

# 基于 Gabor 感知多成份字典的图像稀疏表示算法研究

孙玉宝<sup>1</sup> 肖亮<sup>1</sup> 韦志辉<sup>2</sup> 邵文泽<sup>1</sup>

**摘要** 如何设计合适的能够匹配各层面几何结构的图像稀疏表示过完备字典, 进而形成对图像的稀疏分解是当前研究者关注的热点问题. 根据图像的几何结构特性, 从人类视觉系统特性出发, 建立了匹配各层面图像结构的 Gabor 感知多成份字典, 进而提出一种高效的基于匹配追踪的图像稀疏分解算法. 实验结果表明: Gabor 感知多成份字典具有对图像中平滑、边缘与纹理结构的自适应性, 与 Anisotropic refinement-Gaussian (AR-Gauss) 混合字典相比以较少的原子实现了对图像更为高效的稀疏分解.

**关键词** 稀疏表示, 视觉感知, 几何结构, Gabor 感知多成份字典, 匹配追踪  
**中图分类号** TN911.73

## Sparse Representations of Images by a Multi-component Gabor Perception Dictionary

SUN Yu-Bao<sup>1</sup> XIAO Liang<sup>1</sup> WEI Zhi-Hui<sup>2</sup> SHAO Wen-Ze<sup>1</sup>

**Abstract** It is currently a hot research topic that how to design an effective over-complete dictionary matching various geometric structures of images to provide sparse representation of images. A multi-component Gabor perception dictionary matching various image structures is constructed in terms of geometric properties of the local structures and the perception character of HVS. Furthermore, an effective algorithm based on the matching pursuit method is proposed to obtain sparse decomposition of images with our dictionary. The experimental results indicate that the Gabor multi-component perception dictionary can adaptively provide a precise and complete characterization of local geometry structures, such as plain, edge and texture in images. In comparison with the anisotropic refinement-Gaussian (AR-Gauss) mixed dictionary, our dictionary has a much sparser representation of images.

**Key words** Sparse representation, visual perception, geometrical structure, multi-component Gabor perception dictionary, matching pursuit

调和与分析认为, 图像可以表示为一组基函数的线性组合, 如 Fourier 基函数、小波函数等. 受人类视觉皮层神经元响应的稀疏性启发<sup>[1]</sup>, Olshausen 提出了一种对自然图像更为有效的表示方法: 稀疏表示 (Sparse representation)<sup>[2-3]</sup>. 稀疏表示模型要求

图像线性展开中大部分基函数的系数为零, 只有少数基函数具有较大的非零系数. 这里称基函数为原子, 所有原子信号的集合为字典. 稀疏表示使信号能量只集中于较少的原子, 具有非零系数的原子揭示了信号的主要特征与内在结构. 它在图像处理与计算机视觉中应用广泛, 许多问题的有效解决都依赖于信号的稀疏表示, 如图像恢复、图像压缩、模式分类等.

图像是一种具有多种结构成份的二维复杂信号, 当前的正交变换系统很难对其形成有效表示. 增加字典中原子个数形成非正交的冗余系统可提高匹配的灵活性, 有利于形成对图像这类复杂信号的稀疏表示. 当字典中原子个数  $L$  大于信号维数  $N$  且字典中包含  $N$  个线性无关的向量能够张成整个信号空间时, 字典为过完备的 (Overcomplete) 或冗余的 (Redundant). 信号在此冗余字典下的分解并不唯一, 这为信号的自适应表示提供了可能. 由于稀疏表示的广泛应用, 可依据稀疏性度量标准寻求信号在此字典下的最稀疏表示. 这里采用  $L_0$  范数量度信号的稀疏性, 基于冗余字典的信号稀疏表示模型为

$$\min \|\alpha\|_0 \quad \text{s.t.} \quad f = \Phi\alpha \quad (1)$$

收稿日期 2007-09-20 收修改稿日期 2007-12-25  
Received September 20, 2007; in revised form December 25, 2007

国家高技术研究发展计划 (863 计划) (2007AA12E100), 国家自然科学基金 (60672074), 江苏省自然科学基金 (BK2006569), 中国博士后科学基金 (20060390285), 江苏省博士后科学基金 (200601005B), 教育部高校博士点专项科研基金 (M200606018) 资助

Supported by National High Technology Research and Development Program of China (863 Program) (2007AA12E100), National Natural Science Foundation of China (60672074), Natural Science Foundation of Jiangsu Province of China (BK2006569), China Postdoctoral Science Foundation (20060390285), Postdoctoral Science Foundation of Jiangsu Province of China (200601005B), and Research Fund for the Doctoral Program of Higher Education of China (M200606018)

1. 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 南京 210094 2. 南京理工大学理学院应用数学系 南京 210094

1. Laboratory of Pattern Recognition and Artificial Intelligence, Institute of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094 2. Department of Applied Mathematics, Institute of Science, Nanjing University of Science and Technology, Nanjing 210094

DOI: 10.3724/SP.J.1004.2008.01379

其中,  $f \in \mathbf{R}^N$  为待分解信号, 维数为  $N$ ;  $\alpha \in \mathbf{R}^L$  为系数向量, 维数为  $L$ ;  $L_0$  范数定义为系数向量  $\alpha$  中非零系数的个数  $\|\alpha\|_0 = \#\{i, \alpha[i] \neq 0\}$ . 将冗余字典中的  $L$  个原子作为列向量形成稀疏表示矩阵  $\Phi \in \mathbf{R}^{N \times L}$ . 由于  $L_0$  范数是非凸的, 求解信号在冗余字典下的稀疏表示是一个需要组合搜索的 NP-hard 问题, 不存在求解此最优化问题的已知的多项式时间算法, 需要采用次优的逼近方法. 最近提出了许多获取信号稀疏表示的有效方法, 主要为松弛方法与贪婪方法. 典型的贪婪算法有匹配追踪<sup>[4]</sup>与正交匹配追踪<sup>[5]</sup>. 典型的松弛方法有基追踪算法<sup>[6]</sup>与 FOCUSS<sup>[7]</sup>算法. 最近研究者们又提出了一类基于投影的稀疏分解算法, 如梯度投影算法<sup>[8]</sup>与交叉投影算法<sup>[9]</sup>. 其中, 匹配追踪是一个迭代的贪婪算法, 它在每一次迭代过程中, 从字典中选择最能匹配信号结构的一个原子而构建一种逼近过程, 巧妙地减少了计算复杂性, 是现有稀疏分解算法中复杂度较低的算法, 每一次迭代的运算复杂度为  $O(L)$ , 其中  $L$  为字典中原子个数. 由此可见, 图像匹配追踪稀疏分解的效率由字典中原子个数  $L$  决定.

