

# 隐马尔可夫模型的多序列比对研究

罗泽举<sup>1,2</sup>, 宋丽红<sup>3</sup>

LUO Ze-ju<sup>1,2</sup>, SONG Li-hong<sup>3</sup>

1.重庆工商大学 长江上游经济研究中心,重庆 400067

2.重庆工商大学 计算机科学与信息工程学院,重庆 400067

3.重庆工商大学 经济管理实验教学中心,重庆 400067

1.Research Center of the Economy of the Upper Reaches of Yangtze River,Chongqing Technology and Business University,Chongqing 400067,China

2.School of Computer Science & Information Engineering,Chongqing Technology and Business University,Chongqing 400067,China

3.Economics and Management Center,Chongqing Technology and Business University,Chongqing 400067,China

E-mail:luozeju@126.com

LUO Ze-ju,SONG Li-hong.Multiple sequence analysis of hidden Markov model.Computer Engineering and Applications, 2010,46(7):171-174.

**Abstract:** A new multiple sequence alignment about Hidden Markov Models(HMMs) is researched,using the conservative feature of  $L$  value and consensus sequence,by increasing frequency factor,traditional HMMs learning algorithm is improved.Experiment indicates that not only the stability of the model is improved,but also a average improvement of 3.3% is achieved for protein family recognition by comparing the new algorithm with the traditional one.

**Key words:** hidden markov models;multiple sequence analysis;protein recognition

**摘要:**研究一种关于隐马尔可夫模型的多序列比对,利用值和特征序列的保守性,通过增加频率因子,改进传统隐马尔可夫模型算法的不足。实验表明,新算法不但提高了模型的稳定性,而且应用于蛋白质家族识别,平均识别率比传统隐马尔可夫算法提高了3.3个百分点。

**关键词:**隐马尔可夫模型;多序列分析;蛋白质识别

**DOI:**10.3778/j.issn.1002-8331.2010.07.052 **文章编号:**1002-8331(2010)07-0171-04 **文献标识码:**A **中图分类号:**TP391;TN957.52

## 1 引言

随着21世纪生命科学时代的来临,生命的进化已成为人类研究的一个热点。为了探索进化的轨迹,就必须研究其家族的同源性,而一个基因家族往往由若干条DNA序列组成,因此只有弄清多个序列之间的相互关系,才能真正揭示基因家族的进化特征。如果能知道蛋白质的结构信息,它将比序列信息更适合解释实际发生的遗传事件,因为蛋白质结构除了提供序列信息外,还描绘了分子间相互作用等。然而,遗憾的是,目前得到的蛋白质三维结构数据非常有限,人们还无法弄清许许多多的蛋白质结构,这就迫使人们不得不利用大量已知的一维序列信息进行研究。多序列比对分析正是在这样的背景下产生的,它提供了人们通过研究序列的相似性来探索同源性的重要方法。通过多重序列比对,发现代表它们进化的特征序列(consensus sequence),于是基因家族成员的进化可以看成是由这

个特征序列经过若干代衍变的结果<sup>[1]</sup>。

多重序列分析是一个非常困难的问题,涉及许多模型的选择,Carillo和Lipman引入了基于两两最优化比对分数的多重序列比对方法,并得到了广泛应用,但是这种方法对于计算时间和空间的耗费极大,被证明是NP难题<sup>[2]</sup>。许多研究者利用启发式和近似算法改进了比对分数算法<sup>[3]</sup>,包括Feng和Doolittle的Clustal算法,但这种算法是利用进化树思想先进行两两比对,再根据相似性进行分组比对,直到最终得到比对结果,因此其时间复杂度仍然很高<sup>[4]</sup>。由于一个基因家族的特征序列非常保守,家族的进化可以认为是这个特征序列经过一系列插入、替代、删除的结果,这个过程正好可以用隐马尔可夫模型(Hidden Markov Models, HMMS)来描述。

