

基于 JPEG 序列的图像重建

朱翔 袁杰 都思丹
(南京大学电子科学与工程系 南京 210093)

摘要: 该文提出了一种基于 JPEG 序列的图像重建方法。该方法在已有的单帧图像复原技术的基础之上, 依据超分辨率重建的思想, 将凸集投影(POCS)理论与迭代反投影(IBP)算法相结合, 在频域内降低量化误差, 修复离散余弦系数。此外, 它还利用了最大后验概率(MAP)估计以及相应优化算法的特点, 在去除高斯噪声的同时, 保护边缘和细节信息。实验结果表明, 该方法一方面能够抑制高比率压缩所造成的块效应和振铃效应, 另一方面能较好地恢复图像的细节部分, 有效地提高图像的清晰度。

关键词: 图像处理; 超分辨率重建; 凸集映射; 迭代反投影; 最大后验概率

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2007)08-1841-04

Recovery of JPEG Compressed Image Sequences

Zhu Xiang Yuan Jie Du Si-dan

(Department of Electronic Science and Engineering, Nanjing University, Nanjing 210093, China)

Abstract: A method for the recovery of JPEG compressed image sequences is proposed in this paper. The method is based on the reduction of compression artifacts, and it is inspired with the thoughts of super-resolution reconstruction. To minimize the quantization errors and restore the discrete cosine transformation components, the method utilizes the theory of Projection Onto Convex Sets (POCS) and Iterative Back Projection (IBP) algorithm, which helps computing the projections. It also applies Maximum *A Posteriori* (MAP) estimation techniques to remove Gaussian noise without damaging edges or details of the image. Experimental results show that the proposed method produces good Peak Signal to Noise Ratio (PSNR) results, and could not only remove blocking and ringing artifacts, but also recover details of images, especially in low bit-rate conditions.

Key words: Image processing; Super-resolution reconstruction; Projection Onto Convex Sets (POCS); Iterative Back Projection (IBP); Maximum *A Posteriori* (MAP)

1 引言

超分辨率重建技术是近十年来图像重建领域中发展非常快的一个分支, 它的思想可以概括为利用图像序列在时间上的冗余换取空间上更高的分辨率。虽然基于非压缩图像的重建技术已经逐渐成熟, 但将它们直接应用于压缩图像序列后所获得的效果并不理想。

事实上, 对于我们在日常生活中经常使用的以 JPEG 编码为基础的视频压缩技术, 诸如 Motion-JPEG, MPEG-1, MPEG-2 等, 影响画面质量更多的不在于分辨率, 而在于由于分块和量化处理所造成的块效应与振铃效应。以往在针对单幅 JPEG 图像的复原方面, 人们尝试了大量的方法, 例如 Kim 在文献[1]中应用的自适应滤波就是一种常用的快速处理方式。凸集投影(Projection Onto Convex Sets, POCS)理论因其所具有的良好拓展性而在图像处理中得到广泛的应用。最早将 POCS 用于去除块效应的是 Rosenholtz 和 Zakhor^[2]。此后 Yang 和 Galatsanos^[3]等人在此基础上提出将

频域处理同空间域处理相结合的模式, 取得了比较好的效果。但是, 以上基于单幅图像的重建技术虽然可以在一定程度上去除噪声, 却无法很好地还原由于量化而丢失的信息, 特别是图像细节部分的信息。

本文在文献[3]的基础之上, 结合超分辨率重建的思想, 提出一种基于 JPEG 图像序列的重建方法。该方法吸取了 POCS 和最大后验概率(Maximum *A Posteriori*, MAP)估计的特点, 利用时间上的冗余信息去除块效应和振铃效应, 在一定程度上还原出清晰的图像细节。

