

基于改进 K-L 变换的特征提取技术*

黄晓斌, 万建伟, 王 展

(国防科技大学 电子科学与工程学院, 湖南 长沙 410073)

摘 要 在模式识别中,为解决样本增加时,反复使用传统 K-L 变换进行特征提取耗时多的缺点,提出了一种改进的 K-L 变换方法,该方法利用快速递推算方法来计算协方差矩阵,并使用矩阵扰动理论来求解协方差矩阵的特征值和特征向量。仿真试验表明该方法在确保计算精度的前提下,大大减少了运算时间。

关键词 特征提取 协方差矩阵 矩阵扰动

中图分类号 :TP391.4 文献标识码 :A

An Improved K-L Transform for Feature Extraction

HUANG Xiao-bin, WAN Jian-wei, WANG Zhan

(College of Electronic Science and Engineering, National Univ. of Defense Technology, Changsha 410073, China)

Abstract In order to surmount the disadvantage that much time is consumed for feature extraction with traditional K-L transform, an improved K-L transform method in pattern recognition is presented. This improved method gets covariance matrix with recursion algorithm and computes the eigenvalues and eigenvectors of the covariance matrix with matrix perturbation theory. The simulation results show that compared with the traditional method, the consuming time with this method is sharply decreased, and the computing precision is kept.

Key words feature extraction; covariance matrix; matrix perturbation

对试验样本进行特征提取是试验数据事后处理的一项重要工作,其中 K-L 变换是特征提取常用的方法,该方法有两个重要并且也是十分耗时的步骤^[1] (1)计算样本的协方差矩阵 (2)求解样本协方差矩阵的特征值和特征向量。由于样本容量是随时间推移而不断增加的,每次增加样本都需要重复上述的工作,由此可见计算量很大,为解决这一问题,本文对 K-L 变换中的上述两个步骤进行了改进。

1 问题的提出

尽管样本的增加情况是预先无法获知的,但都可以视样本是逐个增加的。为使分析的问题清晰化,根据样本逐个递增情况,将待解决的问题用以下的数学模型来描述。

在给出模型前,先设定一些数学记号。设有数据矩阵 $\begin{cases} X^{(N)} = [x_1, x_2, \dots, x_l, \dots, x_N] \\ X^{(N+1)} = [x_1, x_2, \dots, x_l, \dots, x_N, x_{N+1}] \end{cases}$ 其中,

$X^{(N)}$ 为原始数据矩阵, $X^{(N+1)}$ 为增加一个样本后的数据矩阵, $x_l = (x_{l,1}, x_{l,2}, \dots, x_{l,k}, \dots, x_{l,M})^T$ 。并令 $\mu^{(N)} = (\mu_1^{(N)}, \mu_2^{(N)}, \dots, \mu_k^{(N)}, \dots, \mu_M^{(N)})^T$ 为 $X^{(N)}$ 的均值向量, $\Sigma^{(N)}$ 为 $X^{(N)}$ 的协方差矩阵, $\mu^{(N+1)} = (\mu_1^{(N+1)}, \mu_2^{(N+1)}, \dots, \mu_k^{(N+1)}, \dots, \mu_M^{(N+1)})^T$ 为 $X^{(N+1)}$ 的均值向量, $\Sigma^{(N+1)}$ 为 $X^{(N+1)}$ 的协方差矩阵。

数学模型如下:

已知 $X^{(N)}$ 、 $X^{(N+1)}$ 、 $\Sigma^{(N)}$, 以及 $\Sigma^{(N)}$ 的特征值 $\{\lambda_k^{(N)}\}_{k=1}^M$ 和对应的特征向量 $\{u_k^{(N)}\}_{k=1}^M$;

求解 $\Sigma^{(N+1)}$, 以及它的特征值 $\{\lambda_k^{(N+1)}\}_{k=1}^M$ 和对应的特征向量 $\{u_k^{(N+1)}\}_{k=1}^M$ 。