由于HMMS模型能节省大量时间和空间,因而越来越引起计算生物学的关注。隐马尔可夫理论最初是由Baum及他的

**基金项目:**国家“十一五”科技支撑计划重大项目资助(the National Great Project of Scientific and Technical Supporting Programs Funded by Ministry of Science & Technology of China During the 11th Five-year Plan. No.2006BAJ05A06);重庆市科委自然科学基金(No.2007BB2205);重庆市科委重点攻关项目(No.2008AC0043)。

**作者简介:**罗泽举(1965-),博士,主要研究方向:机器学习与模式识别,生物信息学;宋丽红(1969-),实验师,主要从事机器学习、数据仓库等研究。

**收稿日期:**2008-09-17 **修回日期:**2008-12-15

同事于 60 年代末 70 年代初提出,并开始用于语音识别<sup>[5]</sup>。最早用于计算生物学是于 80 年代末 90 年代初,目前已经用于 DNA 模型构建,蛋白质二级结构预测,基因预测,横跨膜蛋白识别,其中应用最为普遍的是 Krogh 等人的基于 profile 家族共同特征提取的蛋白质序列分析<sup>[6-7]</sup>。

提出一种基于隐马尔可夫模型的多序列比对算法,利用  $L$  值和特征序列的保守性,通过增加频率因子,改进传统隐马尔可夫模型算法的不足。实验表明,用于蛋白质家族的识别,新算法比传统算法的识别率提高了 3.3 个百分点。

## 2 多重序列比对的数学模型

**定义 1** 三元组  $\Omega=(\Sigma, S, G)$  称为一个多重序列比对序列组,其中  $\Sigma$  是比对序列字母集,  $S=\{S_1, S_2, \dots, S_k\}$  是比对序列的集合,其中  $S_i (i=1, 2, \dots, k)$  是以集合的形式代表一条序列,  $G=(g_{ij})$  是一个比对矩阵,其元素是  $\Sigma$  中的元素。

例如,若对 DNA,  $\Sigma=\{A, T, G, C, -\}$  (其中“-”表示空位或删除状态),对 RNA,  $\Sigma=\{A, U, G, C, -\}$ ,若针对蛋白质,  $\Sigma$  是 20 种氨基酸字母和“-”的集合,即  $\Sigma=\{G, A, L, M, F, W, K, S, N, D, P, V, I, C, Y, H, R, T, Q, E, -\}$ ;对于模型中  $S_i (i=1, 2, \dots, k)$ ,例如  $S_1=\{A, G, A, C, G, T, C\}$ ,代表序列 AGACGTC;对于  $G=(g_{ij})$ ,例如如图 1  $G$  是有 5 个序列的比对。

$$G = \begin{pmatrix} S_1 & A & A & - & T & C \\ S_2 & A & T & G & - & C \\ S_3 & C & T & G & T & G \\ S_4 & C & T & G & - & C \\ S_5 & A & C & - & T & C \end{pmatrix}$$

图 1 一个有 5 条序列的多重序列比对矩阵  $G$

## 3 特征序列

序列的进化可以看成是一个特征序列经过若干代衍变的结果,这个特征序列描绘了这个多重序列共同进化的本质特征,算法的关键就是寻找和这个特征序列相同的匹配。这里利用子序列(Subsequence)方法来描述特征序列,就是从多重序列比对中找出每列出现字符最多的元素,例如图 1 的  $S_1, S_2, \dots, S_5$  的特征序列是:ATGTC。如果多重序列比对里每列元素中出现字符一样多,则随机取其中一个元素。

对特征序列进行研究具有重要意义,首先,利用特征序列可以对一个序列进行数据库搜索,以寻找它所在家族;其次,可以比较不同家族的进化关系;另外,它是构建隐马尔可夫模型的理论基础。如果所进行比对的序列是具有生物学进化意义的相关序列家族,那么每条序列可以看成是这条特征序列经过插入、删除、匹配而进化的结果。