2 基本模型

首先建立一个图像的退化模型。假设已知离散余弦(DCT)域的 $M \times N$ 图像序列: $g^1, g^2, \dots, g^k, \dots, g^n$, 其中 g^k 为参考帧。这里的图像均用列向量表示, 长度即为 MN , 它们与对应的原始空域图像关系为

$$g^i = Q(Bf^i), \quad i = 1, 2, \dots, n \quad (1)$$

其中 B 代表 BDCT 变换矩阵, $Q(\cdot)$ 代表量化。若对 g^k 作逆 BDCT 变换, 则有

$$\tilde{f}^i = B^{-1}Q^{-1}(g^i), \quad i = 1, 2, \dots, n \quad (2)$$

显然由于量化的作用, f^i 与 \tilde{f}^i 并不相同, 两者存在量化误差。假设所有图像相对参考帧 g^k 的运动估计(精确到像素)已知, 它们之间的转换关系就可以表示成

$$f^i = H^{k,i} f^k + u^i + e^i, \quad i \neq k, \quad i = 1, 2, \dots, n \quad (3)$$

式中 $H^{k,i}$ 为从 f^k 到 f^i 的运动补偿矩阵, u^i 为在 f^i 中出现但 f^k 中没有的像素所组成的向量^[4], e^i 是加性噪声。式(1), 式(3)联立可得:

$$g^i = Q[B(H^{k,i} f^k + u^i + e^i)], \quad i \neq k, \quad i = 1, 2, \dots, n \quad (4)$$

本文算法的目标就是根据 $g^1, g^2, \dots, g^k, \dots, g^n$ 获得 \tilde{f}^k , 使其尽可能的接近理想的 f^k 。

3 凸集投影(POCS)法复原

凸集投影(POCS)^[2]的基本思想可以概括为: 将已知的压缩图像, 以及关于所恢复图像的先验知识等信息抽象为希尔伯特空间内的若干凸集(convex set): $C_i, i = 1, 2, \dots, m$ 。如果存在非空集合 $C_0 = \bigcap_{i=1}^m C_i$, 那么理想的恢复图像 $f \in C_0$ 。假设每个 C_i 都有其对应的投影算子 P_i , 通过迭代:

$$f_{n+1} = P_m P_{m-1} \dots P_1 f_n \quad (5)$$

最终可获得收敛到交集 C_0 中的某点, 该点就是理想恢复图像的较好近似。

Yang 和 Galatsanos 在文献[3]中描述了一种凸集: 基于原始压缩图像的凸集, 可以表示为

$$C_T \triangleq \{f \mid F_j^{\min} \leq (Bf)_j \leq F_j^{\max}, \forall j = 1, 2, \dots, M \cdot N\} \quad (6)$$

F_j^{\min} 和 F_j^{\max} 是由量化步长与 BDCT 域的 g 等多种因素共同决定的临界值。一种最直接的计算方法是: 假设变换域图像 g 中第 j 个点的值为 v , 相应步长为 s , 那么其临界值就分别为 $vs \pm s/2$ 。该方法的依据是: 原始图像做 BDCT 变换后(量化之前)该点的数值就在 $vs + s/2$ 和 $vs - s/2$ 之间。

在此基础上我们很自然地联想到将单幅图像的凸集拓展到图像序列的凸集组, 忽略加性噪声, 将式(3)代入式(6)就可以得到 $n-1$ 个新的凸集:

$$C_T^i \triangleq \{f \mid (F^i)^{\min} \leq (B(H^{k,i} f + u^i))_j \leq (F^i)^{\max}, \forall j = 1, 2, \dots, M \cdot N\} \quad (7)$$

有必要将这种拓展与超分辨率重建的思想作比较。基于序列的超分辨率重建之所以比基于单帧的图像放大能获得更多的信息, 关键在于物体在不同帧之间有亚像素(sub-pixel)的移动。同样, 利用压缩图像序列进行重建的根本在于亚块(sub-block)的变化。如果第 k, l 两帧图的区别恰好是以块为单位的平移, 那么 $C_T^k = C_T^l$ 。

在式(6)基础上设计的投影算子 P_T 最常见的形式为

$$P_T f = B^{-1} F \left\{ \begin{array}{l} F_j^{\min}, \quad (Bf)_j < F_j^{\min} \\ F_j^{\max}, \quad (Bf)_j > F_j^{\max} \\ (Bf)_j, \quad \text{其他} \end{array} \right. \quad (8)$$

