}{s}\right| e^{j\zeta t}$$

$Y = (s, u, \zeta) \in R^+ \times R \times R$ , 式中,  $R$  是实数集合;  $R^+$  是正数集合;  $s$  是正尺度因子;  $u$  是时间偏移;  $\zeta$  是调制频率. Matching Pursuit 的过程就是信号  $f(t)$  在函数集合中寻找合适的  $g_Y(t)$ . 其计算过程即求内积

$$p = \langle f(t), g_Y(t) \rangle$$

通过选择  $Y$ , 使获得最大  $p$  值, 然后继续计算

$$R(t) = f(t) - p g_Y(t)$$

最终  $f(t)$  可表示为

$$f(t) \approx \sum_{n=1}^M p_n g_{Y_n}(t)$$

可以证明, 当  $M$  增大时, 能保证上式收敛于  $f(t)$ .

为了能把 Matching Pursuit 用于帧间差值图象, 定义长度为  $N$  的离散一维 Gabor 函数为

$$g_\alpha(i) = K_\alpha \cdot g\left|\frac{i - \frac{N}{2} + 1}{s}\right| \cdot \cos\left|\frac{2\pi\left|i - \frac{N}{2} + 1\right|}{N} + \phi\right|$$

$$i \in \{0, 1, \dots, N - 1\};$$

$\alpha = (s, \zeta, \phi)$ , 它包含了尺度因子  $s$ 、调制频率  $\zeta$ 、相位偏移  $\phi$  而常量  $K_\alpha$  是序列函数的归一化系数. 然后利用上述一维离散函数再构造一个二维离散函数

$$G_{\alpha, \beta}(i, j) = g_\alpha(i) g_\beta(j) \quad i, j \in \{0, 1, \dots, N - 1\}$$

再通过选择一系列的  $\alpha, \beta$  就构成了二维 Gabor 函数字典. 并且这样一个可分离的二维离散函数有助于快速算法的实现<sup>[4]</sup>. 由于字典大小直接影响 Matching Pursuit 的计算量, 因此本文选取了 15 组  $\alpha, \beta$ , 即函数字典大小为 15,  $N = 16$ . 各个参数的选择是根据文献中多次实验的结果<sup>[1]</sup>, 但如何优化是值得研究的课题. 表 2 给出了 Gabor 函数字典大小为 15 时,  $\alpha$  的各参数值.

表 2  $\alpha_m = (s_m, \zeta_m, \phi_m), m \in \{1, 2, \dots, 15\}$

$m$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$s_m$	1	3	5	7	9	100	1	5	100	4	4	8	4	4	4
$\zeta_m$	0	0	0	0	0	0	1	1	1	2	3	3	4	2	4
$\phi_m$	0	0	0	0	0	0	$\pi/2$	$\pi/2$	$\pi/2$	0	0	0	0	$\pi/4$	$\pi/4$

图 1 是根据表 2 中各个参数, 计算  $G_{\alpha, \beta}(i, j) = g_\alpha(i) g_\beta(j)$  后的函数波形结构示意图. 图中任一

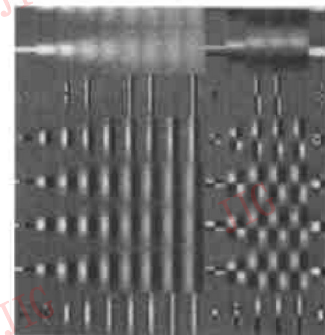


图 1

分都可以用一组索引值 $(h, v)$ 来表示,  $h$  和  $v$  分别表示 Gabor 函数字典结构的水平分量和纵向分量.

## 2 基于 Matching Pursuit 的运动图象压缩

运动图象压缩的具体过程就是通过 Matching Pursuit 对帧间预测误差进行多阶段分解的过程<sup>[1]</sup>, 即将二维字典中的每一个基本函数与误差图象的所有象素进行比较, 当每次获得最佳匹配后, 则从误差图象中减去相应的能量, 再进行下一次匹配的过程. 其能量由 5 个参数确定, 即字典结构的索引 $(h, v)$ , 最佳匹配的位置 $(x, y)$ , 内积值 $p$ . 该 5 个参数构成一个“原子”.

