

# 基于稀疏表示的低比特率可伸缩图像编码算法研究

孙玉宝<sup>1</sup> 肖 亮<sup>1</sup> 韦志辉<sup>1</sup> 胡晰远<sup>2</sup>

(<sup>1</sup> 南京理工大学计算机科学与技术学院模式识别与智能系统实验室, 江苏 南京 210094)

<sup>2</sup> 中科院自动化所 国家专用集成电路设计工程研究中心, 北京 100190

**摘要** 根据图像的几何结构特性,从人类视觉系统特性出发,建立了 Gabor 感知多成份字典,进而模拟人类视觉通路的层次处理机制,构建了稀疏编码网络,能够有效去除图像中的高阶冗余,形成更为稀疏的表示。对稀疏表示系数重组后进行比特平面量化,实现了低比特率的可伸缩编码。实验结果表明,在低比特率下,本文算法压缩后重构图像的感知质量要明显优于 JPEG2000,峰值信噪比也与其相当,并且对于图像中的边缘和纹理等细节保持效果更佳。

**关键词** 稀疏表示; 图像压缩; 感知字典; 可伸缩

中图分类号 TP391 文献标识码 A doi: 10.3788/AOS200828s2.0077

## Low-Rate and Scalable Image Coding with Sparse Representations

Sun Yubao<sup>1</sup> Xiao Liang<sup>1</sup> Wei Zhihui<sup>1</sup> Hu Xiyuan<sup>2</sup>

<sup>1</sup> *Laboratory of Pattern Recognition and Artificial Intelligence, Institute of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China*

<sup>2</sup> *National ASIC Design Engineering Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*

**Abstract** Based on geometric properties of the local image structures and the perception characters of human visual system (HVS), a geometrically motivated Gabor multi-component perception dictionary is constructed. Moreover, inspired from the hierarchical information processing in the human visual path, a sparse coding network is proposed to reduce high order redundancy, thus provides much sparser representations of images. Furthermore, coefficients of repositioned atoms in sparse approximation are quantized by bit-plane quantization. It presents an effective low bit-rate and scalable image compression algorithm. Our simulation results show that, under the low bit-rate, our approach is comparable to the JPEG2000 in terms of PSNR value, while effectively preserves edges and textures structures and exhibits generally a better visual quality.

**Key words** sparse representation; image coding; perception dictionary; scalability

## 1 引 言

从“基函数”的角度看, JPEG 是基于离散余弦变换(DCT 基)的稀疏表示<sup>[1]</sup>, 而 JPEG2000 标准是基于小波基的稀疏表示<sup>[2]</sup>, 自然的问题是下一代标准是基于什么的稀疏表示。相当数量研究表明: 小波变换能够在一定程度上对图像去相关, 但由于小波基的各向同性以及有限的方向分辨率, 幅值较大

的系数仍在边缘、纹理处积聚出现, 并不能对线状奇异性形成稀疏表示<sup>[3]</sup>。同时图像是多种奇异性的混合物, 未必能用单一类型的正交基实现稀疏表示, 因此基于冗余字典(Redundant dictionary)的图像稀疏表示理论是一个新的有效突破, 通过从冗余字典中寻求最佳  $M$  (远小于信号维数) 项线性组合来表示图像, 实现了对图像更为稀疏地逼近与分解, 从而