该式的缺点在于如果量化步长较高, 误差就很明显, 重建的效果不佳。本文根据实验观察到的数值分布情况提出一种改进方案。

JPEG 压缩过程中, BDCT 变换后高频 DCT 系数一般比较低, 再经过量化, 多数为零。相比之下低频系数由于本身数值高, 量化步长小, 量化后相当一部分不为零。以量化后的数值为中心, 考察原系数在其两侧分布情况可以发现: 量化后为零的系数分布同量化后不为零的系数分布有很大区别。如图 1 和图 2 所示。

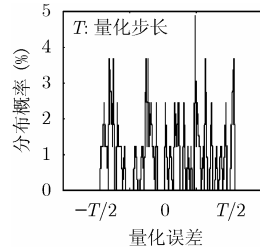


图 1 量化后不为零系数的误差分布

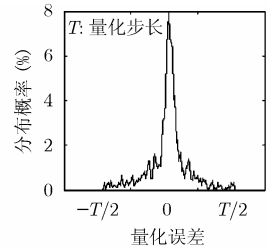


图 2 量化后为零系数的误差分布

图 1, 图 2 是将图像 Lena 按照如下量化矩阵 Q 压缩后, 用 MATLAB 统计所有 DCT 矩阵的系数 AC(3,2) 在量化误差所得到的分布。图 1 是针对量化后不为零的系数, 图 2 是针对量化后为零的系数。可以看出图 1 分布相对均匀; 而图 2 中多数点集中在零附近。

$$Q = \begin{bmatrix} 86 & 59 & 54 & 86 & 129 & 216 & 255 & 255 \\ 64 & 64 & 75 & 102 & 140 & 255 & 255 & 255 \\ 75 & 70 & 86 & 129 & 216 & 255 & 255 & 255 \\ 75 & 91 & 118 & 156 & 255 & 255 & 255 & 255 \\ 97 & 118 & 199 & 255 & 255 & 255 & 255 & 255 \\ 129 & 189 & 255 & 255 & 255 & 255 & 255 & 255 \\ 255 & 255 & 255 & 255 & 255 & 255 & 255 & 255 \\ 255 & 255 & 255 & 255 & 255 & 255 & 255 & 255 \end{bmatrix}$$

统计 DCT 矩阵上每一个量化后为零的系数的误差分布状况, 并统计包含 90% 以上点的差值范围与步长的比例, 可以得到如下参数矩阵 r :

$$r = \begin{bmatrix} 0 & 0.5 & 0.4 & 0.4 & 0.3 & 0.2 & 0.1 & 0.1 \\ 0.4 & 0.4 & 0.4 & 0.3 & 0.2 & 0.1 & 0.1 & 0.1 \\ 0.3 & 0.3 & 0.3 & 0.2 & 0.2 & 0.1 & 0.1 & 0.1 \\ 0.3 & 0.3 & 0.3 & 0.2 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.2 & 0.2 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix}$$

注意到 DC 分量一般都很高, 所以 DC 的分布情况不适合上述分析。需要指出的是, 对于不同的量化矩阵, r 的数值并不相同。但基本上量化步长越高的 DCT 分量, 其对应的 r 上的分量越低, 且基本不超过 0.5。此外, 虽然同样的 Q 用于不同的图像, 所得到的 r 也会有小幅的波动, 但我们认为

由 Lena 得到的 r 能够近似反映其它图像的系数分布情况, 如果这些图像也采用相同 Q 值压缩的话。

据此本文提出一种新的投影算子 P_T^i :

$$P_T^i \mathbf{f} = \mathbf{B}^{-1} \mathbf{F} \begin{cases} F_j^{\min}, & g_j \neq 0 \wedge (\mathbf{B}\mathbf{f})_j < F_j^{\min} \\ F_j^{\max}, & g_j \neq 0 \wedge (\mathbf{B}\mathbf{f})_j > F_j^{\max} \\ F_j^{\min} \times r(p, q), & g_j = 0 \wedge (\mathbf{B}\mathbf{f})_j < F_j^{\min} \\ F_j^{\max} \times r(p, q), & g_j = 0 \wedge (\mathbf{B}\mathbf{f})_j > F_j^{\max} \\ (\mathbf{B}\mathbf{f})_j, & \text{其他} \end{cases} \quad (9)$$