对于以上问题,传统的 K-L 变换在计算 $\Sigma^{(N+1)}$ 时,完全没有利用 $\Sigma^{(N)}$ 的信息,而是重新计算;在计算 $\Sigma^{(N+1)}$ 的特征值和特征向量时,通常采用一些传统的求解算法,如 QR 算法^[2]等。实际上,注意到 $X^{(N)}$

* 收稿日期 2004 - 10 - 08

作者简介:黄晓斌(1978—)男,博士生。

与 $X^{(N+1)}$ 只相差一个样本,因此 $\Sigma^{(N)}$ 与 $\Sigma^{(N+1)}$ 应该具有很强的关联性,具体来说就是一种递推关系。进一步来说,在计算 $\Sigma^{(N+1)}$ 的特征值和特征向量时,根据这种递推关系,可以避免使用那些非常耗时的传统算法,而采用矩阵扰动理论来减少计算量。图 1 和图 2 分别给出了基于传统 K-L 变换和改进 K-L 变换法的特征提取流程框图。

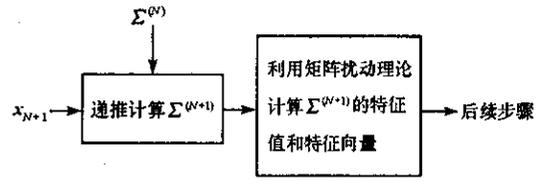
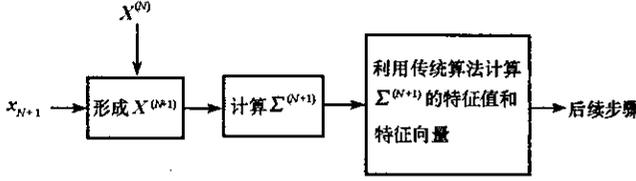


图 1 基于传统 K-L 变换的特征提取流程框图

图 2 基于改进 K-L 变换的特征提取流程框图

fig. 1 Feature extraction based-on traditional K-L transform

Fig. 2 Feature extraction based-on improved K-L transform

注意,特征提取的后续步骤,如特征值的选择、特征子空间的形成^[1],都是比较简单的步骤,且不是本文讨论的范畴,在上述两个框图中省略了。下面详细介绍改进 K-L 变换中的两个关键技术(1)计算协方差矩阵的递推算法(2)利用矩阵扰动理论计算协方差矩阵的特征值和特征向量。

2 计算协方差矩阵的递推算法

$$\begin{aligned} \text{易知: } \mu_k^{(N+1)} &= \frac{1}{N+1} \sum_{l=1}^{N+1} x_{l,k} = \frac{1}{N+1} \sum_{l=1}^N x_{l,k} + \frac{x_{N+1,k}}{N+1} \\ &= \frac{N}{N+1} \left(\frac{1}{N} \sum_{l=1}^N x_{l,k} \right) + \frac{x_{N+1,k}}{N+1} = \frac{N}{N+1} \mu_k^{(N)} + \frac{x_{N+1,k}}{N+1} \end{aligned} \quad (1)$$

$$\text{由此可得: } \mu^{(N+1)} = \frac{N}{N+1} \mu^{(N)} + \frac{1}{N+1} x_{N+1} \quad (2)$$

$$\begin{aligned} \text{另有: } \Sigma^{(N+1)} &= \frac{1}{N} \sum_{l=1}^{N+1} (x_l - \mu^{(N+1)}) (x_l - \mu^{(N+1)})^T \\ &= \frac{1}{N} \sum_{l=1}^N (x_l - \mu^{(N+1)}) (x_l - \mu^{(N+1)})^T + \frac{1}{N} (x_{N+1} - \mu^{(N+1)}) (x_{N+1} - \mu^{(N+1)})^T \end{aligned} \quad (3)$$

