

基于支持向量机的图象插值及错误隐匿策略

王 珏 季 梁

(清华大学自动化系智能技术与系统国家重点实验室, 北京 100084)

摘 要 如何在有损网络环境中传输的视频进行错误隐匿是视频传输研究中的基本问题. 支持向量机(SVM)是一种新兴的通用学习算法, 是国际上机器学习领域新的热点. 为了取得比现有方法更好的错误隐匿效果, 提出了一种新的基于支持向量机回归估计的错误隐匿策略, 首先建立了基于支持向量机回归估计的图象插值算法, 并将其引入到错误隐匿问题中, 然后用空域插值的方法达到错误隐匿的目的. 实验结果表明, 与目前采用的各种错误隐匿策略相比较, 基于支持向量机的错误隐匿策略在错误隐匿效果和推广性能上都具有一定的优越性.

关键词 错误隐匿 支持向量机 非线性插值

中图法分类号: TP391.4 TP181 文献标识码: A 文章编号: 1006-8961(2002)06-0558-07

Image Interpolation and Error Concealment Scheme Based on Support Vector Machine

WANG Jue, JI Liang

(State Key Laboratory of Intelligent Technology and Systems of China Department of Automation, Tsinghua University, Beijing 100084)

Abstract How to prevent quality degradation due to channel errors for images and video transmitting over lossy channels is a fundamental problem in multimedia signal processing. Support Vector Machine (SVM) is a novel powerful learning method and is now a new hotspot in the field of machine learning, and has been successfully used in many pattern recognition problems. To get more satisfying error concealment results, a novel error concealment scheme based on SVM is proposed in this paper. At first, a SVM based image interpolation algorithm is successfully established. SVM learning machines are carefully trained by a large amount of training data exacted from standard images to learn the relationship of neighboring pixels in spatial domain, and these well trained machines are used as specific nonlinear interpolation operators. Comparative results show that this kind of interpolation operator outperforms not only some traditional used interpolation operators such as linear and median operators, but also some operators carefully trained by artificial neural networks. The error concealment problem is placed into a spatial image interpolation framework and the proposed interpolation method is thoroughly used to estimate the missing image blocks according to their neighboring pixels. Experimental results show that compared with some error concealment schemes both in spatial and frequent domain in the literature, especially those based on artificial neural networks, the proposed one has remarkable advantages in error concealment performance and generalization property.

Keywords Error concealment, Support vector machine, Nonlinear interpolation

0 引 言

随着网络技术的飞速发展, 一方面要求多媒体信号在网络上的传输能够适应各种网络带宽条件,

另一方面, 一个新的研究领域正在成为多媒体信号传输中的热点, 这就是在可能出现错误的网络环境下, 如何实现多媒体信号的鲁棒性传输^[1]. 当 MPEG 视频流在出错网络环境(例如 ATM 网络由于拥塞而发生的信元丢失, 无线网络的信号衰减等)

中传输时,解码端解码的以 DCT 压缩的视频图象会发生宏块丢失错误.采用容错编码技术^[2]可以减少传输错误时数据的破坏,但当数据丢失错误比较严重的时候,基于编码端的主动型解决方案可能失效,这时需要在解码端实行被动型的错误隐匿策略.如何在接收端提高受损图象的感观质量,以达到错误隐匿的目的是当前视频传输中研究的热点问题.总体来说,解码端的错误隐匿策略的基本方法主要分为空域插值和频域插值两种^[3].Robie 等用 Hough 变换对丢失宏块进行了空域插值^[4];De Natale 等精心设计了神经网络以及一系列的特定训练样本,在频域对丢失的 DCT 系数进行恢复,取得了较好的效果^[5].然而,这种方法训练样本的生成以及训练过程都非常复杂,算法的推广性能较差.

支持向量机(SVM)是一种新的通用学习方法,它建立在统计学习理论的基础之上,对于有限样本学习问题已经表现出很多优于已有方法的性能,正在成为继神经网络研究之后新的研究热点.基于支持向量机的学习方法能够较好地解决小样本、非线性和高维数的问题,已在模式识别的多个问题中取得了良好的应用效果^[6].

基于训练效果和推广性能的综合考虑,将支持向量机引入到图象插值中,建立了基于支持向量机回归估计的图象插值算法,并将这种算法应用到错误隐匿问题中,提出了一种新的利用图象空间插值的方法进行视频数据错误隐匿的策略.