**基金项目:** 国家 863 计划(2007AA12E100)、国家自然科学基金(60672074)和教育部高校博士点专项科研基金(20070288050)资助课题。

**作者简介:** 孙玉宝(1983—), 男, 博士研究生, 主要从事图像建模与稀疏表示和图像编码等方面的研究。

E-mail: syb8692833@126.com

**导师简介:** 韦志辉(1963—), 男, 教授, 博士生导师, 主要从事基于变分偏微分方程理论及图像处理应用、基于视觉感知的稀疏编码理论及应用等方面的研究。E-mail: gswei@mail.njust.edu.cn

成为新型图像压缩理论的极佳候选者。当前关于过完备稀疏分解的可伸缩编码,特别对低码率下的可伸缩编码逐渐成为人们的研究热点<sup>[4,5]</sup>。

神经科学研究表明人类通过视觉通路中不同功能与结构的神经元的层次处理,在本原视觉皮层实现了对输入图像的稀疏编码<sup>[6,7]</sup>。因此模拟视觉感知的层次处理机制,本文首先对图像进行小波变换,并将各个子带的小波系数划分为维数较低的子块。同时以本原视觉皮层中简单细胞的感受野函数(脉冲响应函数)Gabor 函数作为本文字典的生成函数<sup>[8]</sup>,建立了匹配不同图像结构的 Gabor 感知多成份字典,再应用匹配追踪算法对各个子块在该字典下进行稀疏分解<sup>[9,10]</sup>,从而形成了一个层次的稀疏编码网络,能够有效去除图像中高阶冗余性,形成了更为稀疏的表示。本文对匹配追踪稀疏分解系数重组后进一步进行比特平面量化,再经上下文自适应的算术编码实现了高性能的低比特率可伸缩编码。根据不同带宽需求,输出码流能够在任意点截断,具有比特率可伸缩特性。

## 2 图像的层次稀疏表示

经过长期的发展与进化,人类视觉系统能够自适应于自然图像的统计特性,依次经视觉通路中视网膜、侧膝体与本原视觉皮层中多种具有不同感受野神经元的层次与并行处理,通过侧抑制与局部竞争机制最终在本原视觉皮层中实现了对自然图像的稀疏表示。

本文将模拟视觉通路的层次处理机制来实现对自然图像的稀疏表示。刻画视网膜与侧膝体中神经元的同心圆型感受野的 DOG 函数具有类似小波函数的特性<sup>[11]</sup>,因此本文使用小波函数来建模同心圆型感受野。同时采用本原视觉皮层中简单细胞的感受野函数 Gabor 函数作为生成函数来构造本文的

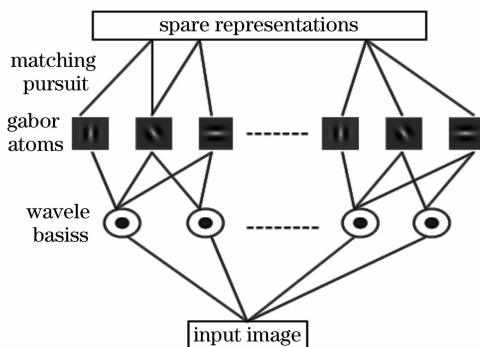


图1 层次的稀疏编码网络

Fig. 1 A hierarchical sparse coding network

Gabor 感知多成份字典. 匹配追踪是一个贪婪的迭代算法,每一次迭代都选择一个与图像结构最匹配的原子,模拟了视觉皮层的局部竞争与侧抑制机制。基于此,首先对图像进行小波变换,以突出图像中的边缘结构,并将各个子带的小波系数划分为维数较低的子块,再应用匹配追踪算法对各个子块在 Gabor 感知多成份字典下进行稀疏分解,从而形成了一个层次的稀疏编码网络(见图 1),能够有效去除图像中高阶冗余性,形成了更为稀疏的表示。

### 2.1 Gabor 感知多成份字典

作为原子的产生函数,二维 Gabor 函数具有七个自由参数,其模型如下所示

$$g_{\lambda, \theta, \varphi, \gamma, x_0, y_0} = K \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \varphi\right), \quad (1)$$

$$x' = (x - x_0) \cos \theta + (y - y_0) \sin \theta, \quad (2)$$

$$y' = -(x - x_0) \sin \theta + (y - y_0) \cos \theta, \quad (3)$$

式中  $K$  为归一化常数,使原子具有单位范数。参数  $(x_0, y_0)$  确定原子的中心位置;标准差决定了原子支撑区域的大小,变换参数可以使原子匹配图像中不同尺度的结构。参数  $\gamma$  确定了原子支撑区域的纵横比。角度参数  $\theta (\theta \in [0, \pi])$  指示了原子中局部几何结构的方向。变换参数  $\theta$  可以使原子匹配图像中不同方向的结构。参数  $\varphi (\varphi \in (-\pi, \pi])$  为余弦调和因子  $\cos\left(2\pi \frac{x'}{\lambda} + \varphi\right)$  的偏移相位,确定了原子  $g$  的对称性。