## 4 隐马尔可夫模型

### 4.1 隐马尔可夫模型的定义

**定义 2** 模型  $\lambda=(S, \Sigma, A, B, \pi)$  称为隐马尔可夫模型。其中  $S=\{S_1, S_2, \dots, S_N\}$  为状态集合,  $\Sigma=\{O_1, O_2, \dots, O_M\}$  是观察符号或观察向量的集合,  $A=(a_{ij})$  为状态转移概率矩阵,记为  $a_{ij}=P(q_{i+1}=S_j | q_i=S_i)$ ,  $1 \leq i, j \leq N$ ;  $B=(b_j(k))$  表示在状态  $S_j$  时产生观察符号  $v_k \in \Sigma$  的离散概率值或连续概率密度矩阵。其中  $b_j(k)=P(v_k | q_i=S_j)$ ,  $1 \leq j \leq N, 1 \leq k \leq M$ ;  $\pi=(\pi_j)$  是初始状态分布矩阵,  $\pi_j=P(q_1=S_j)$ ,  $1 \leq j \leq N, \pi_j \geq 0, \sum \pi_j=1$ 。

上述定义中当观察符号  $v_k$  是离散符号时,叫离散马尔可夫模型;当  $v_k$  是连续矢量时,叫连续马尔可夫模型,其关键参数是  $A, B, \pi$ ,故模型一般简记为  $\lambda=(A, B, \pi)$ 。

**定义 3** (logarithm likelihood Value,  $L$  值) 设由模型  $\lambda$  产生观察序列  $O$  的概率为  $P(O|\lambda)$ ,其自然对数值  $L=\log P(O|\lambda)=\ln P(O|\lambda)$  称为  $L$  值。

当用 HMM 模型进行识别时,首先用 EM 算法对参数  $A, B, \pi$  进行重估,然后利用新参数来计算产生当前序列概率的对数值(logarithm likelihood,  $L$  值),再根据  $L$  值来识别对象所属的类。设当前模型为  $\lambda=(A, B, \pi)$ ,训练重估模型参数后的模型为  $\lambda^*=(A^*, B^*, \pi^*)$ ,则有:

**定理 1** 设  $X$  和  $Y$  是两个随机变量,在模型  $\lambda$  下具有概率分布  $P(X)$  和  $P(Y)$ ,  $P'(X)$  和  $P'(Y)$  是在模型  $\lambda^*$  下的概率分布,若  $P'(X|Y)>0, P(X|Y)>0, P'(X, Y)>0, P(X, Y)>0$  且  $\sum_x P(X|Y) \times$

$$\ln \frac{P'(X, Y)}{P(X, Y)} \geq 0, \text{ 则:}$$

$$P'(Y) \geq P(Y) \quad (1)$$

**证明** 因为  $P'(X|Y)>0, P(X|Y)>0, P'(X, Y)>0, P(X, Y)>0$ ,  $f(x)=\ln x$  在  $(0, +\infty)$  有:  $\ln x \leq x-1$ , 则:

$$\sum_x P(X|Y) \ln \frac{P'(X|Y)}{P(X|Y)} \leq \sum_x P(X|Y) [ \frac{P'(X|Y)}{P(X|Y)} - 1 ] =$$

$$\sum_x P'(X|Y) - \sum_x P(X|Y) = 1 - 1 = 0$$

$$\text{则 } \sum_x P(X|Y) \ln \frac{P(X|Y)}{P'(X|Y)} = - \sum_x P(X|Y) \ln \frac{P'(X|Y)}{P(X|Y)} \geq 0, \text{ 因}$$

$$\sum_x P(X|Y) = 1, \text{ 则}$$

$$\ln P'(Y) - \ln P(Y) = \sum_x P(X|Y) \ln P'(Y) - \sum_x P(X|Y) \ln P(Y) =$$

$$\sum_x P(X|Y) \ln \frac{P'(X, Y)}{P(X, Y)} - \sum_x P(X|Y) \ln \frac{P(X, Y)}{P(X|Y)} =$$