图像稀疏表示理论中的一个关键问题就是如何设计有效的稀疏表示过完备字典. 当前研究者们提出了许多冗余字典, 如小波包字典、小波和正弦函数的级联、局部余弦字典、各向同性 Gabor 字典<sup>[10]</sup>、Anisotropic refinement-Gaussian (AR-Gauss) 混合字典<sup>[11]</sup>等. 前两个字典中都包含小波原子, 由于其可分离性与各向同性, 并不能有效表示图像中的边缘结构. 局部余弦字典虽然能够有效匹配纹理结构, 却不能稀疏表示边缘轮廓结构. 文献 [10] 最早使用 Gabor 字典对二维图像进行稀疏分解, 不过字典中使用的是各向同性的单一空间频率带宽的 Gabor 原子, 不利于捕获图像中边缘轮廓结构的几何正则性. 文献 [11] 中采用 Gauss 函数与 Gauss 函数的二阶导数作为原子的生成函数, 它能够有效匹配图像中边缘轮廓结构. 但该字典中原子数量巨大, 增加了稀疏分解的复杂度, 同时字典中原子不能有效匹配振荡的纹理样式. 如何设计合适的能够匹配各层面几何结构的图像稀疏表示过完备字典, 进而形成对图像的稀疏分解是当前研究者关注的热点问题.

为了设计出有效的能够匹配各层面图像结构的稀疏表示过完备字典, 本文依据 Meyer 的卡通纹理图像模型<sup>[12]</sup>, 建立了相应的多成份字典, 每一子成份字典能够最优匹配图像中特定的结构成份, 从而能够捕获图像中的多个重要结构, 形成对图像的有效表示. 可通过对生成函数进行平移、旋转、伸缩等几何变换生成一系列原子来构造字典, 其关键是如何选择能够匹配图像局部几何结构的生成函数. 人

类视觉系统通过视觉皮层中多种具有不同感受野神经元的层次处理实现了对自然图像的稀疏表示, 其中侧抑制与局部竞争是人类视觉系统选择与视觉刺激模式最匹配神经元的重要方式. Gabor 函数能够有效地刻画简单细胞的响应特性, 同时匹配追踪每一次迭代选择最优匹配原子的过程很好地模拟了侧抑制与局部竞争机制. 因此, 本文模拟人类视觉系统的感知机制, 统一以二维 Gabor 函数作为原子的生成函数, 生成整个匹配图像几何结构特性的 Gabor 感知多成份字典, 采用匹配追踪算法在整个感知多成份字典中寻求图像的稀疏分解. 并且依据视觉皮层中神经元的响应特性、组织方式与自然图像的几何结构特性, 约束生成函数中自由参数的取值范围, 实验结果表明在保持字典性能的同时, 大幅度缩减了字典中原子个数, 提高了匹配追踪稀疏分解效率. 本文的多成份 Gabor 感知字典不同于文献 [10–11] 中的字典, 它是经多方向和各向异性调制的能够有效匹配平滑、边缘轮廓与纹理结构的感知多成份字典. 将本文的 Gabor 感知多成份字典与文献 [11] 的混合字典进行分析比较, 实验结果表明本文字典具有更强的稀疏表示能力, 能够有效表示图像中从平滑、边缘轮廓到纹理等多种几何结构.

## 1 基于多成份字典的图像稀疏表示模型

如何依据图像模型设计有效的表示方法是图像分析中的基本问题. 任何一个有效的图像表示方法都是针对特定的图像模型, 在此信号模型下获取的有效表示. 图像模型是图像表示、处理和分析的基础. 按照 Meyer<sup>[12]</sup> 的观点, 将图像模型为边缘卡通成份 (包括平滑与边缘轮廓结构) 和纹理成份的“和”

$$I \cong I_{\text{cartoon}} + I_{\text{texture}} \cong I_{\text{smooth}} + I_{\text{edge}} + I_{\text{texture}} \quad (2)$$

其中“+”代表图像中含有这三种结构成份. 该模型能够较好地逼近真实的图像信号, 在模型的准确性和易处理性之间取得一个较好的折中. 为了形成对图像的稀疏表示, 字典中原子应能匹配图像中的平滑、边缘、轮廓、纹理等多种局部几何结构, 增加字典中原子种类数形成非正交的冗余系统, 可提高匹配的灵活性, 有利于形成对图像这类复杂信号的稀疏表示. 为此, 依据建立的图像模型, 分别建立适合于表示图像模型中每一结构成份的子成份字典  $D_j$ , 子字典中的原子最优匹配图像中特定的局部几何结构. 这些子字典相互补充, 组合为一个大的多成份字典 (Multi-component dictionary, MCD)  $D = \bigcup_{j=1}^n D_j$ , 其中  $n$  为子成份字典个数. 从而能够匹配图像中的多个重要结构, 形成对图像的有效表示. 设多成份字典  $D = \{\phi_\gamma, \gamma \in \Gamma\}$  且原子  $\phi_\gamma$  具有

单位范数, 信号依据多成份字典的原子分解可表示为

$$\mathbf{f} = \sum_{j=1}^n \sum_{i \in \Gamma_j} \alpha_i \phi_i \quad (3)$$

其中,  $\Gamma_j \in \Gamma$  是子字典  $D_j$  中原子的下标集合且  $\bigcup_{j=1}^n \Gamma_j = \Gamma$ .  $\alpha_i$  为原子  $\phi_i$  的分解系数. 根据本文采用的图像模型, 需要三个子字典  $D_{smooth}$ ,  $D_{edge}$ ,  $D_{texture}$ , 分别对应图像成份  $I_{smooth}$ ,  $I_{edge}$ ,  $I_{texture}$ .