式中 (p, q) 是 j 点在其所属块中的坐标。

如果是针对参考帧 g^k 所对应的凸集, 直接代入式(9)就可以了。但应用到其它图像时由于存在运动补偿, 计算就比较复杂。可以将投影过程分为两步, 首先完成如下变换:

$$\bar{\mathbf{f}}^i = \mathbf{H}^{k,i} \mathbf{f} + \mathbf{u}^i, \quad i \neq k, \quad i = 1, 2, \dots, n \quad (10)$$

前面定义过 \mathbf{u}^i 是 \mathbf{f}^i 中出现, 但 \mathbf{f}^k 涉及不到的像素, $\mathbf{H}^{k,i} \mathbf{f}$ 在这些位置上的数值为零, 实际计算时可以用 $\bar{\mathbf{f}}^i$ 中相应像素的值代替。

按照式(9)将 $\bar{\mathbf{f}}^i$ 投影到 \mathbf{C}_T^i , 得到新的空域图像 $\bar{\mathbf{f}}'^i$, 理论上

$$P_T^i \bar{\mathbf{f}}^i = (\mathbf{H}^{k,i})^{-1} (\bar{\mathbf{f}}'^i - \bar{\mathbf{u}}'^i) \quad (11)$$

式中 $\bar{\mathbf{u}}'^i$ 为 \mathbf{u}^i 投影后的结果。将 $\bar{\mathbf{f}}'^i$ 中这些像素置零, 就得到 $\bar{\mathbf{f}}^i - \bar{\mathbf{u}}'^i$ 。

直接计算 $(\mathbf{H}^{k,i})^{-1}$ 是不现实的。本文采用迭代反投影法 (Iterative Back Projection, IBP) 求近似解^[5,6], 通过以下迭代:

$$\mathbf{f}_{(n+1)} = \mathbf{f}_{(n)} + \mathbf{h}^{\text{BP}} (\bar{\mathbf{f}}^i - \bar{\mathbf{u}}'^i - \mathbf{H}^{k,i} \mathbf{f}_{(n)}) \quad (12)$$

\mathbf{h}^{BP} 称作反投影核 (back-projection kernel), 它决定每点的误差对 $\mathbf{f}_{(n)}$ 中各点灰度值的影响大小。理论上 \mathbf{h}^{BP} 可以有多种选择, 最方便的就是直接采用 $(\mathbf{H}^{k,i})^T$ 。它的物理意义在于 $\mathbf{f}_{(n)}$ 中第 i 个像素对 $\bar{\mathbf{f}}^i$ 的第 j 个像素在变换中贡献越大, 那么 j 像素的误差对 i 像素的修正就越多。

通过求序列的交集, 基本上可以恢复邻近块之间的相关性, 从而很大程度上降低了块效应的影响。

4 最大后验概率 (MAP) 降噪

经过上述映射处理后得到的图像 $\hat{\mathbf{f}}^k$ 虽然更加接近理想的 \mathbf{f}^k , 但其 BDCT 变换 $\hat{\mathbf{g}}^k$ 与 $\mathbf{B}\mathbf{f}^k$ 之间仍存在误差, 本文将这些误差近似看作是零均值高斯分布的。因此, 空间域 $\hat{\mathbf{f}}^k$ 上会留有噪声。由于 DCT 为正交变换, 显然这些噪声也应当成零均值高斯分布。如果序列的帧数目较多, 可以通过分组处理再取中值的方法降噪, 但如果帧数不高, 就不适用了。

Schultz 和 Stevenson 在文献[7]中提出一种基于贝叶斯公式的图像放大技术, 即最大后验概率 (MAP) 估计。本文对该方法加以修改, 可以较好地抑制噪声。首先假定退化模型:

$$\mathbf{y} = \mathbf{z} + \mathbf{n} \quad (13)$$

\mathbf{y} 表示含噪图像, \mathbf{z} 是理想图像, \mathbf{n} 为高斯白噪声。希望获得 $\hat{\mathbf{z}}$, 满足

$$\hat{\mathbf{z}} = \arg \max_{\mathbf{z}} P(\mathbf{z} | \mathbf{y}) \quad (14)$$