最后,参数  $\lambda$  为波形调制因子  $\cos\left(2\pi \frac{x'}{\lambda} + \varphi\right)$  的波长,相应地  $1/\lambda$  为空间频率,是原子的最优响应频率。比率  $\sigma/\lambda$  确定了原子的空间频率带宽,因而也确定了原子支撑域中平行相间的兴奋与抑制条纹区域个数。原子的半幅空间频率带宽  $b$  (Half-amplitude bandwidth) 与比率  $\sigma/\lambda$  具有如下关系

$$b = \text{lb} \frac{(\sigma/\lambda)\pi + \sqrt{\ln 2/2}}{(\sigma/\lambda)\pi - \sqrt{\ln 2/2}}, \quad (4)$$

$$\frac{\sigma}{\lambda} = \frac{1}{\pi} \sqrt{\frac{\ln 2}{2}} \cdot \frac{2^b + 1}{2^b - 1},$$

可见,  $\sigma/\lambda$  比率越大,原子支撑区域中包含的条纹个数越多,相应地空间频率带宽越小。Gabor 原子的方向带宽由原子的空间频率带宽和纵横比因子确定,三者存在如下关系

$$\Delta\theta_{1/2} = \arcsin\left(\gamma \frac{2^b - 1}{2^b + 1}\right), \quad (5)$$

给定空间频率带宽  $b$ , 参数  $\sigma$  与  $\lambda$  是相互依赖的,其中

只有一个为自由参数。本文选取尺度参数  $\sigma$  为自由参数,  $\lambda$  的值由参数  $b$  与参数  $\sigma$  确定。

可通过对生成函数中各自由参数进行平移、伸缩与旋转等几何变换以生成一系列原子来构成字典  $D = \{\varphi_\gamma, \gamma \in \Gamma\}$ , 其中  $\gamma$  为原子索引,  $\Gamma$  为原子索引集合。含有不同条纹个数的 Gabor 原子能够最优匹配图像中不同结构成份。当 Gabor 函数退化为高斯函数, 表现出尺度函数特性, 能够有效地逼近图像中的平滑成份, 组成平滑成份字典  $D_{\text{smooth}}$ 。当支撑区域包含条纹个数较少(本文选取为小于等于 3) 时, 原子能够最优匹配图像中的边缘轮廓成份, 构成边缘成份字典  $D_{\text{edge}}$ 。当原子支撑区域包含条纹个数较多(大于等于 4) 时, 原子表现为强烈的震荡样式, 最优匹配图像中震荡的纹理结构, 对应纹理成份字典  $D_{\text{texture}}$ 。

## 2.2 离散 Gabor 感知多成份字典

在实际应用中, 需建立离散的 Gabor 多成份字典。本文将依据本原视觉皮层中简单细胞的组织方式、响应特性与图像的几何结构特性来约束自由参数的离散采样方式, 使得离散 Gabor 多成份字典在减少原子个数的同时能够保证字典的稀疏表示能力。同时对连续参数的离散采样间隔应足够小, 确保字典中所有原子能够成为整个  $L^2(R^2)$  空间的框架。

下面给出本文离散 Gabor 感知多成份字典各自由参数离散值。均匀一致采样方向参数  $\theta$  ( $\theta \in [0, \pi)$ ), 其采样间隔  $\Delta\theta$  正比于原子的方向带宽。沿着指数序列  $\{a^j\}_{j \in \mathbb{Z}}$  采样尺度参数  $\sigma$ , 伸缩步长充分小, 一般取为  $a = 2^{1/\nu}$ , 其中  $\nu$  是每一对数尺度中的中间尺度个数。对于不同空间频率带宽的原子,  $a$  将取不同的值。相位  $\varphi$  离散为  $0, 0.5\pi$ 。取平移间隔为 1 个像素, 使本文的多成份字典应具有平移不变性。为了自适应于图像中不同类型的局部几何结构, 不同子成份字典对自由参数  $b, \sigma, \gamma$  将采取不同的离散方法。假定待分解图像为  $N \times N$ 。

1) 针对平滑成份字典  $D_{\text{smooth}}$ , Gabor 函数退化为高斯函数。采取各向同性的尺度关系  $\gamma = 1$ , 取  $\nu = 1, a = 2$ , 按指数序列  $\{2^j\}_{j \in \mathbb{Z}}$  均匀采样尺度  $\sigma, j \in [0, [\lg N] - 1]$  限制最大尺度为原图像的一半。