1 用于回归估计的支持向量机^[7]

首先用线性回归函数 $f(x) = \tilde{\omega} \cdot x + b$ 拟合数据 $\{x_i, y_i\}, i = 1, \dots, n, x_i \in \mathbf{R}^d, y_i \in \mathbf{R}$, 假设所有训练数据都可以在精度 ϵ 下无误差地用线性函数拟合,即

$$\begin{cases} y_i - \tilde{\omega} \cdot x_i - b \leq \epsilon \\ \tilde{\omega} \cdot x_i + b - y_i \leq \epsilon \end{cases} \quad i = 1, \dots, n \quad (1)$$

控制函数复杂性的方法是使回归函数最平坦,等价于最小化 $\frac{1}{2} \|\tilde{\omega}\|^2$. 在允许拟合误差的情况下,引入松弛因子 $\xi_i \geq 0$ 和 $\xi_i^* \geq 0$, 则式(1)变为

$$\begin{cases} y_i - \tilde{\omega} \cdot x_i - b \leq \epsilon + \xi_i \\ \tilde{\omega} \cdot x_i + b - y_i \leq \epsilon + \xi_i^* \end{cases} \quad i = 1, \dots, n \quad (2)$$

优化目标变成最小化 $\frac{1}{2} \|\tilde{\omega}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$.

利用 Lagrange 乘子,在条件

$$\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \quad 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, n \quad (3)$$

下,对 Lagrange 因子 α_i, α_i^* 最大化目标函数

$$W(\alpha, \alpha^*) = -\epsilon \sum_{i=1}^n (\alpha_i^* + \alpha_i) + \sum_{i=1}^n y_i (\alpha_i^* - \alpha_i) - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) (x_i \cdot x_j) \quad (4)$$

得回归函数为

$$f(x) = \tilde{\omega} \cdot x + b = \sum_{i=1}^n (\alpha_i^* - \alpha_i) (x_i \cdot x) + b^* \quad (5)$$

这里 α_i, α_i^* 中,将只有小部分不为 0,它们对应的样本就是支持向量.

对于非线性函数的拟合,可以通过非线性变换将其转化为某个高维空间的线性问题,在变换空间求最优拟合函数.但由于在式(5)中只涉及训练样本间的内积运算,根据泛函的有关理论,只要一种核函数 $K(x_i, x_j)$ 满足 Mercer 条件,它就对应某一变换空间的内积,因此高维空间的内积运算可以通过原空间的运算实现.

因此,在以上函数中,采用适当的内积函数 $K(x_i, x_j)$ 就可以实现某一非线性函数拟合,此时目标函数变为

$$W(\alpha, \alpha^*) = -\epsilon \sum_{i=1}^n (\alpha_i^* + \alpha_i) + \sum_{i=1}^n y_i (\alpha_i^* - \alpha_i) - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) K(x_i, x_j) \quad (6)$$

且回归函数变为

$$f(x) = \tilde{\omega} \cdot x + b = \sum_i (\alpha_i^* - \alpha_i) K(x_i, x) + b^* \quad (7)$$

以上的函数拟合方法就是基于支持向量机的函数拟合方法.

2 基于支持向量机回归估计的图象插值算法

图象插值是利用图象像素点之间的相关性,由相邻的已知像素点的信息来估计未知像素点的信息.由于 SVM 良好的学习性能,可以采用 SVM 学习机对图象像素点之间的相关性进行统计学习,并将其用于图象插值.基于 SVM 的图象插值方法如图 1 所示.考虑图象的行扩展和图象的行列同时扩展两种比较典型的插值情况.