通过贝叶斯公式, 可以转化为条件概率与先验概率的乘积。

其中条件概率形式为

$$P(\mathbf{y} | \mathbf{z}) = \frac{1}{(2\pi\sigma^2)^{NM/2}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{z}\|^2}{2\sigma^2}\right) \quad (15)$$

而 \mathbf{z} 的先验概率取决于平滑度。对于每一个像素, 将其与相邻的 8 个像素看成一组, 由式(16)评估:

$$\left. \begin{aligned} d_{k,l,0}^t \mathbf{z} &= z_{k,l+1} - 2z_{k,l} + z_{k,l-1} \\ d_{k,l,1}^t \mathbf{z} &= \frac{1}{2}(z_{k-1,l+1} - 2z_{k,l} + z_{k+1,l-1}) \\ d_{k,l,2}^t \mathbf{z} &= z_{k-1,l} - 2z_{k,l} + z_{k+1,l} \\ d_{k,l,3}^t \mathbf{z} &= \frac{1}{2}(z_{k-1,l-1} - 2z_{k,l} + z_{k+1,l+1}) \end{aligned} \right\} \quad (16)$$

为了将噪声与图像细节区分开来(噪声的起伏一般较小), 本文采用如下函数:

$$\rho_T(x) = \begin{cases} x^2, & |x| \leq T \\ T^2, & |x| > T \end{cases} \quad (17)$$

可看出 x 超过阈值 T 后函数值就不再增加, 这样就可以在后面的优化过程中保护大部分边缘和细节不丢失。与条件概率结合, 最终整个问题简化为式(18)的优化问题:

$$M_\lambda(\mathbf{z}, T) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} \sum_{m=0}^3 \rho_T(d_{k,l,m}^t \mathbf{z}) + \frac{\lambda}{2\sigma^2} \|\mathbf{y} - \mathbf{z}\|^2 \quad (18)$$

λ 的取值影响先验概率与条件概率的贡献比重。可以用最速下降法求解此类问题。

上述降噪方法最大的缺点在于阈值 T 的选择。如果 T 过大可能造成边缘等信息被损害, 过小会使部分噪声得不到抑制, 即平滑程度与细节恢复之间存在矛盾。实际应用中, 不同的场合对这两方面的要求也不一致。因此, 目前我们暂时没有使用某种定量的指导准则来计算 T , 而是通过输出图像的视觉效果选择和调整阈值。

5 实验结果

评判图像重建效果的一个重要指标是峰值信噪比 (Peak Signal to Noise Ratio, PSNR)。以 Lena 为例, 分别比较不同比特率下, 基于单帧的 POCS 重建法, 基于多帧的 POCS (采用低通滤波降噪) 重建法, 以及本文提出的 POCS+MAP 法的 PSNR 分贝值, 如表 1 所示。

后两种重建法分别用了 5 帧图像。可以看到, POCS+MAP 与单纯的 POCS 相比在高比特率的情况下并不具优势, 但是在低比特率下 POCS+MAP 的重建效果优于后者(高出近 1 个分贝)。这主要得益于低信噪比环境中 MAP 卓越的降噪特性, 以及函数 $\rho_T(\cdot)$ 对边缘和细节部分的良好保护。

表 1 不同比特率下重建的 Lena 图像 PSNR 值(dB)

方法	bit-rate			
	1.35 bpp	0.52 bpp	0.34 bpp	0.30 bpp
Original JPEG	32.271	29.272	27.136	26.424
POCS(single frame)	32.362	29.233	26.548	25.868
POCS(5 frames)	34.828	31.186	28.423	28.582
POCS+MAP(5 frames)	34.032	31.244	29.411	29.248

重建的效果除了数值角度的分析,还要从视觉上进行评判。图 3 是某芯片照片及其局部放大。

原始的 JPEG 图像由于高压缩比,芯片上的字迹几乎无法辨别。单帧的 POCS 修复虽然去掉了块效应和振铃效应,但不能很好地恢复细节。在多帧 POCS 所重建的图像上,已经可以看清多数字符,然而局部仍比较模糊。POCS+MAP 法则一定程度上避免了这种因为降噪而产生的模糊现象,图像清晰。