2) 针对边缘成份字典  $D_{\text{edge}}$ , 取空间频率带宽  $b$  为 2.0, 1.5 倍倍频程, 纵横比因子  $\gamma$  分别取值为 0.7, 0.5, 0.3, 使原子为各向异性。取  $\nu = 1/b, a = 2^b$  按指数序列  $\{2^b\}_{j \in \mathbb{Z}}$  均匀采样尺度  $\sigma, bj \in [0, [\lg N] - 3]$  限制原子的最大尺度为图像的 1/8。相位取为正交相位对  $0, \pi/2$ 。当  $b = 1.5, \gamma = 0.7, 0.5,$

0.3 时分别将参数离散为 6、8、12 个方向。当  $b = 2.0, \gamma = 0.7, 0.5, 0.3$  时分别将参数  $\theta$  离散为 4, 6, 12 个方向。

3) 针对纹理成份字典  $D_{\text{texture}}$ , 分别取带空间频率带宽为  $b = 1.0, 0.7, 0.5$ , 只取  $r = 1$ , 原子为各向同性。取  $\nu = 1/b, a = 2^b$  按指数序列  $\{2^b\}_{j \in \mathbb{Z}}$  均匀采样尺度  $\sigma, bj \in [0, [\lg N] - 3]$  限制原子的最大尺度为图像的 1/8。当  $b = 1.0, 0.7, 0.5$ , 分别将参数  $\theta$  离散为 6、8、12 个方向。为了提高原子匹配纹理的准确性, 将参数  $\theta$  统一离散为 12 个方向。

## 2.3 层次的稀疏编码网络

2.2 节从图像的几何结构特性与人类视觉感知特性出发, 设计了结构自适应的 Gabor 感知多成份字典, 缩减了字典中原子个数, 有利于提高稀疏分解效率。进一步, 对图像进行小波变换, 并将每个子带划分为维数较低的子块, 再应用匹配追踪算法对各个子块在 Gabor 感知多成份字典下进行稀疏分解, 从而形成了一个层次的稀疏编码网络。不仅降低了图像的维数同时降低了搜索字典的复杂度, 并适合并行实现。层次的稀疏编码算法具体步骤如下:

- 1) 构建离散 Gabor 感知多成份字典;
- 2) 对给定图像  $I$  进行小波变换与块分解, 记  $B_k$  为分解后的第  $k$  个子块,  $K$  为总块数, 块号集合  $\Lambda = \{1, \dots, K\}$ , 变换系数  $f = \{B_k, k = 1, \dots, K\}$ ;
- 3) 初始化: 令初始残差信号为原始块信号  $R_0^k = B_k, k = 1, \dots, K$ , 每一子块的初始迭代次数  $i = 0$ , 并计算所有原始块信号与字典  $D$  中所有原子的内积:  $\alpha_\gamma^k = \langle R_0^k f, \phi_{\gamma \in \Gamma} \rangle$ ;
- 4) 内积极大化搜索:  $\gamma_i^k = \operatorname{argmax}_{\gamma \in \Gamma, k \in \Lambda} |\alpha_\gamma^k|$ ;
- 5) 残差信号更新:  $R_{i+1}^k f = R_i^k f - \alpha_{\gamma_i^k} \phi_{\gamma_i^k}$ ;
- 6) 内积系数更新:  $\alpha_\gamma^k = \langle R_{i+1}^k f, \phi_{\gamma \in \Gamma} \rangle$ ;
- 7) 迭代停止规则: 如果残差信号能量小于给定的值  $\|R_i^k f\|^2 \leq \xi_{\text{stop}}$ , 转至步骤 8; 否则, 针对相应子块  $k_i, i = i + 1$ , 转至步骤 4);
- 8)  $\Lambda = \Lambda - \{k_i\}$ , 如何集合  $\Lambda$  为空, 则算法结束; 否则转至步骤 4)。

经过  $M$  次迭代后, 图像  $I$  的小波变换系数  $f$  的  $M$  项稀疏逼近为:  $f_M = \sum_{i=0}^{M-1} \alpha_{\gamma_i^i} \phi_{\gamma_i^i}$ , 相应地稀疏逼近误差为  $R^M f = f - f_M$ 。  $M$  趋近无穷时, 在有限维信号空间  $\|R^M f\|$  按指数级收敛  $\lim_{N \rightarrow +\infty} \|R^M f\| = 0$ 。匹配追踪算法的运算复杂度主要由字典中原子个数决定, 可通过使用一个快速算法而实现, 该算法利用一个简单的更新公式与局部字典搜索策略大幅度降低