因此, 多阶段的匹配过程也可以描述为寻找“原子”的过程. 对于一帧  $176 \times 144$  象素的 QCIFT 图象, 若是每一个象素点位置都进行内积计算, 则计算量很大. 为了减少计算量, 其改进措施是先对差值图象进行分块 $(8 \times 8)$ , 然后寻找能量最大的块<sup>[1]</sup>; 再以这一块的中心为中心, 以  $S \times S$  ( $S = 16$ ) 为窗口, 在

其中每一象素点位置上进行匹配计算. 这样, 每一帧原子个数的多少就是码率控制的问题. 图 2 是 Miss America 的第 5 帧和第 6 帧的差值图象, 图 3、图 4、图 5 是寻找原子的结果. 从图中的效果来看, 上述的寻找过程, 既保证了原子从差值图象能量集中的地方逐次分布, 并且又避免了原差值图象所具有的干扰噪声. 当需要码率降低时, 重建图象的细节是逐步丢失的. 从图 6、图 7 和表 3 可见, 当码率从 15kbit/s 降为 10kbit/s 时, 原子数从大约 45 个减少到 25 个, 则 Miss America 的眼睛和脸部就相对变得毛糙和不清楚了.

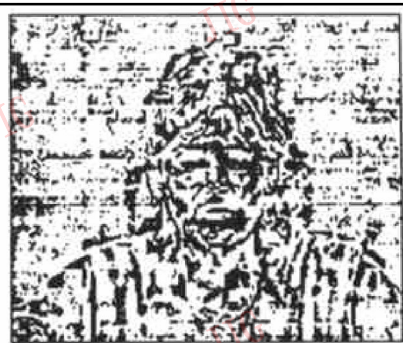


图 2 差值图象

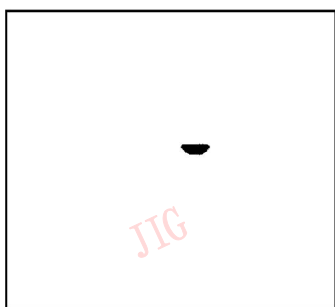


图 3 一个原子

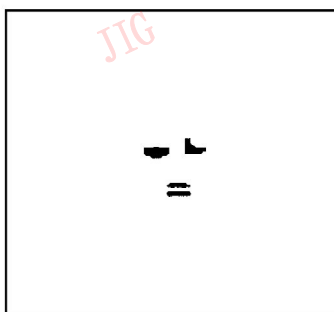


图 4 5 个原子



图 5 45 个原子



图 6 Miss A 重建图象  
15kbit/s 码率下



图 7 Miss A 重建图象  
10kbit/s 码率下

表3 Matching Pursuit 实验结果(Miss America)

码率 (kbit/s)	每帧原子 个数	原子所需 比特数	(h, v)所占 比例(%)	p 所占 比例(%)	(x, y)所占 比例(%)	平均PSNR (dB)
15	44	19.20	39.5	18.6	41.9	38.05
10	25	19.57	37.9	18.9	43.2	37.02

然后就是对原子5个参数的编码. 其中  $p$  是最佳匹配时的最大内积值, 可用固定的线性量化器进行量化, 再进行 Huffman 编码. 函数字典结构索引  $(h, v)$  可以通过对序列图象的统计, 利用固定 Huffman 码表进行编码. 而对位置  $(x, y)$  进行编码时, 则把每帧原子从左到右, 从上到下顺序排列, 其中一组码字用于指示每行第一个原子的横坐标, 另一组码字则用于指示同一行相邻原子之间的水平距离, 在这组码字中, 尚包含一个结束标志的码字, 来指出下一个有效行之间的纵向距离.

实验中, 码率控制方法较为简单, 即首先把目标比特率除以目标帧率, 得到每帧应有的比特数  $F_k$ , 再减去每帧的头部信息和运动矢量编码所需的比特数  $m_k$ , 剩余的就是用于预测误差编码的比特数; 然后, 把预测误差的比特数除以上一帧每个原子平均编码的比特数  $a_{k-1}$ , 就得到当前需要寻找的目标原子数  $N_k$ . 即

$$N_k = \frac{F_k - m_k}{a_{k-1}}$$

实验证明, 通过对这些原子进行编码, 能较好地和目标码率相匹配.