由此可以看出,本文提出的 POCS+MAP 技术具备在低信噪比环境中抑制噪声的同时,较好地还原细节信息的特点,适合于低比特率条件下的图像复原。

6 结束语

本文提出了一种基于 JPEG 序列的图像重建方法。该方

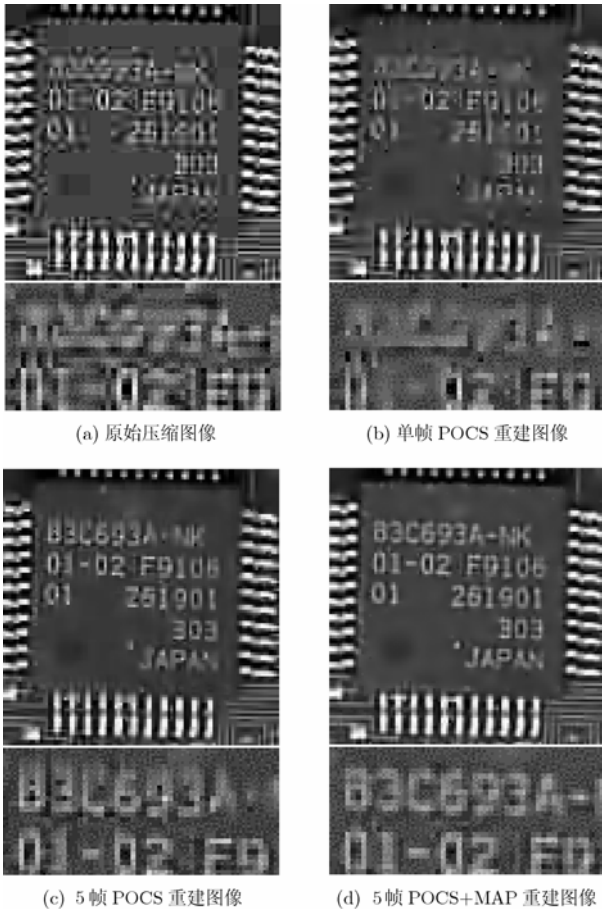


图 3 芯片图像序列的试验结果

法在已有的基于单帧图像复原技术的基础之上,结合超分辨率重建的思想,并综合了 POCS 理论与 MAP 估计的特点。实验结果表明,该方法在去除高比率压缩所造成的块效应和振铃效应的同时,能够较好地恢复图像的细节部分,有效地提高图像的清晰度。

需要指出的是,该方法虽然可以获得比较好的恢复效果,但由于大量应用了迭代算法,处理的速度不高。此外它依赖于对序列精确的运动估计,并在降噪过程中对阈值的选择敏感。未来我们将在这两方面作进一步的研究。

参考文献

- [1] Kim C. Adaptive post-filtering for reducing blocking and ringing artifacts in low bit-rate video coding. *Signal Processing: Image Communication*, 2002, 17(7): 525-535.
- [2] Rosenholtz R and Zakhor A. Iterative procedures for reduction of blocking effects in transform image coding. *IEEE Trans. on Circuits and Syst. for Video Tech.*, 1992, 2(1): 91-95.
- [3] Yang Y and Galatsanos P. Removal of compression artifacts using projections onto convex sets and line process modeling. *IEEE Trans. on Image Processing*, 1997, 6(10): 1345-1357.
- [4] Schultz R and Stevenson R. Extraction of high-resolution frames from video sequences. *IEEE Trans. on Image Processing*, 1996, 5(6): 996-1011.
- [5] Peleg S, Keren D, and Schweitzer L. Improving image resolution using subpixel motion. *Pattern Recognit. Lett.*, 1987, 5(3): 223-226.
- [6] Irani M and Peleg S. Improving resolution by image registration. *CVGIP: Graph. Models Image Process.*, 1991, 53(3): 231-239.
- [7] Schultz R and Stevenson R. A bayesian approach to image expansion for improved definition. *IEEE Trans. on Image Processing*, 1994, 3(3): 233-242.

朱翔: 男, 1982年生, 硕士生, 研究方向为图像处理与模式识别。

袁杰: 男, 1975年生, 副教授, 主要研究方向为数字图像处理。

都思丹: 女, 1962年生, 教授, 博士生导师, 主要研究方向为电路与系统。