了运算复杂度,文献[9,10]中给出了较为详细的说明。

### 3 稀疏表示图像编码器

图2给出了本文基于稀疏表示的编码器(简记为

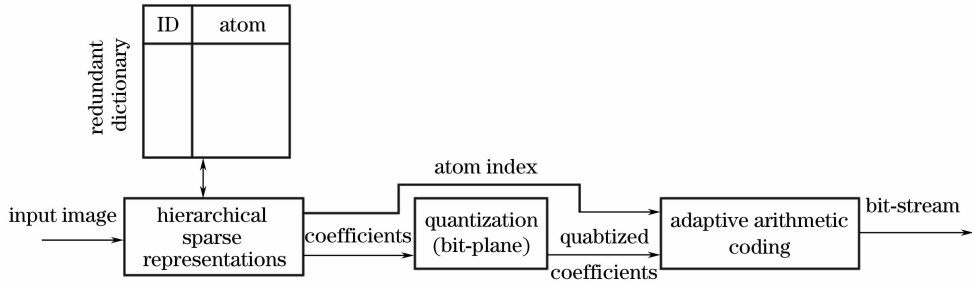


图2 稀疏表示图像编码器结构框图

Fig. 2 Block diagram of the Sparse Representations image coder

具体的量化算法如下所示:

Step1 初始化

1.1) 按块的顺序重新排列稀疏编码网络输出的数据(原子下标与系数,记为PR),块内顺序可以任意;

1.2) 设置初始门限  $T = 2^{\lceil \lg(\text{MaxCoeffs}) \rceil}$ ;

1.3) 建立LSP表并置为空(LSP用来存放已经扫描过并且大于门限的系数及位置和原子号,便于后续的精扫扫描)。

Step2 判断当前PR系数中是否有大于门限T的?

如果有,输出1,进入Step3;否则输出0,转入Step4。

Step3 按顺序对每个子块扫描

3.1) 判断该子块中是否有系数大于T? 如果有输出1,进入3.2);否则输出0,转入3.3)。

3.2) 扫描该子块

i) 编码该子块中大于门限的系数位置,系数的原子号与原子系数符号位;

ii) 将系数的信息(原子号,系数值,位置及块号)移入LSP中;

iii) 将PR中该系数的值置为0。

3.3) 判断是否每个子块都被扫描过? 如果是,进入Step4;否则,子块号加1,转入3.1)。

Step4 判断门限是否大于指定的门限?

如果是,设置门限  $T = T/2$ ,进入精细扫描过程,否则,结束编码过程。

Step5 精细扫描

对于LSP中的每个表项,若不是在刚刚进行的

SPC)结构框图,输入图像首先经稀疏编码网络递归分解后,获取其M项稀疏逼近,为了产生嵌入式码流,采用比特平面量化器量化稀疏逼近中每一个原子的系数.量化后的原子系数与原子索引输入到上下文自适应的算术编码器以产生最终的比特流<sup>[12]</sup>。

扫描过程中新添加的,则输出每个原子系数的二进制表示中第n个最重要的位。精细扫描过程结束后,再转入Step2;

步骤3.2)的i)中,原子号是直接编码的;而对于系数的位置,经zig-zag扫描后,采用游程加算术编码,这样0、1出现的概率分布很不均匀,有利于熵编码。SPC的解码过程是编码的逆过程,编码时编码多少次,解码时也可以解码多少次。解码过程的主要步骤包括:接收编码器发送的解码信息、设置阈值、构造逆量化器、解读编码器输出位流中包含的原子的参数信息与系数信息。随着解码过程的不断深入,重构图像的质量不断得到提高。由于匹配追踪算法每一次迭代都能使对f的逼近更加优化,同时本文采用比特平面量化器可累进逼近原子系数,从而能够产生嵌入式的比特流,具有内在的比特率可伸缩性,根据不同带宽需求,输出码流能够在任意点截断,抛弃后面不重要的原子,从而实现比特率的可伸缩。

### 4 实验结果与分析

为了验证本文图像编码算法的有效性,以512 pixel×512 pixel标准测试灰度图像barbara和baboon为例进行编解码实验,并与JPEG2000编码标准(以下简称JPEG2000)进行了比较。表1列出了不同压缩比下两种算法的重构图像的峰值信噪比(PSNR)值,从中可以看出,在低比特率情况下本文算法的PSNR要高于JPEG2000,这是由于本文的Gabor感知多成份字典具有更强的稀疏逼近能力。图3给出了两种方法对图像Baboon压缩40倍后的