由于用于分解信号的多成份字典是过完备的, 信号表示方式并不唯一, 这为信号的自适应表示提供了可能. 由于信号稀疏表示在图像处理中的广泛应用, 采用  $L_0$  范数作为信号的稀疏性度量标准, 选择信号在此多成份字典下的最稀疏表示. 基于多成份字典的稀疏表示模型如下

$$\begin{aligned} \min \|\alpha\|_0 \\ \text{s.t. } \mathbf{f} = [\Phi_{smooth}, \Phi_{edge}, \Phi_{texture}] \alpha \end{aligned} \quad (4)$$

其中, 分别将各子成份字典  $D_{smooth}$ ,  $D_{edge}$ ,  $D_{texture}$  中的原子作为列向量形成矩阵  $\Phi_{smooth}$ ,  $\Phi_{edge}$ ,  $\Phi_{texture}$ . 这是一个 NP-hard 组合问题, 需要采取次优的逼近方法, 引言中介绍的各种稀疏分解算法如匹配追踪、基追踪算法仍然适用.

依据建立的图像模型, 选择能够最优匹配图像中相应几何结构成份的参数化生成函数  $\phi_i(\mathbf{x})$ :  $\mathbf{R}^2 \rightarrow \mathbf{R}$ ,  $i = 1, 2, 3$ , 并对各参数化产生函数进行平移、各向异性的伸缩和多方向旋转等几何变换形成一系列原子, 构成本文的冗余的多成份字典. 逼近求解多成份字典的稀疏表示模型 (4), 可获取信号在此多成份字典下的稀疏表示, 其核心问题是如何选择原子的生成函数. 下面将进一步分析如何依赖人类视觉系统早期阶段的感知机制选择与几何结构自适应的生成函数.

### 1.1 匹配图像结构的 Gabor 感知多成份字典

人类视觉系统能够自适应于自然环境中输入刺激的统计特性, 通过视觉皮层中具有不同感受野的神经元的层次处理实现了对自然图像的稀疏编码. 本原视觉皮层中的大部分神经元都具有一定的空间局部性、方向选择性与空间频率选择性, 并且只对特定形状结构的刺激才具有最优响应. 例如, 简单细胞的感受野由一些具有一定朝向的平行相间的兴奋与抑制条纹区域组成, 对光点刺激并没有反应, 只有当特定朝向和宽度的条形刺激出现在其感受野内某个特定位置上时, 神经元的响应最强. 而当刺激偏离该朝向时则响应急剧降低, 甚至消失, 它们被形象地称为特征检测器. 本文模拟人类视觉系统的感知机制来形成对图像的稀疏表示, 将字典中的每个原子看作一个神经元, 整个字典则对应本原视觉皮层中

神经元整体, 并且字典中原子具有类似本原视觉皮层中神经元的响应特性: 空间局部性, 方向性和带通性, 并只匹配特定的局部几何结构.

为了更精确地分析神经元的功能, 需要建立刻画其响应特性的可计算模型即感受野函数. 工程上可将感受野函数看作神经元的脉冲响应函数. Daugman 采用二维 Gabor 函数作为简单细胞的感受野函数<sup>[13]</sup>, 刻画其响应特性. 根据 Heisenberg 测不准原理, Gabor 函数具有最小面积的时频窗口, 能够有效编码视觉信号. Jones 证实该函数匹配测量的实验数据<sup>[14]</sup>, 说明了此模型的有效性. 其他神经元如复杂细胞, 超复杂细胞, gratingcell 等的可计算模型也都是基于二维 Gabor 函数, 只是进行了一些相应的后处理. 因此, 本文将统一以二维 Gabor 函数作为原子的产生函数, 生成整个匹配图像几何结构特性的 Gabor 感知多成份字典. 并且依据视觉皮层中神经元的响应特性、组织方式与自然图像的几何结构特性, 约束二维 Gabor 函数中自由参数的取值范围, 缩减字典中原子个数, 提高稀疏分解效率.

作为原子的产生函数, 二维 Gabor 具有七个自由参数, 其模型如下所示

$$\begin{aligned} g_{\gamma} &= K(\gamma) \exp\left(-\frac{\hat{x}^2 + \kappa^2 \hat{y}^2}{2\sigma^2}\right) \cos\left(2\pi \frac{\hat{x}}{\lambda} + \varphi\right) \\ \hat{x} &= (x - x_0) \cos \theta + (y - y_0) \sin \theta \\ \hat{y} &= -(x - x_0) \sin \theta + (y - y_0) \cos \theta \end{aligned} \quad (5)$$

其中,  $\gamma = \{\sigma, \lambda, \theta, \kappa, \varphi, x_0, y_0\}$ ,  $K(\gamma)$  为归一化常数, 使 Gabor 原子具有单位范数. 参数  $(x, y) \in \Omega$  表示图像中像素点的位置;  $(x_0, y_0)$  确定原子的中心位置; 标准差  $\sigma$  决定了原子支撑区域的大小, 类似于神经元的感受野. 变换参数  $\sigma$ , 可以使原子匹配图像中不同空间尺度的结构. 当刺激点的位置距离感受野中心  $(x_0, y_0)$  大于  $2\sigma$  时, 神经元具有非常弱的响应, 通常可以忽略. 模型中的高斯因子很好地刻画了神经元的空间局部性. 参数  $\kappa$  确定了高斯因子的离心率, 从而决定了原子支撑区域的宽度与长度的比率 (纵横比), 也称参数  $\kappa$  为纵横比因子. 为了匹配边缘轮廓结构, 要求  $\kappa < 1$  使原子各向异性. 角度参数  $\theta$  ( $\theta \in [0, \pi)$ ) 定义为法线与平行条纹之间的夹角, 在方程 (5) 中法线为  $x$  轴. 指示了原子中局部几何结构的方向, 表明原子的方向选择性. 变换参数  $\theta$ , 可以使原子具有不同的最优响应方向, 匹配图像中不同方向的结构. 参数  $\varphi \in [-\pi, \pi]$  为余弦调和因子  $\cos(2\pi(\hat{x}/\lambda) + \varphi)$  的偏移相位, 确定了原子  $g$  的对称性: 当  $\varphi = 0$  与  $\varphi = \pi$  时, 函数关于支撑区域中心位置对称; 当  $\varphi = \pm 0.5\pi$  时, 函数  $g$  是反对称的; 其他取值情况下函数是非对称的. 最后, 参数  $\lambda$