$$\sum_x P(X|Y) \ln \frac{P'(X, Y)}{P'(X|Y)} + \sum_x P(X|Y) \ln \frac{P(X, Y)}{P(X|Y)} =$$

$$\sum_x P(X|Y) \ln \frac{P'(X, Y)}{P(X, Y)} + \sum_x P(X|Y) \ln \frac{P(X|Y)}{P'(X|Y)}$$

$$\left( \because \sum_x P(X|Y) \ln \frac{P'(X|Y)}{P(X|Y)} \geq 0 \right) \geq \sum_x P(X|Y) \ln \frac{P(X, Y)}{P(X, Y)} \geq 0$$

故  $P'(Y) \geq P(Y)$ 。

这个定理表明了如果能寻找到一个模型  $\lambda^*$ , 使得相关不等式成立,则在模型  $\lambda^*$  下生成向量  $Y$  的可能性大于等于在模型  $\lambda$  下生成  $Y$  的可能性。因此有必要对模型的参数进行重新训练和进行估计。

### 4.2 隐马尔可夫模型训练算法的改进

训练问题是 HMMS 三个问题中最难估计的。目的是要根据已有模型参数  $\lambda=(S, \Sigma, A, B, \pi)$  来重新估计  $A, B, \pi$ ,传统算法通过巧妙定义向前向后变量  $\xi_i(i, j)$  进行重估<sup>[7]</sup>:

设观察序列是  $O=O_1^* O_2^* \dots O_k^*$ , 相应的状态序列为:  $Q=q_1 q_2 \dots q_k$ ; 定义变量:

$$\alpha_i(i) = P(O_1^* O_2^* \dots O_i^*, q_i = S_i | \lambda) \quad (2)$$

$$\beta_i(i)=P(O_{i+1}^*O_{i+2}^*\cdots O_n^*,q_i=S_i,\lambda) \quad (3)$$

$$\xi_i(i,j)=P(q_i=S_i,q_{i+1}=S_j|O^*,\lambda)=\frac{\alpha_i(i)a_{ij}b_j(O_{i+1}^*)\beta_{i+1}(j)}{P(O^*|\lambda)}$$

$$=\frac{\alpha_i(i)a_{ij}b_j(O_{i+1}^*)\beta_{i+1}(j)}{\sum_{i=1}^N\sum_{j=1}^N\alpha_i(i)a_{ij}b_j(O_{i+1}^*)\beta_{i+1}(j)} \quad (4)$$

$$\gamma_i(i)=P(q_i=S_i|O^*,\lambda)=\sum_{j=1}^N\xi_i(i,j) \quad (5)$$

从而得出三个参数的估计式:

$$\pi_i^*=\gamma_i(i) \quad (6)$$

$$a_{ij}^*=\frac{\sum_{i=1}^{k-1}\xi_i(i,j)}{\sum_{i=1}^k\gamma_i(i)} \quad (7)$$

$$b_j^*(k)=\frac{\sum_{i=1}^k\gamma_i(j)}{\sum_{i=1}^k\gamma_i(j)} \quad (8)$$

但是上述算法的弊端是建立在单个训练序列算法基础上的,只包含了小量的训练状态,不考虑联合概率和频率因子,因此模型的可信程度受到限制。应当增加训练序列,因为当训练序列增多时,由于样本量的增加,训练算法将更稳定,置信度也得到相应的提高,通过联合概率和增加多重观察序列频率因子来修正上述模型参数  $a_{ij}^*$  和  $b_j^*(k)$ ,改进算法如下:

将训练序列增加到  $k$  个,设观察序列为:

$$O=[O^1,O^2,\dots,O^k] \quad (9)$$

其中  $O^l=[O_{(1)}^l,O_{(2)}^l,\dots,O_{(M)}^l],l=1,2,\dots,k$  是多重序列中的第  $l$  个观察序列,设各观察序列间相互独立,定义联合概率如下:

$$P(O|\lambda)=\prod_{l=1}^kP(O^l|\lambda)=\prod_{l=1}^k p_l \quad (10)$$

考虑各条序列的训练权重,增加频率因子  $p_l \prod_{i=1}^{M-1} \alpha_i^l(i)$ ,修正参数的估计式为:

$$a_{ij}^*=\frac{\sum_{l=1}^k p_l \sum_{i=1}^{M-1} \alpha_i^l(i)a_{ij}b(O_{i+1}^l)\beta_{i+1}^l(j)}{\sum_{l=1}^k p_l \sum_{i=1}^{M-1} \alpha_i^l(i)\beta_i^l(j)} \quad (11)$$

$$b_j^*(r)=\frac{\sum_{l=1}^k p_l \sum_{i=1}^{M-1} \alpha_i^l(i)\beta_i^l(j)}{\sum_{l=1}^k p_l \sum_{i=1}^{M-1} \alpha_i^l(i)\beta_i^l(j)} \quad (12)$$

由于  $\pi_1^*=1,\pi_i^*=0,i \neq 1$ ,故对于初始态  $\pi_i^*$  不用进行重新估计。经实验证明,用联合概率和增加频率因子的方法得到的模型训练稳定且识别效果比传统算法显著提高。

## 5 实验结果

### 5.1 训练模型的建立

Pfam 数据库<sup>[8]</sup>是以序列谱为基础,用 profile 隐马尔可夫模

型方法组建的蛋白质家族多重序列比对数据库,可用于识别一未知序列是否属于某个已知的蛋白质家族,甚至即使其同源性非常微弱也能识别。不同于标准的双序列比对数据库搜索(如 BLAST,FASTA),Pafm 数据库注重于多重蛋白质域搜索。从与 Pafm 数据库相连的 SCOP 蛋白质数据库中选择了类为隔膜和细胞表面蛋白,折叠为单个横跨膜螺旋,超族和族为光合体系 II 中心组的已知结构蛋白质序列 100 条作为训练序列,先进行多重序列比对,后建立多重序列的 HMMS 模型,用改进算法进行计算,经过 200 个循环的训练,得到表 1 的 HMMS 模型。

表 1 HMMS 的训练模型参数

模型长度			L 值		
插入节点	删除节点	匹配节点	最高	最低	平均
100	100	100	521.8	363.5	438.2

表 1 中的训练分数是用改进算法后得到的分数,分数高表明序列的保守性越强。

图 2 是进行上述多重序列训练时得到的最大值、最小值和平均值的三条训练曲线,可见当训练接近 200 个循环时已经非常稳定,说明提出的算法是有很有效的和稳定的。

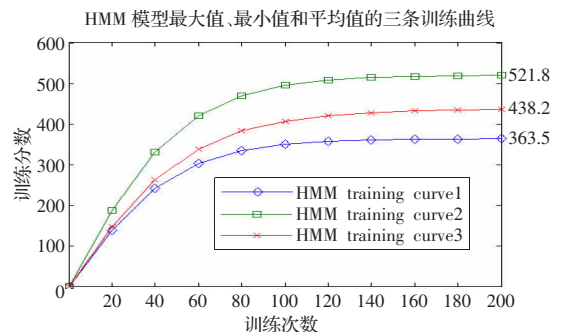


图 2 HMM 的 L 值(训练分值)训练曲线图

### 5.2 各类蛋白质家族的 L 值分布

选取三个蛋白质家族作为样本,分别是横跨膜蛋白、转移隔膜蛋白和球菌蛋白。并为每一类建立一个多重序列,考虑训练分数(表 2)。从表 2 可以看出这三个蛋白质家族各有不同的隐马尔可夫 L 值,它们代表不同的家族类,三类不同的蛋白有不同的映射谱,各类残基序列的训练值相当保守。