### 3 快速算法的实现<sup>[4]</sup>

综上所述, 二维离散 Gabor 函数有助于快速算法的实现, 这可类似于 FFT、小波变换的快速算法, 从而使 Matching Pursuit 具有实用性. 一般, 对于  $S \times S$  窗口计算, 若函数字典大小为  $B \times B$  个一维基本函数, 长度分别为  $N$ , 而每一次内积又是在  $N \times N$  的像素区域上进行, 因此整个窗口需要计算  $S^2 B^2 N^2$  次. 若  $S = 16, B = 15, N = 16$ , 则获得一个原子需要计算 14 745 600 次. 若利用二维离散 Gabor 函数, 来寻找原子, 由于  $G_{\alpha, \beta}(i, j)$  是可分离的, 因此

$$\sum_{i=1}^N \sum_{j=1}^N G_{\alpha, \beta}(i, j) \cdot I(i, j) = \sum_{i=1}^N g_{\alpha}(i) \sum_{j=1}^N g_{\beta}(j) \cdot I(i, j)$$

其中,  $I(i, j)$  是位置  $(i, j)$  处的像素值. 参照图 8, 可

以看到, 若先用某一个  $g_{\beta}$  进行垂直的一维内积计算, 把结果存在数组中, 接着用所有  $g_{\alpha}$  的循环进行水平的一维内积计算, 然后再换一个  $g_{\beta}$  继续计算. 其计算量是<sup>[5]</sup>:

$$BSN(S + N) + B^2 S^2 N$$

那么, 这样寻找一个原子要计算 1 044 480 次, 即比前者减少了 14 倍左右. 而 Matching Pursuit 的计算量要明显多于 DCT 变换, 但可用专用芯片实现. 具体应用时, 需预先定义一个二维数组, 以存储  $G_{\alpha, \beta}$  的各个数值, 这样在内积时可避免重复计算高斯函数的值. 并且在一维 Gabor 函数中, 当尺度因子较小时, 函数值近似为零, 计算时就可不必再作这些乘积.

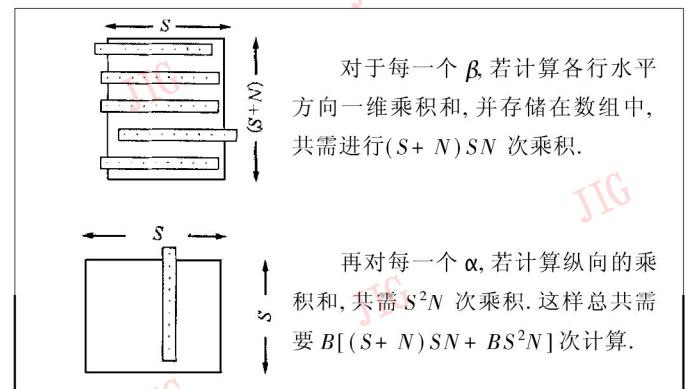


图8 快速算法

## 4 实验结果

本实验在挪威 Telenor 的 H. 263 仿真软件 tmn 基础上, 对帧间预测差值改用 Matching Pursuit 进行处理, 图象格式为 QCIF. 选用标准测试序列图象 Miss America 和 Claire, 其码率分别为 10kbit/s 和 15kbit/s, 帧率为 15 帧/s. 为了能和 H. 263 在相同条件下比较, 预测误差也是采用 H. 263 的双线性内插方法获得.

信噪比 PSNR 定义为

$$PSNR = 10 \log_{10} \left| \frac{255^2}{MSE} \right|$$

MSE 是重建图象的均方误差.

由表 3 和表 4 可以看到, 码率在 10kbit/s 和 15kbit/s 时, H. 263 每一帧传送的非零 DCT 系数个

表4 H. 263 实验结果(Miss America)

码率 (kbit/s)	每帧 DCT 系数	DCT 系数所需 比特数	平均 PSNR(dB)
15	70.65	11.04	37.48
10	29.79	15.37	36.05

数要比 Matching Pursuit 传送的原子个数多, 但 Matching Pursuit 的每个原子需要 20bit 左右, 大于 DCT 系数所需的比特数, 因此用 H. 263 时, DCT 方法与 Matching Pursuit 方法相比较, 其对帧间预测误差编码所需比特数接近. 表 4 还提供了原子各个参数所占的比例, 其中原子的坐标位置编码需要更多比特. 当码率为 10kbit/s 时, 平均信噪比要高于 H. 263 约 1dB.