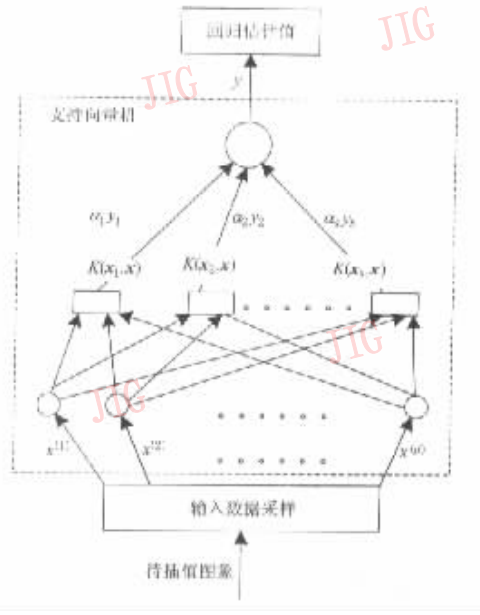


图1 基于支持向量机的图像插值方法

图象的行扩展是指在图象的相邻两行之间插入新的一行. 在实验中首先将原始图象的所有奇数行去掉, 然后用插值算法恢复. 对于不同方法恢复的图象, 通过计算恢复图象的峰值信噪比 (PSNR) 来比较不同方法的优劣.

SVM 学习机的输入和输出样本选择如图 2 所示. 由于 SVM 学习机是多输入单输出学习机, 因此采用 3 个 30 维输入向量的学习机进行训练和插值. 为了证实 SVM 学习机的推广性能, 训练样本和测试样本分别从不同的标准图象库中随机提取. SVM 学习机的核函数采用如下的高斯核函数

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{s^2}\right) \quad (8)$$

式中参数取 $s=900$.

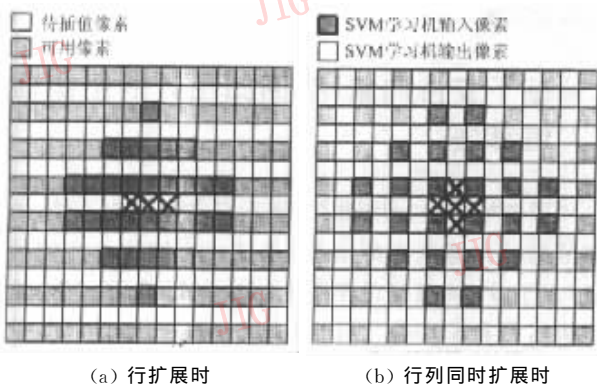


图2 SVM 学习机输入输出像素点位置选择

为了确定合适的训练样本数, 从训练图象库中

随机选择了 400~8 000 个共 11 组数目不等的训练样本集进行训练并将训练好的 SVM 学习机用于图象插值. 用 Lena 图象做为测试图象, 其插值恢复结果与训练样本数量的关系如图 3 所示, 可见, 当训练样本总数超过 5 000 个时, 训练结果已经趋于平稳. 因此在后继实验中均取 6 000 个样本作为训练集.

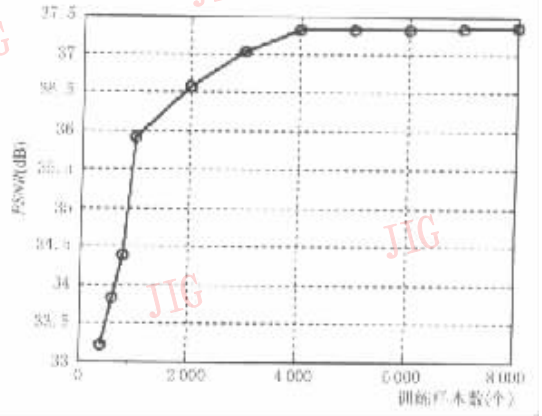


图3 SVM 学习机插值效果与训练样本数的关系

用训练好的 SVM 学习机对 Lena, Boats, Poster, Boy 和 Kodak 5 幅图象进行插值, 并将结果与线性插值、中值插值及两种人工神经网络的插值方法^[8]进行比较, 结果如表 1 所示.

表1 图象行扩展情况下不同插值算法的比较

插值图象	插值后图象的峰值信噪比 PSNR(dB)				
	线性插值	中值插值	神经网络 1 (Backprop)	神经网络 2 (RLS)	SVM 学习机
Lena	32.72	30.95	32.33	32.76	37.33
Boats	26.41	25.88	26.61	27.06	31.71
Poster	24.94	24.38	25.16	26.28	30.51
Boy	25.54	25.57	26.08	26.78	31.42
Kodak	28.59	26.26	28.65	28.78	32.52