对(3)式的第一部分进行转换:

$$\begin{aligned} &\sum_{l=1}^N (x_l - \mu^{(N+1)}) (x_l - \mu^{(N+1)})^T \\ &= \sum_{l=1}^N \left((x_l - \mu^{(N)}) + \frac{1}{N+1} (\mu^{(N)} - x_{N+1}) \right) \left((x_l - \mu^{(N)}) + \frac{1}{N+1} (\mu^{(N)} - x_{N+1}) \right)^T \\ &= \sum_{l=1}^N (x_l - \mu^{(N)}) (x_l - \mu^{(N)})^T + \frac{2}{N+1} (\mu^{(N)} - x_{N+1}) \sum_{l=1}^N (x_l - \mu^{(N)})^T \\ &\quad + \frac{N}{(N+1)^2} (\mu^{(1,N)} - x_{N+1}) (\mu^{(1,N)} - x_{N+1})^T \\ &= \sum_{l=1}^N (x_l - \mu^{(1,N)}) (x_l - \mu^{(1,N)})^T + \frac{N}{(N+1)^2} (\mu^{(1,N)} - x_{N+1}) (\mu^{(1,N)} - x_{N+1})^T \end{aligned} \quad (4)$$

对(3)式的第二部分进行转换:

$$\begin{aligned} &(x_{N+1} - \mu^{(N+1)}) (x_{N+1} - \mu^{(N+1)})^T \\ &= \left((x_{N+1} - \mu^{(N)}) + \frac{1}{N+1} (\mu^{(N)} - x_{N+1}) \right) \left((x_{N+1} - \mu^{(N)}) + \frac{1}{N+1} (\mu^{(N)} - x_{N+1}) \right)^T \\ &= \frac{N^2}{(N+1)^2} (x_{N+1} - \mu^{(N)}) (x_{N+1} - \mu^{(N)})^T \end{aligned} \quad (5)$$

将(4)(5)式代入(3)式得:

$$\Sigma^{(N+1)} = \frac{1}{N} \sum_{l=1}^N (x_l - \mu^{(1,N)}) (x_l - \mu^{(1,N)})^T + \frac{1}{(N+1)^2} (\mu^{(1,N)} - x_{N+1}) (\mu^{(1,N)} - x_{N+1})^T$$

$$\begin{aligned}
 & + \frac{N}{(N+1)^2} (\mathbf{x}_{N+1} - \boldsymbol{\mu}^{(N)}) \chi (\mathbf{x}_{N+1} - \boldsymbol{\mu}^{(N)})^T \\
 & = \frac{N-1}{N} \boldsymbol{\Sigma}^{(N)} + \frac{1}{N+1} (\mathbf{x}_{N+1} - \boldsymbol{\mu}^{(N)}) \chi (\mathbf{x}_{N+1} - \boldsymbol{\mu}^{(N)})^T
 \end{aligned} \quad (6)$$

因此协方差矩阵的递推公式如下:

$$\begin{cases} \boldsymbol{\mu}^{(N+1)} = \frac{N}{N+1} \boldsymbol{\mu}^{(N)} + \frac{1}{N+1} \mathbf{x}_{N+1} \\ \boldsymbol{\Sigma}^{(N+1)} = \frac{N-1}{N} \boldsymbol{\Sigma}^{(N)} + \frac{1}{N+1} (\mathbf{x}_{N+1} - \boldsymbol{\mu}^{(N)}) \chi (\mathbf{x}_{N+1} - \boldsymbol{\mu}^{(N)})^T \end{cases} \quad (7)$$

初始条件为: $\begin{cases} \boldsymbol{\mu}^{(1)} = \mathbf{x}_1 \\ \boldsymbol{\Sigma}^{(1)} = \mathbf{0} \end{cases}$