图 9 是 Miss America 第 70 帧和 Claire 第 96 帧

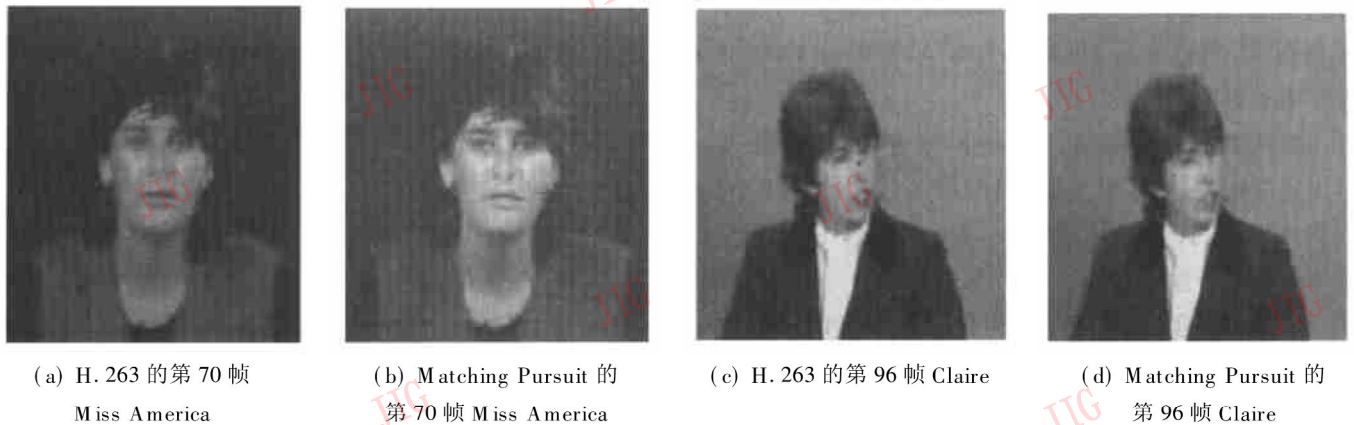


图 9 H. 263 和 Matching Pursuit 编码效果比较

图 10、图 11、图 12 是 Miss America 和 Claire 的信噪比图, 图上实线是 Matching Pursuit 的结果, 虚线是 H. 263 的结果. 每种情况下 Matching Pursuit 的信噪比基本上均高于 H. 263, 且起伏的趋势相象. Miss America 和 Claire 都是典型的会议电视序列图象, Miss America 只有头部和肩部的缓慢平移运动, Claire 则有非平移的头部转动. 在后面的八、九十帧, 由于两者都有大幅度的运动, 因此信噪比迅速下降, 但 Matching Pursuit 的信噪比依然高于 H. 263. 当码率为 10kbit/s 时, Miss America 在个别帧处 H. 263 与 Matching Pursuit 两者的信噪比很接近; 在 15kbit/s 时, 这种情况较明显, 平均信噪比差 0.57dB. 当码率很低时, 由于 Matching Pursuit 的原子总是首先集中在预测差值图象的高

帧的 H. 263 与 Matching Pursuit 两者主观观察效果比较. H. 263 是用 DCT 对差值图象进行变换, 因此可以看到图 9(a) 在嘴唇、眼睛、脸部周围有块的痕迹, 相对于 Matching Pursuit 的图 9(b) 而言, 细节也不够多. Claire96 帧处的头部转动, 不属于平移. 同样 Matching Pursuit 的图 9(d) 比 H. 263 的图 9(c) 有更多的细节和较少的块效应, 而且整体上更显得清楚.

能量处, 即出现在很需要它们的地方, 而基于 DCT 变换的系统的比特则扩散到了差值图象的大部分区域, 没能体现出主要的运动变化, 此时 Matching Pursuit 的编码率较高.

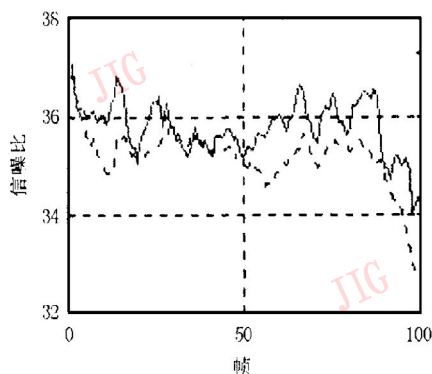


图 10 Claire 图象信噪比曲线(10kbit/s 码率时)