为波形调制因子  $\cos(2\pi(\hat{x}/\lambda) + \varphi)$  的波长, 相应地  $1/\lambda$  为余弦因子的空间频率, 表征原子的最优响应频率. 神经生理学研究表明参数  $\sigma$  与  $\lambda$  是相互依赖的, 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形区域个数. 原子的半幅空间频率带宽 (Half-amplitude bandwidth, 采用对数尺度的倍频程度量)  $b$  与比率  $\sigma/\lambda$  具有如下关系

$$b = \log_2 \left( \frac{\frac{\sigma}{\lambda} + \sqrt{\frac{\ln 2}{2}}}{\frac{\sigma}{\lambda} - \sqrt{\frac{\ln 2}{2}}} \right), \quad \frac{\sigma}{\lambda} = \frac{1}{\pi} \sqrt{\frac{\ln 2}{2}} \frac{2^b + 1}{2^b - 1} \quad (6)$$

可见, 比率  $\sigma/\lambda$  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. 二维 Gabor 原子的方向带宽 (度) 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

$$\Delta\theta_{\frac{1}{2}} = \arcsin \left[ \kappa \frac{2^b - 1}{2^b + 1} \right] \quad (7)$$

该式表明变换参数  $\kappa$  能够调节原子方向带宽与空间频率带宽的分配.  $\kappa$  较小时, 原子方向带宽较窄, 具有较强的方向选择性.  $\kappa$  较大时, 原子空间频率带宽较窄, 具有较强的空间频率选择性. 当固定纵横比因子  $\kappa$ , 方向带宽随空间频率带宽的增加而增加. 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.

变换空间频率带宽可以使原子的感受野中呈现不同的条纹个数. 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, 隶属于相应的子成份字典.

1) 平滑成份子字典  $D_{smooth}$

如果原子波形调制因子的振荡频率为 0, 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 (见图 1(a)), 组成平滑成份子字典.

2) 边缘成份子字典  $D_{edge}$

当支撑区域中包含条纹个数较少 (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, 构成边缘成份子字典  $D_{edge}$ . 此时, 如果  $\varphi = 0$  或  $\varphi = \pi$ , 则原子是对称的, 沿波形调制因子的传播方向具有类似墨西哥帽小波 (高斯函数的二阶导数) 的特性, 能够捕获图像中的屋顶型边缘 (见图 1(b)). 如果  $\varphi = \pm 0.5\pi$  时, 原子则是反对称的, 沿波形调制因子的传播方向具有类似高斯小波 (高斯函数的一阶导数) 的特性, 能够捕获图像中的阶跃型边缘 (见图 1(c)). 同时, 在垂直于波形调制因子

传播的方向, 原子呈现平滑特性. 因此 Gabor 原子过边缘方向是奇异的, 沿边缘方向则是光滑的, 较好地匹配了图像边缘轮廓的局部几何正则性, 能够有效地表示图像中的边缘轮廓结构.

3) 纹理成份子字典

当原子支撑区域中包含条纹个数较多 (本文选取为大于等于 4) 时, 原子表现为强烈的振荡样式, 最优匹配图像中振荡的纹理结构 (见图 1(d)), 对应纹理成份子字典  $D_{texture}$ .

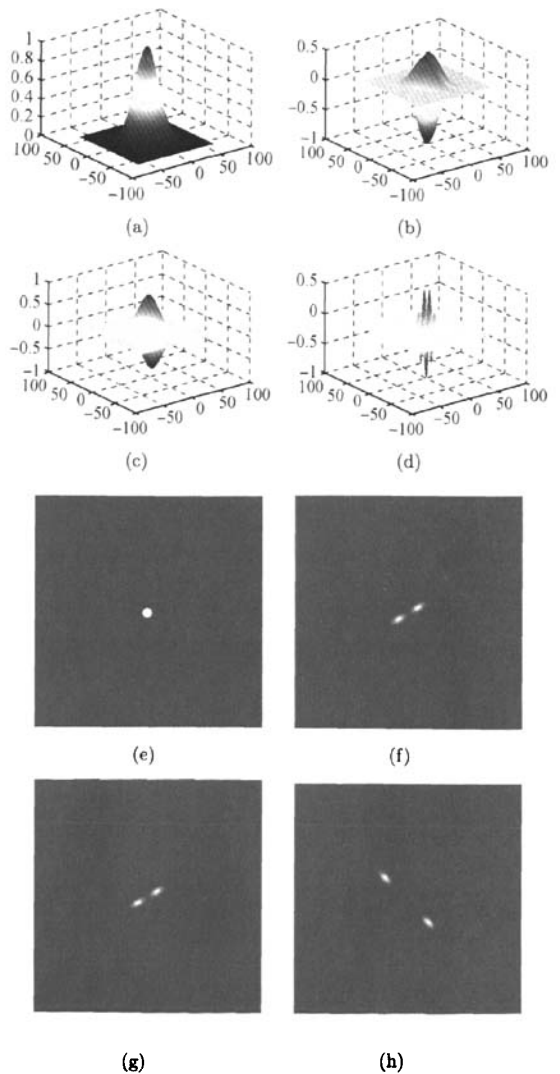


图 1 不同结构类型 Gabor 原子的三维图形与相应的傅里叶能量谱 ((a)~(d) 给出了不同类型原子的三维图形; (e)~(h) 为相应原子的傅里叶能量谱强度映射平面图.)  
Fig. 1 The surfaces of 2D Gabor atoms with different structures and the power spectrum intensity map images of them ((a)~(d) display the surface of them and (e)~(h) display their power spectrum correspondingly.)