3 利用矩阵扰动理论计算特征值和特征向量

注意到式(7)中 $\boldsymbol{\Sigma}^{(N+1)}$ 是由两个矩阵组成的, 当样本来自同一分布, 并且 $N \gg 1$ 时, 后一个矩阵 $\frac{1}{N+1} (\mathbf{x}_{N+1} - \boldsymbol{\mu}^{(N)}) \chi (\mathbf{x}_{N+1} - \boldsymbol{\mu}^{(N)})$ 的各元素要远小于前一个矩阵 $\frac{N-1}{N} \boldsymbol{\Sigma}^{(N)}$ 的各元素, 因此, 后者可以看作是对前者的扰动, 并且由 $\boldsymbol{\Sigma}^{(N)}$ 的特征值和特征向量很容易推得 $\frac{N-1}{N} \boldsymbol{\Sigma}^{(N)}$ 的特征值和特征向量为 $\left\{ \frac{N-1}{N} \lambda_k^{(N)} \right\}_{k=1}^M$ 和 $\{\mathbf{u}_k^{(N)}\}_{k=1}^M$ 。显然, 可以利用矩阵扰动理论来计算 $\boldsymbol{\Sigma}^{(N+1)}$ 的特征值和特征向量。

为方便, 对数学记号重新做标定。令 $\mathbf{A} = \boldsymbol{\Sigma}^{(N+1)}$, $\mathbf{A}_0 = \frac{N-1}{N} \boldsymbol{\Sigma}^{(N)}$, $\Delta \mathbf{A} = \frac{1}{N+1} (\mathbf{x}_{N+1} - \boldsymbol{\mu}^{(N)}) \chi (\mathbf{x}_{N+1} - \boldsymbol{\mu}^{(N)})^T$ 。并另记 $\frac{N-1}{N} \boldsymbol{\Sigma}^{(N)}$ 的特征值和特征向量为 $\{\lambda_0^{(k)}\}_{k=1}^M$ 和 $\{\mathbf{u}_0^{(k)}\}_{k=1}^M$, $\boldsymbol{\Sigma}^{(N+1)}$ 的特征值和特征向量为 $\{\lambda^{(k)}\}_{k=1}^M$ 和 $\{\mathbf{u}^{(k)}\}_{k=1}^M$ 。请注意, 这时的特征值和特征向量的上标符号和前面记法的下标符号含义一致。因此, 求解 $\boldsymbol{\Sigma}^{(N+1)}$ 特征值和特征向量的问题可以转化为以下的数学模型:

已知 $\mathbf{A} = \mathbf{A}_0 + \Delta \mathbf{A}$, 以及 \mathbf{A}_0 的特征值 $\{\lambda_0^{(k)}\}_{k=1}^M$ 和特征向量 $\{\mathbf{u}_0^{(k)}\}_{k=1}^M$;

求解 \mathbf{A} 的特征值 $\{\lambda^{(k)}\}_{k=1}^M$ 和特征向量 $\{\mathbf{u}^{(k)}\}_{k=1}^M$ 。

3.1 基于矩阵扰动理论的特征值与特征向量数值计算方法

关于矩阵扰动情况下的特征值和特征向量计算问题, 文献[3~5]进行了一系列讨论, 给出了许多算法, 如 RAL1、RAL3、RAL4、NRT1、LIT。就计算特征值的精度而言, LIT 方法是最好的, 但结合本文的问题, RAL3 方法无疑是最合适的, 因为它能将特征向量和特征值同时计算出来, 无需额外的计算, 从而进一步减小计算量。下面就介绍 RAL3 算法^[5]。

步骤 1 按式(8)~(10)计算特征向量的扰动量

$$e_{kj} = \frac{(\mathbf{u}_0^{(j)})^T \Delta \mathbf{A} \mathbf{u}_0^{(k)}}{\lambda_0^{(k)} - \lambda_0^{(j)}}, \quad k \neq j \text{ 且 } k, j = 1, \dots, M \quad (8)$$

$$e_{kk} = - \sum_{\substack{j=1 \\ j \neq k}}^M e_{kj} \mathbf{u}_0^{(j)}, \quad k = 1, \dots, M \quad (9)$$

$$\Delta \mathbf{u}^{(k)} = \sum_{j=1}^M e_{kj} \mathbf{u}_0^{(j)}, \quad k = 1, \dots, M \quad (10)$$