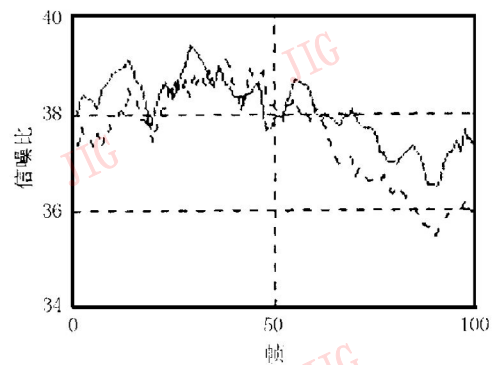


图 11 Miss America 信噪比曲线(10kbit/s 码率时)

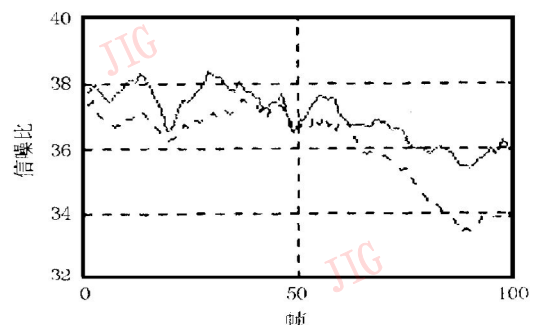


图 12 Miss America 信噪比曲线(15kbit/s 码率时)

## 5 结 论

当要求极低码率时, DCT 变换在图象中移动对象的边缘有高频噪声, 而方块效应出现在平滑变换的地方. Matching Pursuit 因为函数字典不是基于块的, 所以没有块效应, 而且它的原子总是首先出现在有显著变化的地方, 因此可随时根据低码率的要求, 来进行原子多少的选择. 会议电视和电视电话等序列图象以头肩像为主, 由实验结果看到, 选用这种序列图象, 当码率为 10kbit/s 时 Matching Pursuit 要较明显好于 H. 263.

但随着码率的提高, Matching Pursuit 不能显示出它的优势, 由于原子数目的增多, 因而匹配时的内积计算量就会变得相当大, 从而降低了它在较高码率下的可行性<sup>[5]</sup>.

Matching Pursuit 可作为极低码率下的一种选择, 其中, 搜索窗口  $S$  的大小, 字典中一维函数的个数  $B$ , 函数长度  $N$  都直接影响它的计算量. 实验中  $S=16$  是个折衷. 而文献[6]采用了不同函数及长度, 这是进一步优化函数字典, 减少计算量, 提高效率的良好方法. 并且文献[1]中采用平滑运动模式和 Matching Pursuit 相结合, 能更好地消除块效应. 本文受条件限制, 没能用硬件来实现算法.

### 参 考 文 献

- 1 Ralph Neff, Avidesh Zekhor, Martin Vetterli. Very low bit rate video coding using Matching Pursuits. In: Proceeding of SPIE Conference on Visual Communication and Image Processing, Chicago, 1994, 2308(1): 47~ 60.

- 2 Musmann H G. A unified point of view on video compression: A model-based perspective. In: Int. Workshop, Very Low Bit Rate Video Compression. Urbana, IL, May 1, 1993.
- 3 Mallat S, Zhang Z. Matching pursuit with time-frequency dictionaries. IEEE Transactions in Signal Processing, 1993, 41 (12): 3397~ 3415.
- 4 Neff R. A very low bit rate video coding system using matching pursuits [Master's Thesis], University of California at Berkeley, August 1994.
- 5 Ralph Neff, Avidesh Zekhor. Matching pursuit video coding at very low bit rates. In: IEEE Data Compression Conference, Snowbird Utah, 1995.
- 6 Ralph Neff, Eugene Miloslavsky. Experiment T3: matching pursuit coding of prediction errors current status and results. In: MPEG Meeting in Maceio Brazil, 1996.



**陈康力** 1992年毕业于上海同济大学电气工程系, 1999年获信号与信息处理专业硕士学位. 现任同济大学电子信息工程学院讲师, 上海图形图象学会会员. 目前从事基于小波、分形图象压缩编码及网络多媒体技术相关课题的研究工作.



**王绍霖** 1955年毕业于上海交通大学电讯工程系. 现为同济大学电气工程系教授, 上海图形图象学会理事, 中国电子学会图象通信委员会委员. 研究兴趣包括信号处理、图象处理、神经网络、目标识别等.

## 通 告

为适应我国信息化建设需要, 扩大作者学术交流渠道, 本刊已加入《中国学术期刊(光盘版)》和“中国期刊网”. 作者著作权使用费与本刊稿酬一次性付给. 如作者不同意将文章编入该数据库, 请在来稿时声明, 本刊将做适当处理.