不同结构类型 Gabor 原子对与其尺度、方向、中心位置、相位、结构类型相一致的图像内容形成最优响应。变换参数  $\sigma, \kappa, \theta$  可以使 Gabor 原子最优匹配图像中不同尺度、不同纵横比的、不同方向的边缘轮廓或纹理结构。原子的各向异性与多方向性增强了字典的稀疏表示能力。组合调制这些自由参数可生成整个多成份字典。最终字典中的每个原子可通过一组参数  $\sigma, \kappa, \theta, \varphi, b, x_0, y_0$  来标定。

图 1 展示了不同类型二维 Gabor 原子的三维图形与相应的傅里叶能量谱。其中傅里叶能量谱强度映射图中暗的灰度表示相应空间频率的能量较小或接近于 0, 亮的灰度则代表较大的频谱能量。图 1(a) 给出了当 Gabor 函数退化为高斯函数的三维图形, 其具体参数值为  $\sigma = 16, \kappa = 1.0, \theta = 0^\circ$ 。图 1(b)~(d) 中所示 Gabor 原子各具体参数取值分别为  $b = 1.5, \sigma = 8, \kappa = 0.5, \theta = 30^\circ, \varphi = 0$ ;  $b = 1.5, \sigma = 8, \kappa = 0.5, \theta = 30^\circ, \varphi = 0.5\pi$  与  $b = 0.5, \sigma = 8, \kappa = 0.5, \theta = 135^\circ, \varphi = 0$ 。二维 Gabor 函数能够有效检测与建模图像中的边缘轮廓与纹理结构, 被广泛地应用于图像处理与分析, 如边缘轮廓检测与纹理分析等, 因此 Gabor 多成份字典中的原子能够有效匹配图像中边缘与纹理结构。同时如图 1(f)~(h) 所示, 二维 Gabor 函数的傅里叶变换为两个高斯函数的和, 并且它们的中心关于原点对称, 这也正说明了二维 Gabor 函数具有空间频率与方向选择性, 能够刻画简单细胞的响应特性, 进一步证实了以二维 Gabor 函数为生成函数生成的原子能够有效刻画人类的视觉感知特性与图像的几何结构特性。

针对 Meyer 的卡通纹理图像模型, 建立了视觉感知与几何结构驱动的 Gabor 感知多成份字典。将自然图像的局部几何结构特性引入原子中, 使其能够自适应于图像中的边缘轮廓、纹理等结构。下一小节将讲述如何离散化 Gabor 生成函数中的各自由参数, 生成离散的 Gabor 感知多成份字典。

## 1.2 离散的 Gabor 感知多成份字典

本文的二维 Gabor 产生函数具有七个自由参数:  $\sigma, \kappa, \theta, \varphi, b, x_0, y_0$ 。实际应用中, 需要对产生函数的各自由参数进行离散采样。如果遍历各自由参数的所有可能取值范围, 则会产生非常庞大的字典, 无法实际应用于信号分解。依据视觉皮层中神经元的响应特性、组织方式与自然图像的几何结构特性, 约束自由参数的取值范围, 则会在保持字典性能的同时, 大幅度减少原子个数, 提高信号稀疏分解效率。

神经生理学研究表明简单细胞的感受野具有较大的动态范围, 最大的感受野宽度比最小的要大 30 多倍。相对而言, 简单细胞感受野的纵横比  $\kappa$  则具

有较小的变化区间  $0.2 \leq \kappa \leq 1^{[14]}$ 。神经生理学研究还发现生物视觉皮层中简单细胞的响应带宽取值范围较分散。猫视觉皮层中全体简单细胞的空间频率带宽的变化范围为 0.5~2.5 倍频程, 加权平均带宽为 1.32 倍频程, 加权方向半带带宽为  $17.5^\circ$ 。而短尾猿空间频率带宽为 0.4~2.6 倍频程, 中值带宽为 1.4 倍频程, 中值方向半带带宽为  $20^\circ^{[13]}$ 。同时, Pollen 与 Ronner 发现视觉皮层中临近的简单细胞对不管绝对相位为何值, 两者相位相差  $\pi/2$ , 存在正交相位关系<sup>[15]</sup>。依据视觉皮层中神经元的响应特性与图像的局部几何正则性, 对原子产生函数中的各自由变量的约束如下:

- 1) 原子纵横比小于等于 1;
- 2) 原子中包含的平面波 (余弦或正弦波) 沿着椭圆形高斯的短轴传播;
- 3) 存在正交相位的原子对;
- 4) 原子的半幅空间频率带宽范围为 0.4~2.6 倍频程, 按照不同倍频程铺叠二维空间频率平面;
- 5) 原子的方向带宽随着空间频率带宽的增加而增加。

在构造离散 Gabor 感知多成份字典时, 自由参数的采样应遵循上面的约束。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空间  $L^2(\mathbf{R}^2)$  的框架。直观上, 为构造框架, 所有离散原子的 Heisenberg 盒子至少要能够覆盖整个空间与空间频率平面。调节原子的空间频率带宽与方向带宽, 使其具有不同的空间频率分辨率与方向分辨率, 能够灵活地匹配图像中的多种结构。本文按照多组不同倍频程带宽原子铺叠二维的空间频率平面, 并选取足够小的方向与尺度采样间隔, 保证每组相同倍频程带宽原子都能形成对二维空间频率平面的完全铺叠。多组不同倍频程带宽原子形成一个冗余的过完备多成份离散字典。

依据上面给出的多个约束条件, 下面给出本文离散 Gabor 感知多成份字典各自由参数离散值。由式 7, 表 1 (见下页) 列出了取不同空间频率带宽与纵横比因子时, 原子具有的方向带宽。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, 伸缩  $a > 1$  步长应充分小, 一般取为  $a = 2^{1/\nu}$ , 其中  $\nu$  是每一波段中的中间尺度个数。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, 形成对信号的平移不变表示。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将采取不同的离散方法。假定待分解图像的大小为  $N \times N$ , 下面给出多成份字典中各参数的具体离散值。

表 1 不同空间频率带宽与纵横比因子下 Gabor 原子具有的方向带宽

方向带宽 (度)	空间频率带宽 $b = 1.5$	空间频率带宽 $b = 1$	空间频率带宽 $b = 0.7$	空间频率带宽 $b = 0.5$
纵横比因子 $\kappa = 1.0$	57.1	38.9	27.5	19.8
纵横比因子 $\kappa = 0.7$	39.1	27.0	19.2	13.8
纵横比因子 $\kappa = 0.5$	27.6	19.2	13.7	9.8
纵横比因子 $\kappa = 0.3$	16.5	11.5	8.2	5.9

1) 针对平滑成份子字典  $D_{smooth}$ , Gabor 函数退化为高斯函数, 以高斯函数为其生成函数. 采取各向同性的尺度关系  $\kappa = 1$ , 取  $\nu = 1, a = 2$ , 按指数序列  $\{2^j\}_{j \in \mathbb{Z}}$  均匀采样尺度  $\sigma, j \in [0, \lfloor \log_2 N \rfloor - 1]$  限制最大尺度为原图像的一半. 由于高斯函数关于中心对称, 变换参数  $\theta$  没有意义. 平滑成份子字典中包含的原子个数复杂度为  $O(N^2 \log_2 N)$ .

