基于伪 PCA 的手写数字识别算法

韩 旭 刘 强 许 瑾 谌海云
(西南石油大学电气信息学院 四川 南充 637001)

摘 要 PCA(Principal Component Analysis)是最重要的数据降维算法之一,针对降维过程出现的信息丢失问题,学术界说法不一。基于此,文中提出了一种新的改进算法(Similar Principal Component Analysis,SPCA),新算法在处理过程中保留了部分细节信息。以手写数字(MNIST)数据库为例,将原始向量组进行临近特征筛选,得出多维复合非正交特征向量组;将训练库所得的向量组与测试集的向量组进行比对,识别出所测试的手写数字。结果表明,该算法能够以较少量的训练样本实现对测试样本的较为完全的识别。

关键词 主成分分析,手写数字识别,非正交特征,临近特征
中图法分类号 TP391.41 **文献标识码** A

Handwritten Numeral Recognition Algorithm Based on Similar Principal Component Analysis

HAN Xu LIU Qiang XU Jin CHEN Hai-yun
(School of Electrical Information Engineering, Southwest Petroleum University, Nanchong, Sichuan 637001, China)

Abstract Principal component analysis (PCA) is one of the most important data reduction algorithms, there is much-maligned views in the process of handling data. A novel improved similar principal component analysis (SPCA) algorithm which is based on principal component analysis (PCA) algorithm was proposed in this paper. This algorithm can keep some detail information in the process. Taking the MNIST handwritten numeral database as an example, the near feature vector is chosen in original vectors to get the groups of non-orthogonal feature vectors. Then, the vectors of training library is compared with the vectors of testing library, and the recognition rate is calculated. Recognition results indicate that the algorithm can make high identification of the testing samples through a small number of training samples.

Keywords Principal component analysis (PCA), Handwritten numeral recognition, Non-orthogonal feature vectors, Near feature vector

1 引言

PCA 方法[1]由 Pearson 于 1901 年首次提出,主要作为数据降维手段对高维数据进行处理,因此有着广泛的运用。Sirovich 等[2]首次将 PCA 用于图像识别问题,之后 Turk 与 Pentland 等[3-4]将其应用于人脸识别领域。PCA 算法的改进算法也有很多,如 MPCA[5]、KPCA[6-7]、稀疏主成分[8-9]等。文中结合闫老师[10]的观点,首先对 PCA 进行分析,然后再对伪主成分分析(SPCA)进行理论推导,并给出实验证明。

手写数字识别一直是图像识别领域的热门话题,该技术主要用于手写邮政编码分类[11]、手写账户识别[12]等。本文受文献[2]的启发,结合文献[13-14]中对非线性的处理方法,提出基于多种特征组合的方法[15]——伪主成分分析(SPCA)法。SPCA 将特征集中化,同时保留部分细节信息,以构成高效的特征约束信息族。经验证,本文算法具有较强的识别力与鲁棒性。

本文第 2 节分析与推导 SPCA;第 3 节给出实验结果并分析;最后总结全文。

2 SPCA 算法的理论推导与分析

2.1 PCA 算法

为了方便描述与引入,文中参考文献[10]中的形式。设 N 个样本矩阵集合为$\{A_k\}$,其中 $k=1,2,\cdots,N$,A_k 为 $m\times n$ 的样本矩阵,显然 A_k 是 $m\times n$ 维的张量空间[16](Tensor Space)。记:

$$A_k=\sum_{i=1}^{n}x_i \tag{1}$$

根据文献[2-3],对式(1)进行中心化处理:

$$x_i\leftarrow x_i-\frac{1}{n}\sum_{i=1}^{n}x_i \tag{2}$$

其中,$X'=\frac{1}{n}\sum_{i=1}^{n}x_i$ 为图像的均值,由式(1)和式(2)得到中心化之后的图像:

$$\overline{X}=\sum_{i=1}^{n}x_i \tag{3}$$

然后计算样本的协方差矩阵:

$$S=\overline{X}\,\overline{X}^{\mathrm{T}}=\sum_{i=1}^{n}(x_i-X')(x_i-X')^{\mathrm{T}} \tag{4}$$