图象的行列同时扩展以及这种情况下 SVM 学习机的输入输出数据如图 2(b) 所示. 与文献^[8]一样, 为保证缩小后的图象的感观质量, 先对原始图象进行如下的竖直和水平方向两次线性滤波

$$f_v(x, y) = \frac{1}{128} [64p(x, y) + 38(p(x, y-1) + p(x, y+1)) - 6(p(x, y-3) + p(x, y+3))] \quad (9)$$

$$f_H(x, y) = \frac{1}{128} [64p(x, y) + 38(p(x-1, y) + p(x+1, y)) - 6(p(x-3, y) + p(x+3, y))] \quad (10)$$

然后去掉图象的奇数行和列, 生成待插值图象. 通过训练 4 个 24 维输入的 SVM 学习机进行插值

恢复,其插值结果以及与其他算法的插值结果相比较,如表 2 所示.

表 2 图象行列同时扩展时不同插值算法的比较

插值图象	插值后图象的峰值信噪比 PSNR(dB)				
	线性插值	中值插值	神经网络 1 (Backprop)	神经网络 2 (RLS)	SVM 学习机
Lena	36.14	33.81	38.35	40.09	44.98
Boats	32.00	31.09	36.06	37.04	41.87
Poster	35.71	31.06	35.48	36.00	39.86
Boy	31.65	30.70	36.27	37.41	41.78
Kodak	36.04	28.64	36.86	37.63	42.43

从表 1 和表 2 的结果可以看出,基于 SVM 学习机的图象插值算法得到的插值效果优于所比较的其他插值算法.值得一提的是,实验中的 SVM 学习机采用的训练样本和测试样本是完全不重合的,而表 1 和表 2 中两种神经网络得到的数据^[8]是在训练样本和测试样本重合的情况下得到的.由此也可以看出,基于支持向量机的图象插值算法具有很好的

推广性能.基于这样的考虑,将 SVM 学习机引入到错误隐匿问题中.

3 基于支持向量机的错误隐匿策略

视频解码端的错误隐匿策略的核心思想,是利用丢失的图象块与相邻图象块之间的相关性来尽可能恢复丢失图象块的信息,从而提高图象的感观质量.图 4 所示为原始 Lena 图象以及发生了大量 DCT 块(8×8)丢失的错误图象.和文献[5]一样,首先考虑单个图象块间隔丢失情况下的错误隐匿,然后再考虑连续图象块丢失的情况.对每一个丢失的像素块,利用它周围的 8 个相邻像素块的像素进行空间插值.实验中共设计输入输出向量如图 5 所示的 3 个 50 维输入的 SVM 学习机,利用训练样本集进行训练,这里训练样本仍然是随机地从标准图象库中抽取.



(a) 原始 Lena 图象

(b) 发生块丢失的错误图象

图 4

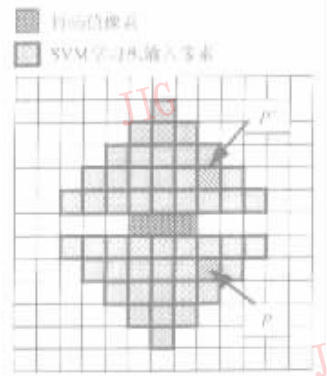


图 5 用于错误隐匿的 SVM 学习机的输入输出像素

对丢失的像素块进行插值,其算法如下:

(1) 如图 6 所示,将像素块按空间位置分为 NE,SE,SW,NW 4 个部分,对每一部分计算其相邻像素块的灰度平均值

$$G_{\phi_r} = \frac{1}{N_{\phi_r}} \sum_{P_i \in \phi_r} P_i \quad (11)$$

其中, $r \in \{NE, SE, SW, NW\}$, ϕ_r 表示其相邻像素块



图 6 丢失图象块邻域示意图

集合, N_{ϕ_r} 表示集合 ϕ_r 的像素个数, P_i 表示像素灰度值.

(2) 首先对图象块的首行和末行的像素值用 SVM 学习机进行插值(如图 7 所示).输入数据按以下情况确定:

若 SVM 学习机所取的输入像素 P_i 没有丢失,则取输入数据为

$$I_i = P_i \quad (12)$$

若 SVM 学习机所取的输入像素丢失,并且没有进行插值估计,则取

$$I_i = (G_{\phi_r} + P'_i)/2 \quad (13)$$

其中, P'_i 是 P_i 在输入样本集中的对称像素点,见图 5 所示.