2) 针对边缘成份子字典  $D_{edge}$ , 取空间频率带宽  $b$  为 1.5 倍频程, 纵横比因子  $\kappa$  分别取值为 0.7, 0.5 和 0.3, 使原子为各向异性. 取  $\nu = 1/b$ , 按指数序列  $\{2^{bj}\}_{j \in \mathbb{Z}}$  均匀采样尺度  $\sigma, bj \in [0, \lfloor \log_2 N \rfloor - 3]$  限制最大尺度为原图像的  $1/8$ . 相位取为正交相位对  $0, \pi/2$ . 根据表中所列方向带宽为采样间隔均匀采样方向参数  $\theta$ . 当  $\kappa = 0.7, 0.5, 0.3$  时分别将参数  $\theta$  离散为 6、8、12 个方向. 组合这些离散值生成整个边缘成份子字典. 边缘轮廓成份子字典中原子个数复杂度为  $O(N^2(\log_2 N - 2))$ .

3) 针对纹理成份子字典  $D_{texture}$ , 分别取带空间频率带宽为  $b = 1, 0.7, 0.5$ , 只取  $\kappa = 1$ , 原子为各向同性. 取  $\nu = 1/b, a = 2^b$ , 按指数序列  $\{2^{bj}\}_{j \in \mathbb{Z}}$  均匀采样尺度  $\sigma, bj \in [0, \lfloor \log_2 N \rfloor - 3]$  限制最大尺度为原图像的  $1/8$ . 相位取正交相位对  $0, \pi/2$ . 当  $b = 1, 0.7, 0.5$ , 依据表 1 所列的原子方向带宽, 应分别将参数  $\theta$  离散为 6、8、12 个方向. 为了提高原子匹配纹理的准确性, 将参数统一离散为 12 个方向. 组合这些离散值生成整个纹理成份子字典. 纹理成份子字典中包含的原子个数复杂度为  $O(N^2(\log_2 N - 2))$ .

上面给出了本文采用的各自由参数离散化方法. 采用文献 [11] 中的离散计划, AR-Gauss 混合字典中原子个数复杂度为  $O(N^2 \log_2 N^2)$ , 本文的 Gabor 感知多成份字典中原子个数复杂度仅为  $O(N^2 \log_2 N)$ , 有效地降低了字典中原子个数. 由于依据视觉皮层神经元的特性以及图像的局部几何结构特性限制约束字典中原子个数, 整个多成份字典仍然具有较少的原子个数, 提高了图像稀疏分解的效率. 经实验, 上面的离散计划在字典的大小与其稀疏表示性能方面取得了一个较好的折中. 也可以采取更精细的采样, 但字典中原子的数量增多, 增加了

稀疏分解时搜索复杂度.

2 匹配追踪算法

傅里叶变换具有 FFT 快速算法, 小波变换具有滤波器组迭代快速算法, 基于冗余字典的信号稀疏表示也需要相应的快速算法, 或者说应当提高信号稀疏分解的效率. 基于冗余字典的稀疏表示本身是一个 NP-hard 问题, 在过去的一些年里, 研究者提出了许多获取信号稀疏表示的逼近方法, 主要为松弛方法与贪婪方法. 最近研究者们又提出了一类基于投影的稀疏分解算法. 匹配追踪是一种贪婪算法, 算法复杂度是现有稀疏分解算法中复杂度较低的, 也是目前使用较为广泛的稀疏分解算法.

匹配追踪算法运用贪婪技巧降低了计算复杂度. 它是一个迭代的递归算法, 每一步从字典  $D = \{\phi_\gamma | \|\phi_\gamma\| = 1, \gamma \in \Gamma\}$  中选择一个与残差信号最匹配的原子, 每一步都使得信号的逼近更为优化. 算法具体步骤如下.

步骤 1. 初始化: 令初始残差信号为原信号  $R^0 f = f$ , 初始迭代次数  $i = 0$ , 并计算原始信号与字典中所有原子的内积:  $\alpha_\gamma = \langle R^0 f, \phi_{\gamma \in \Gamma} \rangle$ ;

步骤 2. 内积极大化搜索:  $\gamma_i = \arg \max |\alpha_{\gamma \in \Gamma}|$ ;

步骤 3. 残差信号更新:  $R^{i+1} f = R^i f - \alpha_{\gamma_i} \phi_{\gamma_i}$ ;

步骤 4. 内积系数更新:  $\alpha_\gamma = \langle R^{i+1} f, \phi_{\gamma \in \Gamma} \rangle$ ;

步骤 5. 停止规则: 如果残差信号能量小于给定的值  $\|R^i f\|^2 \leq \xi_{stop}$ , 停止迭代. 否则,  $i = i + 1$ , 转至步骤 2.

经过  $N$  次迭代后, 信号  $f$  可被稀疏分解为:  $f = \sum_{i=0}^{N-1} \alpha_{\gamma_i} \phi_{\gamma_i} + R^N f$ . 当  $N$  趋近无穷时, 在有限维信号空间  $\|R^N f\|$  按指数级收敛. 相应地, 获得信号分解为  $f = \sum_{i=0}^{+\infty} \alpha_{\gamma_i} \phi_{\gamma_i}$ . 匹配追踪算法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, 该算法的详细说明可参见文献 [10].

最新的早期视觉感知模型指出视觉皮层中各邻近神经元相互抑制<sup>[16]</sup>, 与视觉刺激结构最匹配的神经元的响应得到增强, 而不匹配的神经元则受到抑制, 通过局部竞争机制选择与视觉刺激结构最匹配的神经元. 将匹配追踪算法步骤 3 的

残差信号更新公式两边分别与  $\phi_\gamma$  作内积可得:  $\langle R^{i+1}f, \phi_\gamma \rangle = \langle R^i f, \phi_\gamma \rangle - \langle R^i f, \phi_{\gamma_i} \rangle \langle \phi_{\gamma_i}, \phi_\gamma \rangle$ . 用神经科学的术语来说, 它是通过所选择的权重为  $\langle \phi_{\gamma_i}, \phi_\gamma \rangle$  的模式而对  $\langle R^i f, \phi_\gamma \rangle$  产生的一个抑制, 其中权重是模式与  $\phi_\gamma$  的相关性度量, 并通过内积极大化搜索选择最优匹配的原子. 因此匹配追踪模拟了视觉皮层的侧抑制与局部竞争机制.