再求解上述矩阵 S 的前 q 个最大特征值

本文受南充市校科技战略合作专项项目(NC17SY4011)资助。

韩 旭(1991-),男,硕士生,主要研究方向为机器学习、模式识别,E-mail:201621000193@stu.swpu.edu.cn;**刘 强**(1992-),男,硕士生,主要研究方向为图像识别、模式识别;**许 瑾**(1980-),女,硕士,讲师,主要研究方向为智能控制;**谌海云**(1967-),男,硕士,教授,主要研究方向为机器人自动化控制。

$$V=\{v_j \mid j=1,2,\cdots,q\} \tag{5}$$

与对应的特征向量

$$D=\{d_j \mid j=1,2,\cdots,q\} \tag{6}$$

D 即为所求得的主成分方向集合，也即特征空间（Eigenspace）。如果将集合$\{A_k\}$记为张量空间，A_k 量化为一个列向量 x_i，则上述所得到的 D 为样本集合$\{A_k\}$的特征子空间。每一幅图像都可以投影到该特征子空间中，进而将原图 $m\times n$ 维向量组降维成 $m\times q$ 维向量组，其中 $q\leqslant n$，实现了数据降维处理。

2.2　PCA 算法图解

假设一组 $m\times n$ 维的数据 $A_l=[x_{ij}]$，$i=1,2,\cdots,m$，$j=1,2,\cdots,n$，$l\notin\{k\}$。为了方便描述，本文中的图只示意性地画出前 3 维。在初始坐标 $x_1x_2x_3o$ 系下，A_l 在空间中的分布如图 1 和图 2 所示。

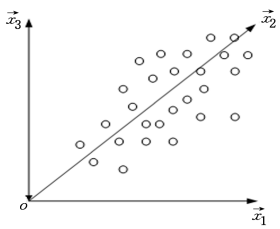

图 1　原始数据示意图

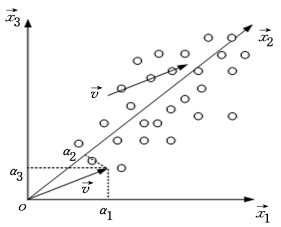

图 2　原始向量

A_l 中的某一列向量$\vec{v}$可以表示为：

$$\vec{v}=\alpha_1x_1+\alpha_2x_2+\cdots+\alpha_nx_n \tag{7}$$

经过 PCA 算法降维，新的特征空间 $x_1'x_2'x_3'o'$的方向如图 3 所示。

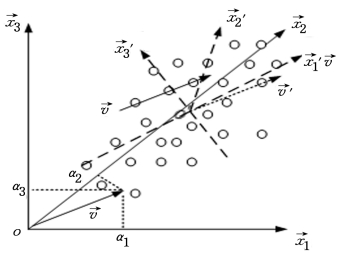

图 3　向量投影坐标转换

假设降维处理过程中，$q=n$，即数据维数不改变，投影只改变原来向量的表达形式，即：

$$\vec{v}=\vec{v}'=\alpha_1'x_1'+\alpha_2'x_2'+\cdots+\alpha_n'x_n' \tag{8}$$

此时向量$\vec{v}$的信息没有丢失，细节信息全部保留。要实现降维，$q<n$，此时式(8)应改为：

$$\vec{v}\approx\vec{v}'=\alpha_1'x_1'+\alpha_2'x_2'+\cdots+\alpha_q'x_q' \tag{9}$$