表 2 多类蛋白质家族的 HMMS 分类

家族	残基序列	L 值
横跨膜蛋白	1rzh SEQRES	363.5
	1rzhL SEQRES	521.9
	1rzhM SEQRES	547.1
转移隔膜蛋白	1ppjI SEQRES	120.9
	1ppjO SEQRES	100.5
	1ppjJ SEQRES	131.8
球菌蛋白	1rh5B SEQRES	4.1
	1rh5C SEQRES	19.0
	1rh5A SEQRES	-19.5

### 5.3 识别结果

在 scop 数据库中分别选取横跨膜蛋白 78 条、转移隔膜蛋白 78 条和球菌蛋白 60 条,结合 HMM 训练分值并考虑 L 值的保守性,得到表 3 所示的识别结果,由于传统隐马尔可夫模型算法只从两两比对入手,得到进化树进行比对,不考虑多序列



训练权重,不但耗时,而且由于只根据单个序列进行训练,置信度也低。提出的算法考虑了具有一致序列的多重序列,增加了多序列频率因子,直接根据 HMMS 的  $L$  值分数进行分析,提高了训练的稳定性和可靠性。从表 3 可见新算法比传统方法平均提高了 3.3 个百分点,本类和非本类的区分程度较高,因而训练模型是有很有效的。

表 3 蛋白质家族的识别结果

家族	训练序列	测试序列	正确识别率/(%)	
			传统 HMM 算法	新算法
横跨膜蛋白	39	39	92.3	94.9
转移隔膜蛋白	39	39	89.7	92.3
横跨膜蛋白	39	39	92.3	94.9
球菌蛋白	30	30	90.0	93.3
转移隔膜蛋白	39	39	92.3	94.9
球菌蛋白	30	30	86.7	93.3

## 6 讨论

提出一种训练隐马尔可夫模型参数的新算法,利用特征序列的一致性,增加了训练的序列数,用联合概率和频率因子修正了模型参数,经过实验证明训练后的模型具有更好的稳定性和识别效果。隐马尔可夫模型由于能对基因序列符号进行编解码,其插入、删除、匹配正好模拟了基因的进化过程,使它在生命科学领域倍受欢迎。目前国际上已经建立著名的基于隐马尔可夫模型蛋白质家族识别数据库(基于多重序列比对)Pfam