由上分析, 模拟人类视觉系统的感知机制, 本文选用匹配追踪算法在整个视觉感知与几何结构驱动的 Gabor 多成份字典中追踪图像的稀疏分解. 匹配追踪在字典中迭代选取与图像结构最匹配的 Gabor 原子, 精确地刻画了图像中边缘与纹理结构的位置、尺度、方向、幅度、相位、频率等信息. 算法复杂度主要由字典中包含的原子个数决定, 个数越多运算复杂度越高. 由于使用了文献 [10] 中的快速算法, 在本文原子个数复杂度为  $O(N^2 \log_2 N)$  的 Gabor 感知多成份字典中追踪图像的稀疏分解的运算复杂度降低为  $O(N \log_2 N)$ , 提高了分解效率.

### 3 实验结果与分析

将本文的 Gabor 感知多成份字典与 AR-Gauss 混合字典进行比较, 分析两者的匹配追踪稀疏表示 (逼近) 能力. 混合字典采用文献 [11] 中的离散计划, 本文的 Gabor 感知多成份字典采用 2.2 节给出的离散形式. 表 2 (见下页) 给出了针对不同标准测试图像 Barbara (256 × 256)、Lena (128 × 128)、Pepper (128 × 128) 两个字典的原子总数以及使用相同原子个数稀疏逼近重构图像的峰值信噪比 (Peak signal to noise ratio, PSNR). Pepper 为边缘轮廓结构丰富的图像, 而 Barbara、Lena 图像则具有纹理、边缘轮廓等多种结构特征. 由于依据视觉皮层神经元的特性以及图像的几何特性限制约束字典中原子个数, 本文的整个多成份字典在保持稀疏表示性能的同时仍然具有较少的原子个数, 提高了追踪效率. 图 2 列出了对标准测试图像 Barbara (256 × 256) 分别采用本文字典与 AR-Gauss 混合字典使用相同原子个数的稀疏逼近重构图像, 重构 PSNR 分别为 33.845 dB 与 35.592 dB, 本文字典高出 1.747 dB.

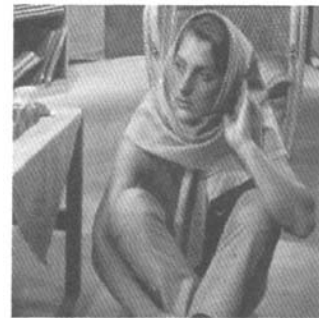
图 3 (见下页) 展示了图 2 中重构图像的局部放大对比图, 可以看出本文的 Gabor 感知多成份字典能够对纹理结构成份形成有效的表示, 这是因为本文的多成份字典包含能够有效匹配纹理结构的子成份字典  $D_{texture}$ . 图 4 (见下页) 给出了两个字典针对表 2 中标准测试图像 Barbara 的重构 PSNR 随着追踪的原子个数增加而增加的函数图形. 从图 4 与表 2 中可以看出, 对于边缘轮廓丰富的图像, 两个字典具有非常接近的稀疏表示能力; 而对于纹理丰富的图像, 本文字典具有更强的稀疏逼近能力. 实验

结果表明了本文字典高效的稀疏表示 (逼近) 能力, 能够有效表示图像中的边缘、轮廓、纹理等多种几何结构.



(a) 标准测试图像 Barbara (256 × 256)

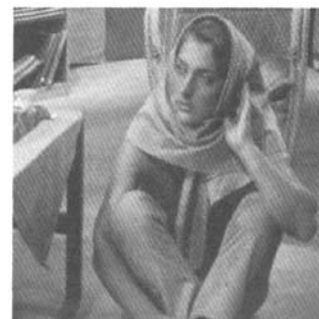
(a) A standard test image: Barbara (256 × 256)



(b) 采用本文字典的匹配追踪重构图像 (PSNR = 35.592 dB)

(b) Reconstruction image by our proposed dictionary

(PSNR = 35.592 dB)



(c) 采用 AR-Gauss 混合字典的匹配追踪重构图像 (PSNR = 33.845 dB)

(c) Reconstruction image by AR-Gauss mixed dictionary (PSNR = 33.845 dB)

图 2 不同字典相同原子个数 (2949 个) 的匹配追踪重构图像

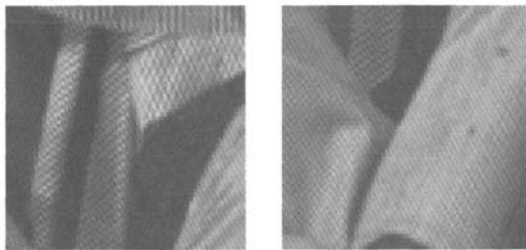
Fig. 2 The reconstruction images of matching pursuit approximation using the same number of atoms (2949) by different dictionaries



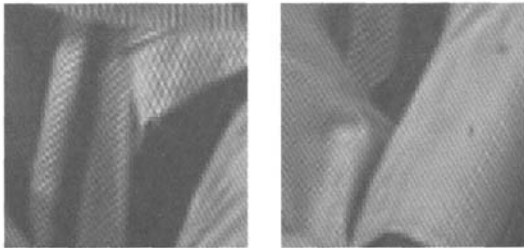
表 2 本文字典与 AR-Gauss 混合字典中的原子个数与匹配追踪重构图像质量比较

Table 2 Comparison of the numbers of contained atoms and matching pursuit approximation quality between our dictionary and AR-Gauss mixed dictionary

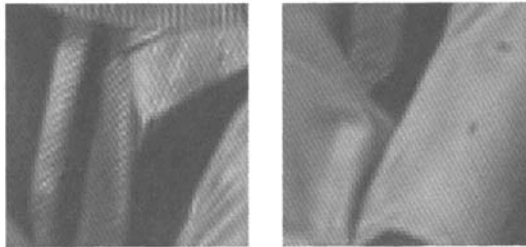
图像	本文字典原子数	混合字典原子数	重构原子数	混合字典重构 PSNR	本文重构 PSNR
Barbara (256 × 256)	49 676 288	161 087 488	2 949	33.845	35.592
Lena (128 × 128)	10 272 768	27 000 832	1 014	31.045	32.416
Pepper (128 × 128)	10 272 768	27 000 832	1 198	31.735	32.013