则在原来坐标系 $x_1x_2x_3o$ 中，向量$\vec{v}'$表达为：

$$\vec{v}'=\alpha_1''x_1+\alpha_2''x_2+\cdots+\alpha_r''x_r \tag{10}$$

显然，$\vec{v}'$就是 $\vec{v}$ 在原坐标系的近似表达（反投影到原来坐标系下），此时满足约束：

$$\alpha_i''\leqslant\alpha_i,\ r<n \tag{11}$$

可以看出，原数据信息有所丢失，细节信息未被全部保留。更难的是，无法评估丢失的细节信息 ξ，即无法计算丢失的 ξ 对原信息的影响。

2.3　PCA 算法的改进——SPCA

注意到式(11)中存在两种细节丢失：1）投影系数缩减 $\alpha_i''\leqslant\alpha_i$；2）投影坐标丢失 $r<n$。

分析第一类丢失问题：根据式(9)，由于数据降维，$q<n\Rightarrow\vec{v}\approx\vec{v}'$，根据据式(10)，$\alpha_i''$是由$\vec{v}'$向原坐标系投影时产生的，故只要实现数据降维，投影系数必然缩减，因此第一类丢失问题是由算法本身所决定的。分析第二类丢失问题：$r<n$，根据 PCA 识别算法，原始数据向特征子空间进行投影，与原坐标系相比，坐标轴个数减少，如果改进“投影”方法，则可以避免投影坐标丢失问题。

本文改进思想：分别计算原向量空间中与 D 空间中主轴向量$\{d_j\}$最靠近的向量$\{d_j'\}$（伪主成分（Similar Principal Component）），然后用$\{d_j'\}$替换$\{d_j\}$，构成伪特征子空间 D'，称该方法为伪 PCA 算法。

算法伪代码如算法 1 所示。

算法 1　主成分分析法伪代码

输入：记 $\phi_{i,j}(m,n)=A=\{x_1,x_2,\cdots,x_n\}$；低维空间维数（特征量个数）q

输出：矩阵 $D=\{w_1,w_2,\cdots,w_q\}$

1. 对样本进行中心化：$x_i\leftarrow x_i-\frac{1}{n}\sum_{i=1}^{n}x_i$；
2. 计算协方差矩阵：$\overline{X}\,\overline{X}^T$；
3. 对上述矩阵求解特征值和对应的特征向量；
4. 取最大的 q 个特征值所对应的特征向量 $w_1,w_2,\cdots,w_q$.

提取临近特征向量，伪代码如算法 2 所示。

算法 2　求临近特征的伪代码

输入：A，D，q

输出：矩阵 $D'=\{x_1,x_2,\cdots,x_q\}$

1. 计算向量 w_i 与 x_j 的夹角 θ_{ij}，其中 $i\in(1,q)$，$j\in(1,n)$，$\theta_{ij}\in(0,\pi)$；
2. 对 θ_i. 排序，取最小值，并确定此时对应的 j；
3. 以 x_j 来代替 w_i，记为 x_i.

SPCA 算法分析：根据式(9)，假设计算向量$\vec{v}$的临近向量为$\vec{u}$，则：

$$\vec{u}=\alpha_1'x_1+\alpha_2'x_2+\cdots+\alpha_n'x_n \tag{12}$$

投影坐标未丢失，$r=n$。伪特征子空间 D' 现在满足：1）在新的坐标系下已降维。2）在原坐标系下的维度信息未丢失，很好地保留了细节信息（在这里必须指出，D 满足：①在新的坐标系下已降维；②在原坐标系下发生了两类信息丢失）。

改进后出现了 3 个问题：1）投影系数缩减使得原来投影中的“主要”主成分“影响”降低（降维所决定的）；2）投影坐标未丢失，使得原来的细节得以保留，原来投影中的“非主要”主成分“影响”会增大——类间非线性相似度增大（由实验结果证明）；3）无法对原图进行重构。