步骤 2 按式(11)获得真实特征向量 $\mathbf{u}^{(k)}$ 的估计值, 用 $\hat{\mathbf{u}}^{(k)}$ 表示

$$\hat{\mathbf{u}}^{(k)} = \mathbf{u}_0^{(k)} + \Delta \mathbf{u}^{(k)}, \quad k = 1, \dots, M \quad (11)$$

步骤 3 使用 Rayleigh 商的形式按式(12)获得真实特征值 $\lambda^{(k)}$ 的估计值, 用 $\hat{\lambda}^{(k)}$ 表示

$$\hat{\lambda}^{(k)} = \frac{(\hat{\mathbf{u}}^{(k)})^T \mathbf{A} \hat{\mathbf{u}}^{(k)}}{(\hat{\mathbf{u}}^{(k)})^T \hat{\mathbf{u}}^{(k)}}, \quad k = 1, \dots, M \quad (12)$$

众所周知,利用数值方法来计算特征值和特征向量,有两个基本问题需要讨论:①特征值与特征向量的稳定性分析;②所采用的数值方法的精度分析。

3.2 特征值与特征向量的稳定性分析

特征值与特征向量的稳定性只与求解问题本身有关,而与所采用的数值方法是无关的,但它却对数值计算方法提供了指导。只有问题本身具有较好的稳定性,采用数值计算方法才具有意义。具体到上述模型,就是要研究微小的 ΔA 扰动,是否会引起 A 的特征值和特征向量的剧烈变化。

在分析特征值稳定性问题中,需要使用如下的定理^[6]。

定理 1 设 Hermite 矩阵 P 和 Q 的特征值分别为 $\{\alpha^{(k)}\}_{k=1}^M$ 和 $\{\beta^{(k)}\}_{k=1}^M$, 则 $\forall k$ 有:

$$|\alpha^{(k)} - \beta^{(k)}| \leq \|P - Q\|_2 \quad (13)$$

式中, $\|\cdot\|_2$ 表示矩阵 \cdot 的谱范数,如 $\|P\|_2 = \max_k \{\alpha^{(k)}\}$ 。

易知 A 和 A_0 都是 Hermite 矩阵,将它们分别代替式(13)中的 P 和 Q , 则可得:

$$|\lambda^{(k)} - \lambda_0^{(k)}| \leq \|A - A_0\|_2 = \|\Delta A\|_2 \quad (14)$$

由于 ΔA 的特征值为:
$$\begin{cases} \alpha^{(1)} = \frac{1}{N+1} \|x_{N+1} - \mu^{(N)}\|^2, & k=1 \\ \alpha^{(k)} = 0, & 2 \leq k \leq M \end{cases}$$
 式中, $\|\cdot\|$ 表示向量 \cdot 的欧式长度。

因此
$$|\lambda^{(k)} - \lambda_0^{(k)}| \leq \frac{1}{N+1} \|x_{N+1} - \mu^{(N)}\|^2 \quad (15)$$

由于 $\|x_{N+1} - \mu^{(N)}\|^2$ 较小,并且 $N \gg 1$, 所以 $|\lambda^{(k)} - \lambda_0^{(k)}|$ 较小,由此可知特征值是稳定的。

有关特征向量的稳定性是个非常复杂的问题,这里只给出一个定性的结论^[21]:特征值的分离度影响特征向量的稳定性,邻近的特征值所对应的特征向量是不稳定的。

3.3 RAL3 算法算法的精度分析

由(15)式可得利用 RAL3 算法计算特征值的精度为:

$$\begin{aligned} |\hat{\lambda}^{(k)} - \lambda^{(k)}| &= |\hat{\lambda}^{(k)} - \lambda_0^{(k)} + \lambda_0^{(k)} - \lambda^{(k)}| \leq |\hat{\lambda}^{(k)} - \lambda_0^{(k)}| + |\lambda_0^{(k)} - \lambda^{(k)}| \\ &\leq \left| \frac{(\hat{u}^{(k)})^T A \hat{u}^{(k)}}{(\hat{u}^{(k)})^T \hat{u}^{(k)}} - \lambda_0^{(k)} \right| + \frac{1}{N+1} \|x_{N+1} - \mu^{(N)}\|^2 \end{aligned} \quad (16)$$