图像的局部对比图。可以发现,采用 JPEG2000 压缩后图像中 baboon 的毛发与胡须部分的纹理都已经模糊成球状,但是本文方法编码的图像对于该部分的纹理结构依然保持得较好,具有更优的视觉感知质量。总体而言,在低比特率情形下本文的

PSNR 值要高于 JPEG2000,而在较高比特率情形下 JPEG2000 编码器的 PSNR 值要高于本文的方法。同时从实验结果图中可以发现,本文的方法对于保持图像的纹理和细节信息方面要优于 JPEG2000 编码器,具有更优的感知质量。

表 1 本文算法 SPC 与 JPEG200 在不同压缩率下 PSNR 的比较

Table 1 PSNR comparison of SPC and JPEG2000 at different compression ratio

Image	Compressionratio Encoder	32:1	40:1	64:1	80:1	100:1	160:1
		Barbara	JPEG2000	27.36	26.36	24.59	23.97
	SPC	27.38	26.96	25.63	25.12	24.14	23.06
Baboon	JPEG2000	22.81	22.13	21.32	21.01	20.70	20.24
	SPC	23.18	22.69	21.83	21.25	20.96	20.43

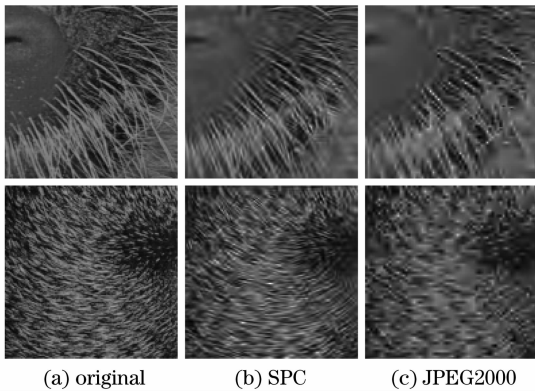


图 3 编码器 SPC 与 JPEG2000 在比特率 0.2 下对 512 pixel×512 pixel Baboon 压缩后的细节图像

Fig. 3 Detail view, 512 pixel×512 pixel Baboon encoded at 0.2 bpp by SPC and JPEG2000

参 考 文 献

1 Gregory K W. The JPEG still picture compression standard[J]. *IEEE Trans. on Consumer Electronics*, 1992, **38**(1): 18~34  
 2 A. Skodras, C. Christopoulos, T. Ebrahimi. The JPEG 2000 still image compression standard[J]. *IEEE Signal Processing*, 2001, **18**(5): 36~58

3 Crouse M S, Nowak R D, Baraniuk R G. Wavelet-based statistical signal processing using hidden Markov model [J]. *IEEE Trans. Signal Process*, 1998, **46**(4): 886~902  
 4 R. Figueras, I Ventura, P. Vandergheynst *et al.*. Low-rate and flexible image coding with redundant representations[J]. *IEEE Trans. Image Processing*, 2006, **15**(3): 726~739  
 5 L. Peotta, L. Granai, P. Vandergheynst. Image compression using an edge adapted redundant dictionary and wavelets[J]. *Signal Processing*, 2006, **86**(3): 444~456  
 6 Olshausen B A, Field D J. Sparse coding with an overcomplete basis set: a strategy employed by V1? [J]. *Visual Research*, 1997, **37**(33): 11~25  
 7 Vinje W E, Gallant J L. Sparse coding and decorrelation in primary visual cortex during natural vision [J]. *Science*, 2000, **287**(5456): 1273~1276  
 8 Daugman J G. Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters[J]. *J. Opt. Soc. Am. A*, 1985, **2**(7): 1160~1169  
 9 S. Mallat, Z. Zhang. Matching pursuits with time-frequency dictionaries [J]. *IEEE Trans. on Signal Processing*, 1993, **41**(12): 3397~3415  
 10 Bergeau F, Malt S. Match pursuit of images[C]. *Proceedings of IEEE-SP*, Philadelphia, PA, USA, 1994. 330~333  
 11 Rodieck R W, Stone J J. Analysis of receptive fields of cat retina ganglion cells[J]. *J. Neurophysiology*, 1965, **28**(14): 833~849  
 12 I. H. Witten, R. M. Neal, J. G. Cleary. Arithmetic coding for data compression[J]. *Commun. ACM*, 1987, **30**(6): 520~540