为了降低“非主要”主成分的影响，此处定义新的概念

定义 1（重叠系数 υ）　每一类与所有类的均值的比值。

$$\upsilon=\frac{\Phi}{\frac{1}{N}\sum\Phi}\in(0,1) \tag{13}$$

其中，Φ 表示某一类学习结果 D'，$\frac{1}{N}\sum\Phi$ 表示 N 类学习结果的均值。因为无法衡量丢失的细节信息对整体的影响，所以重叠系数 υ 在式(13)下会有所波动。可以预见：q 增大，“非主要”主成分“影响”增大，则 υ 必然增大，可以通过式(14)来减弱“非主要”主成分增大带来的影响。

$$D_{ti}''=D_{ti}'-\upsilon\cdot V_{\text{mean}},\ \upsilon\in(0,1) \tag{14}$$

其中，D_{ti}'表示第 i 类训练集所学习的伪特征子空间。结合上文所述，非线性相似度无法彻底抵消，会给最终的分类决策带来极大的影响。为了更好地分类，本文引入视觉跟踪领域中用到的皮尔森相关（Pearson Correlation，PC）系数[17]来进行分类，同时使用均方误差（Mean Squared Error，MSE）最小的

方法来对皮尔森相关函数判别的结果进行比较。

2.4 分类与优化

结合式(14),0－9 共 10 类学习样本的均值为:

$$V_{\text{mean}}=\frac{1}{10}\sum_{i=1}^{10}D'_{ti} \tag{15}$$

同时,测试集样本也需要减弱“主要”主成分“影响”,增大“非主要”主成分“影响”,以提高识别率,如式(16)所示:

$$W''_{ti}=W_{ti}-\upsilon'\cdot V_{\text{mean}},\upsilon\in(0,1) \tag{16}$$

其中,W_{ti} 为第 i 类测试集样本。

本文采用皮尔森相关(Pearson Correlation,PC)系数来进行分类判别。算法伪代码如算法 3 所示。

算法 3 为 PCA 分类与优化的伪代码

输入:D''_{ti},W_t'',q

输出:矩阵 $W_l=\{i\}$

1. 记 $D''_{ti}=\{D_0'',D_1'',\cdots,D_9''\}$,$W_t''=\{y_1,y_2,\cdots,y_t\}$
2. 计算 D''_i 和 y_j 的皮尔森相关系数 r_{ij},其中 $i\in(1,9)$,$j\in(1,t)$,$t=10000$;
3. 对 $r._j$ 排序,确定最大 r_{ij},记录此时的 i 即为比较所得结果.

2.5 算法的整体步骤

算法的整体步骤如下:

Step1 根据算法 1 计算数据集 A 的 q 个主成分;

Step2 根据算法 2 计算 Step1 所得到的伪主成分,并对每一类结果求平均值;

Step3 将 Step2 所得的每一类(总共 10 个类别)的结果与每一类测试样本集按照式(14)处理;

Step4 将 Step3 所得结果按照算法 3 进行分类。

3 实验

本节采用公开的 MNIST 手写数字数据库进行算法验证,识别原理参考文献[18]中的启发式方法——同一类的两张图像的 L2 范数应该是最小的。

3.1 数据预处理

在该数据库中,手写数字 0－9 的训练样本共有 60000 张图像,测试样本共 10000 张图像。每个样本图像由像素为 28＊28 的 0－255 的数字构成,样本图如图 4 所示。

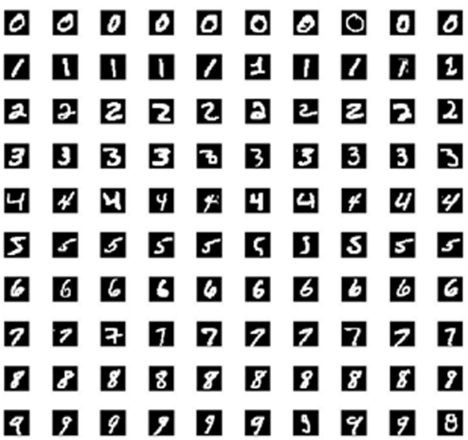

图 4 示例样本

记样本矩阵为 $\phi_{i,j}(m,n)$,$m=n=28$。由于图片中存在全零行或列,会增大类间相似度,因此考虑到接下来的成分分析,必须对样本中全零行或列做处理——本文对全零行填充幅值为 0～1 的高斯噪声,同时将样本图中非零数化为 1(不改变笔迹),以便于运算。示意图如图 5 所示。