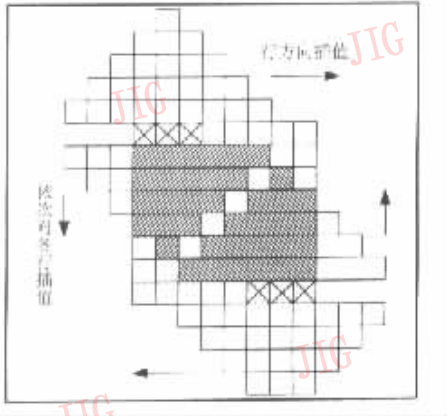


图 7 基于 SVM 的错误隐匿过程

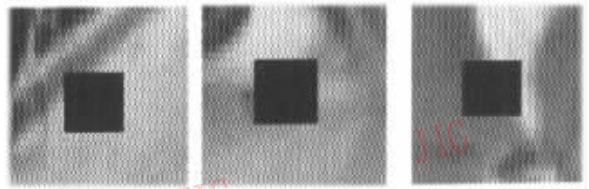
若 SVM 学习机所取的输入像素丢失,并且已经进行了插值估计,则取

$$I_i = P_i^{SVM} \quad (14)$$

其中, P_i^{SVM} 是该像素点的基于 SVM 学习机的估计值.



(a) 错误隐匿后的 Lena 图象



(b) 部分错误的局部放大图象



(c) 相应的基于 SVM 的恢复图象

图 8

表 3 图象发生不连续块丢失时不同错误隐匿算法的 MSE 比较

单位: dB

错误隐匿方法	Lena	Peppers	Maskerad
DCT1	43.56	40.74	86.81
DCT2	49.90	50.65	98.50
DCT3	42.29	40.09	86.76
DCT4	36.40	37.20	82.83
Max-sm	19.37	14.63	53.12
Neural Pred.	18.8	12.6	50.0
Neural Pred. +LP filter	13.98	11.14	38.90
SVM	7.13	9.12	33.87
SVM+refinement	7.02	8.87	32.25

当视频流的每一个宏块包含多个 DCT 块,并且发生错误时,就会发生图象块连续丢失的情况,在

(3)重复步骤 2,直到丢失图象块完全被恢复为止.

(4)为了达到更好的插值效果,当图象块恢复完毕以后,可以重复步骤 1~3,对估计值进行调整.在调整过程中,由于输入数据更加接近像素的实际灰度值,因此能够进一步提高插值的准确性.

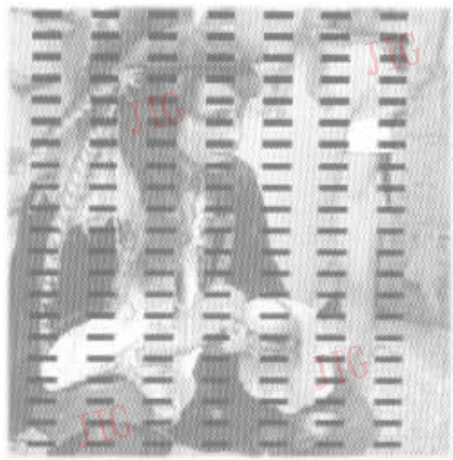
按照以上的算法,对发生均匀 DCT 块丢失的 Lena,Peppers 和 Maskerad 3 幅图象进行错误隐匿.图 8 所示为图 4 中的错误图象经过错误隐匿后的恢复图象,以及部分丢失图象块和相应的插值恢复后的局部图象放大图.对于错误隐匿后的图象通过计算图象的平方误差和(MSE)来评价错误隐匿的效果,并与文献[5]中采用不同方法得到的错误隐匿结果相比较,其结果如表 3 所示.其中 DCT1、DCT2、DCT3 和 DCT4 4 种方法是频域的插值方法,Max-sm 是一种基于图象平滑性的空域插值算法[9].

这种情况下仍然可用 SVM 学习机进行错误隐匿,只是对于每一个连续丢失的图象块增加了插值的次数.Maskerad 图象发生连续块丢失错误以及错误隐匿的结果如图 9 所示,图 10 所示为部分局部错误隐匿细节.与文献[5]提出的频域神经网络插值法进行比较的结果如表 4 所示.