注意到,上述的精度公式由于使用了特征向量的估计值,故在计算前是无法定量确定的。当计算完毕后,可以用该公式评估利用 RAL3 算法计算特征值的精度。

有关特征向量的计算精度非常复杂,本文略。

4 仿真试验及分析

按照图 1 和图 2 的流程框图进行仿真对比分析,分别使用 $N=10\ 000, 20\ 000, \dots, 100\ 000$, 共 10 个; $M=2, 5, 10, 20$, 来随机产生原始数据矩阵 $X^{(N)}$, 增量样本 x_{N+1} 也随机设定,对每个参数配对 (N, M) 独立进行 50 次试验,并对试验结果求平均值。传统 K-L 变换中使用 QR 分解法来获得特征值和特征向量,并以它为精确值(看作 $\lambda^{(k)}$ 和 $u^{(k)}$)来分析改进 K-L 方法所得特征值和特征向量(即 $\hat{\lambda}^{(k)}$ 和 $\hat{u}^{(k)}$)的偏离度。为定量分析,给出如下的偏离度定义:

$$\text{平均特征值偏离度 } \Delta = \frac{1}{M} \sum_{k=1}^M \frac{|\hat{\lambda}^{(k)} - \lambda^{(k)}|}{\lambda^{(k)}};$$

$$M-1 \text{ 维特征子空间偏离度 } \phi = \text{span}\{\hat{u}^{(1)}, \hat{u}^{(2)}, \dots, \hat{u}^{(M-1)}\} \otimes \text{span}\{u^{(1)}, u^{(2)}, \dots, u^{(M-1)}\}.$$

式中, $\text{span}\{\hat{u}^{(1)}, \hat{u}^{(2)}, \dots, \hat{u}^{(M-1)}\}$ 表示由 $\hat{u}^{(1)}, \hat{u}^{(2)}, \dots, \hat{u}^{(M-1)}$ 所张成的 $M-1$ 维估计特征子空间, $\text{span}\{u^{(1)}, u^{(2)}, \dots, u^{(M-1)}\}$ 表示由 $u^{(1)}, u^{(2)}, \dots, u^{(M-1)}$ 所张成的 $M-1$ 维精确特征子空间,符号“ \otimes ”表示求解两个特征子空间的夹角。

图 3 和图 4 分别给出用本文方法计算特征值和特征向量的精度(图 3 的 Y 轴是对数坐标)。从图中可以看出,当样本维数 M 固定时,计算精度随原始样本数 N 的增加而提高;当原始样本数 N 固定时,计算精度随样本维数 M 的增加而降低,但总体上都维持在一个很高的精度水平。

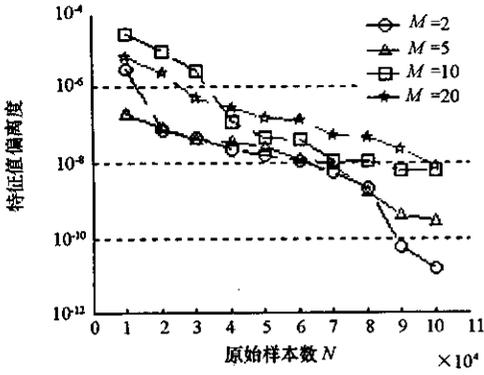


图 3 特征值偏离度随 (N, M) 变化关系
Fig. 3 Eigenvalue departure along with (N, M)