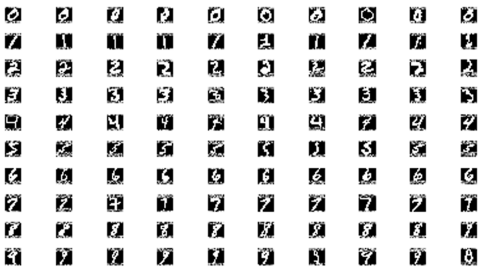

图 5 添加高斯噪声

3.2 伪主成分(临近特征)的提取

参照之前的伪代码,对各类样本集的 L 张样本进行临近特征学习,产生一张“临近特征”样本图。

例如,对数字“1”的 L 张样本进行训练,得到伪特征子空间 D_1',再将空间 D_1'中的向量求平均,得到 k 维本类训练结果。其余以此类推,得到了 0－9 10 张学习结果。由图 6 可以看出,实验结果验证了前文对于改进后出现问题的分析——“主要”主成分“影响”降低,“非主要”主成分“影响”增大。

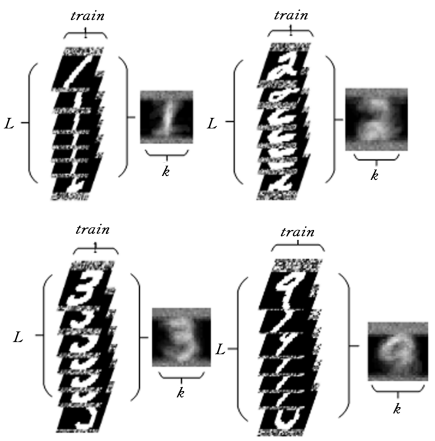

图 6 训练与学习结果

3.3 实验及结果

3.3.1 数据初始化

将训练数据(TrainingDigits)按 0－9 分成 10 组,记为 Δ_i,每组内样本编号为 δ_j,此处 j 的最大值为组内样本个数;将测试数据(TestDigits)按 0－9 分成 10 组,记为 Δ_i',每组内样本编号为 δ_j',此处 j 的最大值为组内样本个数。记每组需要进行训练的个数为 L_train,每组需要进行测试的个数为 L_test。为了使结果更加可靠,L_train 随机选取,但不超过所有 δ_j 的最小值,L_test 取 10000,即为 10000 列测试数据矩阵,以包含所有的测试样本。其中第 1－980 列:测试 0;第 981－2115 列:测试 1,共 1135 列;第 2116－3147 列:测试 2,共 1032 列;第 3148－4157 列:测试 3,共 1010 列;第 4158－5139 列:测试 4,共 982 列;第 5140－6031 列:测试 5,共 892 列;第 6032－6989 列:测试 6,共 958 列;第 6990－8017 列:测试 7,共 1028 列;第 8018－8991 列:测试 8,共 974 列;第 8992－10000 列:测试 9,共 1009 列。

3.3.2 实验流程

将训练样本经过 SPCA 算法计算得到训练样本的临近特征(伪特征子空间 D'_{ti}),再与测试样本减去一定量的所有训练样本的均值,得到 D''_{ti} 与测试样本集 W''_{ti},然后根据上文算法进行识别计算。实验流程如图 7 所示。

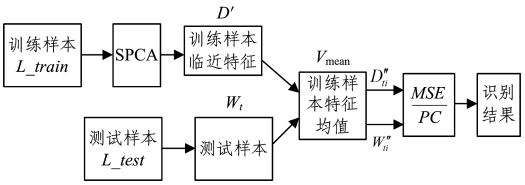

图 7　实验流程图

3.3.3　实验结果

由上述分析可知，本算法控制参数包括：特征子空间维数（特征向量个数）$k=q$；每组需要进行训练的个数 L_train，这里取 $trainl$；重叠系数 $\upsilon(\upsilon'=\upsilon)$。

每组取 $trainl$ 张图片作为训练样本，每次训练提取 k 个特征，每次测试 10000 张测试集图片，$\upsilon=0.3$，结果如图 8 所示。图 9 为验证结果。取 $trainl=350$ 时的实验数据如表 1 所列。