表 4 图象发生连续块丢失时不同错误隐匿算法的 MSE 比较

单位: dB

Image	Neural Pred.	Neural+smooth	SVM	SVM+refinement
Lena	26.93	21.65	21.36	19.46
Peppers	18.63	15.78	16.47	15.54
Maskerad	77.66	63.06	46.45	44.61



(a) 发生连续丢失的 Maskerad 图象



(b) 基于 SVM 进行错误隐匿后的恢复图象

图 9



(a) Maskerad 错误图象的部分局部放大图象



(b) 相应的基于 SVM 的局部恢复放大图象

图 10

从以上的结果可以看出,基于 SVM 学习机的错误隐匿策略能够得到比传统的神经网络预测插值更好的错误隐匿效果. 另外,文献[5]中所设计的神经网络需要依赖于专门为其设计的特定训练样本,才能达到比较好的效果,而本文中 SVM 学习机的训练数据是随机地取自标准图象库. 由此可见,基于 SVM 学习机的错误隐匿策略在推广性能方面也具有一定的优势.

4 结 语

实验结果表明,基于支持向量机的图象插值和错误隐匿算法不仅能够得到比传统的神经网络学习算法更优越的结果,而且在算法的推广性上也要优

于基于神经网络的方法. 这主要是因为从统计学习理论上讲,支持向量机是一种能够从理论上保证其具有最优推广性能的统计学习方法,而神经网络虽然能够在一些问题上得到比较好的学习效果,却无法从理论上保证其推广性能;从另一方面讲,基于支持向量机和神经网络的算法得到的实验结果都要好于基于常用的插值算子得到的结果,这说明采用机器学习的方法得到的高维非线性插值算子能够更好地抓住图象相邻像素之间的关系,也说明了机器学习方法的有效性.

对于机器学习方法而言,一个重要的前提就是训练集和测试集的相关性. 在实验中,训练图象和测试图象都是一般的自然图象,而自然图象之间都存在着某种程度的相关性,因此该算法才能够得到比

较好的效果,但是,如果使用相关性更强的图象,例如训练和测试都使用人脸图象,该算法是否能够得到更好的效果,或者使用相关性更弱的图象,算法的效果是否会变差,这些问题需要做进一步的探讨和分析。

算法成功的应用说明,统计学习的最新方法能够很好地应用到图象处理和计算机视觉问题中.随着目前统计学习理论不断发展,相信更多的图象处理问题可以通过统计学习的方法得到更好的解决。

参 考 文 献

- 1 Koenen R. Overview of the MPEG-4 standard (Maui Version) [S]. ISO/IEC JTC1/SC29/WG11 N3156, Maui, Dec. 1999.
- 2 Kozintsev I, Ramchandran K. Robust image transmission over energy-constrained time-varying channels using multiresolution joint source-channel coding [J]. IEEE Transactions on Signal Processing, 1998,46(4):1012~1026.
- 3 Wang Y, Zhu Q F. Error control and concealment for video communication: A review[J]. Proc. of the IEEE. , 1998,86(5): 974~997.
- 4 Robie D L, Mersereau R M. The use of hough transforms in spatial error concealment[A]. In:Proc. of 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'00)[C]. Istanbul, Turkey, 2000;2131~2134.

- 5 De Natale F G B, Perra C, Vernazza G. DCT information recovery of erroneous image blocks by a neural predictor[J]. IEEE Journal On Selected Areas In Communications, 2000,18(6): 1111~1121.
- 6 Constantine P, Tomaso P. A Pattern classification approach to dynamical object detection[A]. In:Proceedings of ICCV [C], Kerkyra, Greece, 1999;1223~1228.
- 7 张学工. 关于统计学习理论和支持向量机[J]. 自动化学报, 2000,26(1):32~42.
- 8 Plaziac N. Image interpolation using neural networks[J]. IEEE Trans. On Image Processing, 1999,8(11):1647~1651.
- 9 Wang Y, Zhu Q F, Shaw L. Maximally smooth image recovery in transform coding [J]. IEEE Trans. On Communication, 1993,41(10):1544~1551.



王 珏 1977年生,现为清华大学自动化系硕士研究生.主要研究方向为计算机视觉,图象、视频处理以及模式识别。



季 梁 1940年生,现为清华大学自动化系教授.主要研究方向为中医药现代化研究、生物信息学、图象处理和模式识别。