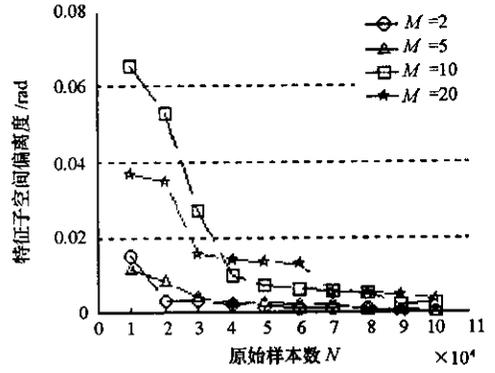


图 4 特征子空间偏离随 (N, M) 变化关系
Fig. 4 Sub-space departure along with (N, M)

图 5 和图 6 分别给出了两种方法运行时间对比。从图中可以看出,当样本维数 M 固定时,传统方法运行时间随原始样本数 N 的增加而线性快速增加,改进算法的运行时间存在一个跳变点,在跳变点的两侧基本不随 N 变化;当原始样本数 N 固定时,传统方法和改进方法的运行时间都随样本维数 M 的增加而增加。从横向比较中可以明显看出,改进算法的运行时间比传统方法大为减少。

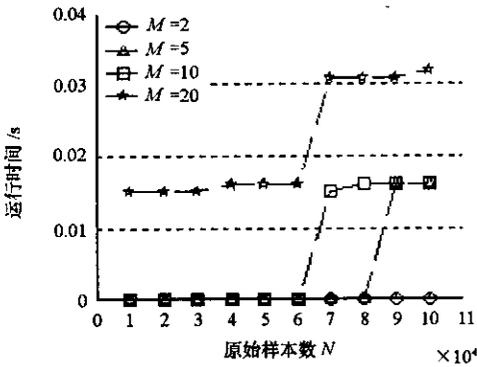


图 5 改进 K-L 方法效率随 (M, N) 变化关系
Fig. 5 Time consuming along with (N, M) based-on improved K-L transform

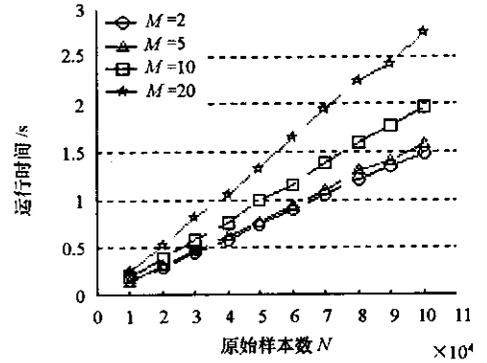


图 6 传统 K-L 方法效率随 (M, N) 变化关系
Fig. 6 Time consuming along with (N, M) based-on traditional K-L transform

5 结束语

本文基于单个样本增加的情况,使用协方差矩阵递推算法和矩阵扰动理论对传统的特征提取算法——K-L 变换法进行了改进。仿真试验表明本文的方法是有效的。但注意到,本文在进行递推时,都是假设已经知道前一步的精确值。实际上,如果将估计出来的特征值和特征向量作为当前值进行下一步递推,这样做将更具有意义。

参考文献:

- [1] 边肇祺,张学工,等. 模式识别[M]. 北京:清华大学出版社,1999.
- [2] 戈卢布 G H,范洛恩 C F. 矩阵分析[M]. 北京:科学出版社,2002.
- [3] Rogers L C. Derivatives of Eigenvalues and Eigenvector[J]. AIAA J. 1970 8(5):943 - 944.
- [4] Murthy D V,Haftka R T. Approximations to Eigenvalues of Modified General Matrices[J]. Int. J. Numerical Methods in Engineering, 1988, 26(2):293 - 331.
- [5] 刘济科,彭林芯. 一般矩阵小扰动的特征值近似计算[J]. 中山大学学报(自然科学版),2000, 7(4):23 - 26.
- [6] 吕炯兴. 关于 Hermite 矩阵厄任意扰动[J]. 南京航空航天大学学报,1998, 4(2):122 - 125.