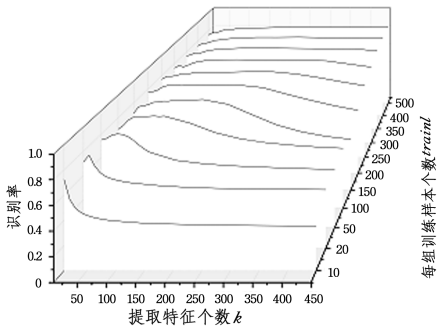

图 8　测试结果

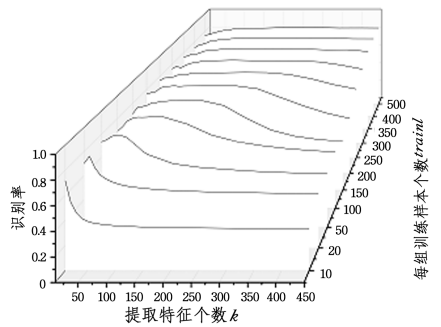

图 9　对测试进行验证

表 1　在 $trainl=350$ 条件下的数据

k	PC	MSE	time/s
10	0.7514	0.7555	20.158
20	0.7671	0.7707	21.581
30	0.7849	0.7893	23.001
40	0.7843	0.7847	23.882
50	0.8007	0.8005	25.338
60	0.8177	0.8150	26.940
70	0.8230	0.8203	27.589
80	0.8272	0.8265	29.083
90	0.8308	0.8275	30.441
100	0.8351	0.8322	31.522
150	0.8430	0.8411	37.761
200	0.8435	0.8402	41.692
250	0.8400	0.8370	48.152
300	0.8372	0.8346	55.041
350	0.8298	0.8295	64.296
400	0.8312	0.8311	72.053
450	0.8120	0.8105	73.281

由表 1 可知，k 所在的列表示提征数取特，PC 所在的列表示采用皮尔森相关函数进行分类的识别率，MSE 所在的列表示采用 MSE 进行分类验证的识别率，time 所在的列表示运行所花费的时间，单位为秒（s）。在每组输入 $trainl=350$ 个样本数据的条件下，对提取特征数 k（见表 1）进行测试与验证，结果如图 10 所示。

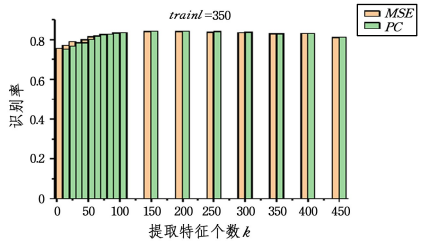

图 10　在 $trainl=350$ 条件下的数据

当 k 取 60～400 之间的值时，总体识别率在 82% 以上。当 $k=200$ 时，对 0－9 的实验测试结果如图 11 所示。

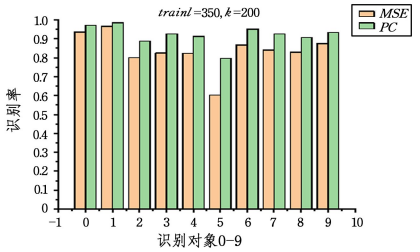

图 11　在 $trainl=350$，$k=200$ 条件下的数据

图中识别率为某组成功匹配的个数与该组待测试的样本总数的比值。由图 11 可知，当 $k=200$ 时，本文所提算法对个别数字的识别率较高，比如对数字 0，1 和 6 的识别率接近 100%，而对有些数字，例如数字“5”，识别率较低。通过实验发现，本算法在临近特征描述上，控制参数难以优化普及到所有样本。由图 10 和图 11 可以得出，PC 判别略优于 MSE 判别。

4　结论

本文认真分析了 PCA 算法并提出了改进：

1）在传统方法上，引入动态视觉领域中的“相关”作为判别函数，并且用二范数判别函数进行验证。由实验结果可知，两者结果略有差别，但总体来说，PC 判别略优于 MSE 判别。