(a) 原局部图像  
(a) The original local images



(b) 采用本文字典的局部重构图像  
(b) Local images reconstructed by our proposed dictionary



(c) 采用 AR-Gauss 字典的局部重构图像  
(c) Local images reconstructed by AR-Gauss dictionary

图 3 不同字典匹配追踪重构图像局部放大对比图

Fig. 3 Comparison of local reconstruction images using different dictionaries

4 结论

本文依据 Meyer 的图像模型与人类视觉系统早期阶段的感知机制, 基于二维的 Gabor 函数设计了视觉感知与图像几何结构驱动的 Gabor 感知多成份

字典, 并依据视觉皮层中神经元的响应特性与自然图像的局部几何结构特性, 调制约束二维的 Gabor 生成函数中自由参数的取值范围, 大幅度降低了字典中原子个数, 从而降低了稀疏分解的复杂度. 为了说明本文字典的稀疏表示性能, 与 AR-Gauss 混合字典进行了比较, 实验结果表明本文字典具有更强的稀疏表示能力, 能够自适应于图像中的边缘、轮廓、纹理等多种几何结构. 下一步工作可将本文的 Gabor 感知多成份字典应用于图像编码、去噪等领域.

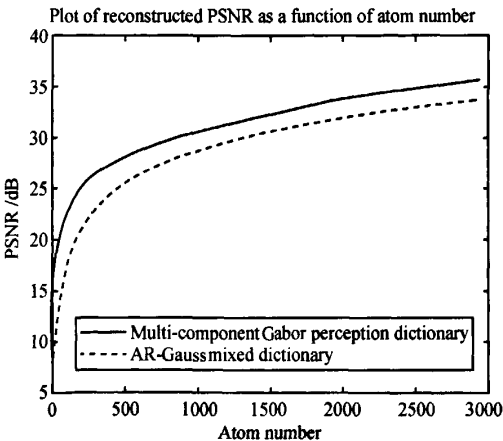


图 4 本文字典与 AR-Gauss 字典匹配追踪稀疏逼近率比较  
Fig. 4 Comparison of matching pursuit approximation rates between our dictionary and AR-Gauss mixed dictionary

References

1 Vinje W E, Gallant J L. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 2000, 287(5456): 1273–1276

2 Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse coding for natural images. *Nature*, 1996, 381(6583): 607–609

3 Olshausen B A, Field D J. Sparse coding with an overcomplete basis set: a strategy employed by V1? *Visual Research*, 1997, 37(33): 3311–3325



- 4 Mallat S G, Zhang Z F. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 1993, 41(12): 3397–3415
- 5 Davis G M, Mallat S G, Zhang Z F. Adaptive time-frequency decompositions. *SPIE Journal of Optical Engineering*, 1994, 33(7): 2183–2191
- 6 Chen S S, Donoho D L, Saunders M A. Atomic decomposition by basis pursuit. *SIAM Journal of Scientific Computing*, 1999, 20(1): 33–61
- 7 Gorodnitsky I F, Rao B D. Sparse signal reconstruction from limited data using FOCUSS: are-weighted minimum norm algorithm. *IEEE Transactions on Signal Processing*, 1997, 45(3): 600–616
- 8 Figueiredo M A T, Nowak R D, Wright S J. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 2007, 1(4): 586–598
- 9 Mancera L, Portilla J. L0-norm-based sparse representation through alternate projections. In: *Proceedings of IEEE International Conference on Image Processing*. Washington D. C., USA: IEEE, 2006. 2089–2092
- 10 Bergeau F, Malt S. Match pursuit of images. In: *Proceedings of the 1995 International Conference on Image Processing*. Washington D. C., USA: IEEE, 1995. 53
- 11 Ventura R F, Vandergheynst P, Frossard P. Low-rate and flexible image coding with redundant representations. *IEEE Transactions on Image Processing*, 2006, 15(3): 726–739
- 12 Meyer Y. *Oscillating Patterns in Image Processing and Non-linear Evolution Equation*. Boston: American Mathematical Society, 2001. 122
- 13 Daugman J G. Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society America A: Optics, Image Science, and Vision*, 1985, 2(7): 1160–1169
- 14 Jones J P, Palmer L A. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in the cat striate cortex. *Journal of Neurophysiology*, 1987, 58(6): 1233–1258
- 15 Pollen D A, Ronner S F. Phase relationships between adjacent simple cells in the visual cortex. *Science*, 1981, 212(4501): 1409–1411
- 16 Malik J, Perona P. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society America A: Optics, Image Science, and Vision*, 1990, 7(5): 923–932



孙玉宝 南京理工大学博士研究生. 主要研究方向为图像建模与稀疏表示, 图像压缩与通信.

E-mail: syb8692833@126.com

(SUN Yu-Bao Ph. D. candidate at the Institute of Computer Science and Technology, Nanjing University of Science and Technology. His research interest covers image modeling and sparse representation, image compression and communication.)



肖亮 副教授. 主要研究方向为变分偏微分方程在图像处理中的应用、模式识别、运动估计与跟踪、虚拟现实与系统仿真. E-mail: xtxiaoliang@163.com

(XIAO Liang Associate professor. His research interest covers variational PDE application in image processing, image modeling, pattern recognition,

motion estimation and tracking, virtual reality, and system simulation.)



韦志辉 博士, 教授. 主要研究方向为图像处理, 图像建模, 小波分析, 多尺度变换理论, 数字水印, 编码与压缩. 本文通信作者.

E-mail: gswei@mail.njust.edu.cn

(WEI Zhi-Hui Professor, Ph. D.. His research interest covers image processing, image modeling, wavelet analysis, multi-scale analysis, digital watermark, and image coding and compressing. Corresponding author of this paper.)



邵文泽 博士研究生. 主要研究方向为图像建模, 图像恢复, 图像插值, 超分辨率复原, 数学反问题.

E-mail: shaowenze@qianlong.com

(SHAO Wen-Ze Ph. D. candidate. His research interest covers image modeling, image restoration, image interpolation, super-resolution image reconstruction, and mathematical inverse problem.)