(<http://www.sanger.ac.uk/>), 已逐渐成为全球研究人员探索生命进化的共享数据库。

## 参考文献:

- [1] Lazarl M N, Wacker M, Schegg B, et al. The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation[J]. *Glycobiology*, 2005, 15(4): 361-367.
- [2] Eld D G, Cien E. Methods for multiple sequence alignment with guaranteed error bounds[J]. *Bulletin of Mathematical Biology*, 1993, 55: 141-154.
- [3] Landan G, Graur D. Local reliability measures from sets of co-optimal multiple sequence alignments[C]// *Pacific Symposium on Bio-computing*, 2008, 13: 15-24.
- [4] Mukawa S, Goto C. In vivo characterization of two granuloviruses in larvae of *Mythimna separata* (Lepidoptera: Noctuidae)[J]. *Journal of General Virology*, 2008, 89(4): 915-921.
- [5] Baum L E, Petrie T. Statistical inference for probabilistic functions of finite state Markov chains[J]. *Annmath Stat*, 1996, 37: 1554-1563.
- [6] Krogh A. Hidden Markov models in computational biology applications to protein modeling[J]. *J Mol Biol*, 1994, 235: 1501-1531.
- [7] Knapp K, Chen Y P. An evaluation of contemporary hidden Markov model gene finders with a predicted exon taxonomy[J]. *Nucleic Acids Research*, 2007, 35(1): 317-324.
- [8] University of Oxford and Mahidol University. 蛋白质结构分类数据库[D/OL]. [2008-08-30]. <http://www.sanger.ac.uk/>.
- [9] 报, 1992, 18(2): 233-237.
- [6] Martinez A, Kak A. PCA versus LDA[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(2): 228-233.
- [7] Pentland A, Moghaddam B. Starner view-based and modular eigenspaces for face recognition[C]// *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 1994: 84-91.
- [8] Hong Z Q. Algebraic feature extraction of image recognition[J]. *Pattern Recognition*, 1991, 24(3): 211-219.
- [9] 高全学, 梁彦, 潘泉, 等. SVD 用于人脸识别存在的问题及解决方法[J]. *中国图象图形学报*, 2006, 12(11): 1785-1791.
- [10] 杨欣, 费树岷, 陈丽娟. 基于类矩阵和特征融合的加权自适应人脸识别[J]. *中国图象图形学报*, 2008, 5(13): 931-936.
- [11] 王蕴红, 谭铁牛, 朱勇. 基于奇异值分解和数据融合的脸像鉴别[J]. *计算机学报*, 2000, 23(6): 649-653.
- [12] 王虹, 朱双燕. 基于整体与局部奇异值分解的人脸识别[J]. *计算机工程与应用*, 2007, 43(22): 235-237.
- [9] 邢藏菊, 曲延锋, 徐健, 等. 一种用于抑制椒盐噪声的多窗口中值滤波器[J]. *电子与信息学报*, 2002, 24(12): 1912-1916.
- [10] Hwang H, Haddad R A. Adaptive median filters: New algorithms and results[J]. *IEEE Transaction on Image Processing*, 1995, 4(4): 499-502.
- [11] 关新平, 赵立兴, 唐英干. 图像去噪混合滤波方法[J]. *中国图象图形学报*, 2005, 10(3): 322-337.
- [12] 蔡靖, 杨晋生, 丁润涛. 分类均值-加权中值混合滤波器[J]. *天津大学学报*, 1999, 32(5): 551-554.
- [13] 张政, 张宇. 基于均值操作的快速自适应滤波器[J]. *中国图象图形学报*, 2000, 5(6): 530-534.
- [14] 王振华, 窦丽华, 陈杰. 一种尺度自适应调整的高斯滤波器设计方法[J]. *光学技术*, 2007, 32(3): 395-397.

(上接 164 页)

## 参考文献:

- [1] Turk M, Pentland A. Eigenfaces for recognition[J]. *Journal of Cognitive Neuroscience*, 1991, 3(1): 71-86.
- [2] Belhumeur V, Hespanda J, Kiregeman D. Eigenfaces vs. fisherfaces: recognition using class specific linear projection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(7): 711-720.
- [3] Bartlett M S, Movellan J R, Sejnowski T J. Face recognition by independent component analysis[J]. *IEEE Transactions on Neural Networks*, 2002, 13(6): 1450-1464.
- [4] Kim K I, Jung K, Kim H J. Face recognition using kernel principal component analysis[J]. *IEEE Signal Processing Letters*, 2002, 9(2): 40-42.
- [5] 洪子泉, 杨静宇. 用于图像识别的图像代数特征抽取[J]. *自动化学*

(上接 170 页)

- [3] 张宇, 王希勤. 自适应中心加权的改进均值滤波算法[J]. *清华大学学报: 自然科学版*, 1999, 39(9): 76-78.
- [4] 张恒, 雷志辉, 丁晓华. 一种改进的中值滤波算法[J]. *中国图象图形学报*, 2004, 9(4): 408-411.
- [5] 张旭明, 徐滨士, 董世运, 等. 自适应中值-加权均值混合滤波器[J]. *光学技术*, 2004, 30(6): 652-655.
- [6] 邢卓异, 朱齐丹, 定宇. 自适应-模糊滤波算法[J]. *哈尔滨工程大学学报*, 2006, 27(6): 844-847.
- [7] 周刚, 贾振红, 覃锡忠. 一种新的图像去噪混合滤波方法[J]. *激光杂志*, 2007, 28(1): 57-59.
- [8] 胡浩, 王明照, 杨杰. 自适应模糊加权均值滤波器[J]. *系统工程与电子技术*, 2002, 24(2): 15-17.