2）通过实验结果可以判断，伪特征子空间会将细节饱和，后期继续增大训练样本或者增加特征子空间维数不会对识别率的提高做出明显的贡献，从而减少了样本依赖。

由测试结果可知，所提算法的识别效果较为理想，进而证明了前文对 SPCA 的改进方案是有效的。

本文在 MNIST 数据库上进行了算法验证。由于 MNIST 数据库不受光照、表情、旋转等因素的干扰，因此实验结果具有一定的普适性和可参考性。SPCA 算法能够以较少量训练样本[19-20]（所有训练样本在 5000 张左右）实现对测试样本（10000 张）的较完全识别，识别时间较短（在 $trainl=350$，$k=200$ 的条件下，平均每张测试样本耗时约为 4 ms）。本算法仍有待改进，通过对中间变量的分析，发现提取的临近特征保留了细节信息，但仍不能完整地描述识别对象的局部特征。同时在研究中发现，测试样本中的某些数字样本在特征表达上与人的识别不符，这些样本很容易造成误判，会降低识别率。表 2 列出本文测算的相关数据。

（下转第 307 页）

应路由算法[J]. 计算机科学，2015，42(4)：94-100.

[4] 池凯凯，杜文洁，李燕君，等. 能量捕获无线传感网的吞吐量最大化路由[J]. 计算机科学，2015，42(S2)：242-244.

[5] YANG S，MCCANN J A. Distributed Optimal Lexicographic Max-Min Rate Allocation in Solar-Powered Wireless Sensor Networks[J]. Acm Transactions on Sensor Networks，2014，11(1)：1-35.

[6] KUO T W，TSAI M J. On the construction of data aggregation tree with minimum energy cost in wireless sensor networks：NP-completeness and approximation algorithms[C]// IEEE INFOCOM. IEEE，2012：2591-2595.

[7] HE J，JI S，PAN Y，et al. Constructing Load-Balanced Data Aggregation Trees in Probabilistic Wireless Sensor Networks[J]. IEEE Transactions on Parallel & Distributed Systems，2014，25(7)：1681-1690.

[8] TUFAIL A，KHAYAM S A，ALI A，et al. On the reliability of backbone-assisted end-to-end transmissions in WSNs[C]// International Conference on Ubiquitous and Future Networks. IEEE Press，2009：22-27.

[9] LEE J H，HUR K，EOM D S. WTE-MAC：Wakeup time estimation MAC For improving end-to-end delay performance In WSN [C] // Military Communications Conference (2011-Milcom). IEEE，2012：902-907.

[10] ENGMANN F，ABDULAI J D，AZASOO J Q. Enhancing the Reliability of WSN Through Wireless Energy Transfer[C]// International Conference on Computational Science and Its Applications. Springer International Publishing，2016：610-618.

[11] LANDOLSI M A，STARK W E. On the accuracy of Gaussian approximations in the error analysis of DS-CDMA with OQPSK modulation[J]. IEEE Transactions on Communications，2002，50(12)：2064-2071.

[12] IMON S K A，KHAN A，FRANCESCO M D，et al. Energy-efficient randomized switching for maximizing lifetime in tree-based wireless sensor networks[J]. IEEE/ACM Transactions on Networking，2015，23(5)：1401-1415.

[13] VURAN M C，AKYILDIZ I F. Error Control in Wireless Sensor Networks：A Cross Layer Analysis[J]. IEEE/ACM Transactions on Networking，2009，17(4)：1186-1199.

[14] SEAH W K G，ZHI A E，TAN H. Wireless sensor networks powered by ambient energy harvesting (WSN-HEAP)-Survey and challenges[C]// International Conference on Wireless Communication，Vehicular Technology，Information Theory and Aerospace & Electronic Systems Technology，2009. Wireless Vitae. IEEE，2010：1-5.

(上接第 281 页)

表 2　测算相关数据

过程参数	特征量个数 k	每组训练个数 $trainl$	重叠系数	测试集	学习类别
	m(组)	n(组)	υ	10000	10
测算次数	$10^5 \cdot m \cdot n$				

其中，m(组)表示随机选取并测试 m 个不同的 k 值，n(组)表示随机选取并测试 n 个不同的 $trainl$ 值。例如，$m=50$，$n=20$，则在测试验证阶段的计算次数为 10^8(次)。实验结果还指出，当 υ 变化时，识别曲线类似图 8 和图 9。

结束语　本文针对 PCA 算法提出了一点改进，并根据改进后出现的问题提出了新的解决方案并引入新的判别方法。虽然本文解决了一些问题，但产生了一些新的问题。PCA 的改进方法多样，本文中的改进还有待继续优化，实验方法还可以结合其他算法，或者利用其他数据库来发现改进中的不足。

参 考 文 献

[1] PEARSON K. On lines and planes of closest fit[J]. Philosophical Magazine，1901，6：559-572.

[2] KIRBY M，SIROVICH L. Application of the Karhunen-Loeve procedure for the characterization of human faces[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence，1990，12(1)：103-108.

[3] TURK M，PENTLAND A. Eigenfaces for recognition[J]. Journal of Cognitive Neuroscience，1991，3(1)：71-86.

[4] TURK M A，PENTLAND A P. Face recognition using eigenfaces[J]. Journal of Cognitive Neurosience，1991，7：71-86.

[5] LU H，PLATANIOTIS K N，VENETSANOPOULOS A N. MPCA：Multilinear Principal Component Analysis of Tensor Objects[J]. IEEE Transactions on Neural Networks，2008，19(1)：18.

[6] YANG J，FRANGI A F，YANG J Y，et al. KPCA plus LDA：a complete kernel Fisher discriminant framework for feature extraction and recognition[J]. IEEE Trans. Pattern Anal. Mach. Intell.，2005，27(2)：230-244.

[7] CAO L J，CHUA K S，CHONG W K，et al. A comparison of PCA，KPCA and ICA for dimensionality reduction in support vector machine[J]. Neurocomputing，2003，55(1)：321-336.

[8] 任佳佳. 稀疏主成分分析的相关研究[D]. 西安：西安工程大学，2017.

[9] 谭亚芳，刘娟，王才华，等. 一种稀疏可控的主成分分析方法[J]. 计算机科学，2017，44(1)：243-246.

[10] 闫荣华，彭进业，汶德胜. PCA 与 2DPCA 的关系[J]. 计算机科学，2017，44(S2)：202-206，216.

[11] HASTIE T，FRIEDMAN J，TIBSHIRANI R. The Elements of Statistical Learning[J]. Technometrics，2001，45(3)：267-268.

[12] http://yann.lecun.com/exdb/mnist.

[13] HSU C W，LIN C J. A comparison of methods for multiclass support vector machines[J]. IEEE Transactions on Neural Networks，2002，13(2)：415-425.

[14] CHANG C C，LIN C J. LIBSVM：A Library for Support Vector Machines[J]. Acm Transactions on Intelligent Systems and Technology，2011，2(3)：1-27.

[15] 张慧，王坤峰，王飞跃. 深度学习在目标视觉检测中的应用进展与展望[J]. 自动化学报，2017，48(8)：1289-1305.

[16] 唐西林，刘燕娜. 几类张量空间的乘积基的结构[J]. 佛山科学技术学院学报(自然科学版)，2018，36(2)：1-6.

[17] https://en.wikipedia.org/wiki/Correlation_and_dependence.

[18] LEE K C，HO J，KRIEGMAN D J. Acquiring Linear Subspaces for Face Recognition under Variable Lighting[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence，2005，27(5)：684-698.

[19] 田绍兴，陈劲杰. 基于 KNN 的手写数字的识别[J]. 农业装备与车辆工程，2017，55(10)：96-97.

[20] 刘荣荣. 基于卷积神经网络的手写数字识别软件的设计与实现[D]. 包头：内蒙古大学，2015.