基于流形正则化的非光滑非负矩阵分解

姜 伟[1] 陈 耀[1] 杨炳儒[2]
(辽宁师范大学数学学院 大连 116029)[1] (北京科技大学计算机与通信工程学院 北京 100083)[2]

摘 要 经典的非光滑非负矩阵分解方法只能发现数据中的全局统计信息,对于非线性分布数据无能为力,而流形学习方法在探索高维非线性数据集真实几何结构方面具有明显优势。鉴于此,基于流形正则化思想,提出了一种新颖的基于流形正则化的非光滑非负矩阵分解方法。该方法不仅考虑了数据的几何结构,而且对编码系数矩阵和基矩阵同时进行稀疏约束,并将它们整合于单个目标函数中。构造了一个有效的乘积更新算法,并在理论上证明了算法的收敛性。标准数据集上的实验表明了 MRnsNMF 的有效性。

关键词 非负矩阵分解,非光滑,流形正则化

中图法分类号 TP181 **文献标识码** A

Manifold Regularized-based Nonsmooth Nonnegative Matrix Factorization

JIANG Wei[1] CHEN Yao[1] YANG Bing-ru[2]
(School of Mathematics, Liaoning Normal University, Dalian 116029, China)[1]
(School of Computer & Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China)[2]

Abstract The classical Nonsmooth Nonnegative Matrix Factorization(nsNMF) method discovers only the global statistical information of data and fails in dealing with nonlinear distributed data, while the manifold learning algorithms show great power in exploring the faithful intrinsic geometry structures of high dimensional data set. To address this issue, based on manifold regularization, we developed a novel algorithm called Manifold Regularized-based Nonsmooth Nonnegative Matrix Factorization(MRnsNMF). It not only considers the geometric structure in the data representation, but also introduces sparseness constraint to both coding coefficient and basis matrix simultaneously, and integrates them into one single objective function. An efficient multiplicative updating procedure was produced along with its theoretic justification of the algorithmic convergence. The feasibility and effectiveness of MRnsNMF were verified on several standard data sets with promising results.

Keywords Non-negative matrix, Nonsmooth, Manifold regularization

1 引言

依据视觉感知机理,主视皮层 V1 区神经元对信息感知能产生“稀疏表示”。非负矩阵分解[1,2] (Nonnegative Matrix Factorization, NMF)的非负约束在一定程度上产生基于局部的、稀疏的非负矩阵,因此客观上为我们提供了一种用“计算”来“表达”视觉感知过程的数学模型。标准的 NMF 方法的稀疏度并不令人满意,因此,近年来,许多科学工作者对提高 NMF 方法的稀疏性展开了全面深入的研究,取得了一系列重要研究成果。Li 等[3]在标准的 NMF 目标函数的基础上提出了局部非负矩阵分解(Local Nonnegative Matrix Factorization, LNMF),通过对编码向量加稀疏限制和基的局部性约束获得目标对象基于部分的表示。与 LNMF 方法类似,Hoyer[4]根据稀疏编码原则提出了一种非负稀疏编码(Nonnegative Sparse Coding, NNSC),采用对编码系数矩阵 ℓ_1 范数稀疏性约束,使分解后的编码矩阵具有比较好的稀疏性,该方法的一个缺点是基向量由于加性迭代,不能很好地保持非负特性。Liu 等[5]修改了 Hoyer[4]中描述的方法,用 KL 散度代替了欧式距离,构造了稀疏非负矩阵分解(Sparse Nonnegative Matrix Factorization, SNMF),基向量由于乘性迭代,克服了 NNSC 缺点,很好地保持了数据非负性。Hoyer[6]对 Liu[5]的研究进行了拓展,提出了可精确控制稀疏性的算法(Nonnegative Matrix Factorization with Sparse Constraints, NMFSC),此算法最具创新性的特点是通过一个稀疏函数精确控制编码矩阵和基矩阵的稀疏性。

以上算法通过构建稀疏性限制,应用非线性映射技术,增强了稀疏性。单独对基向量与编码向量稀疏性进行限制,取得了良好的效果。但是,由于非负矩阵分解模型的局限性,基矩阵的稀疏必将导致编码矩阵的不稀疏,反之亦然。Pascual 等[7]基于此分析构造了非光滑非负矩阵分解(Nonsmooth Nonnegative Matrix Factorization, nsNMF),该算法创造性地引入一个“光滑”矩阵到模型中,促使基矩阵和编码矩阵同时稀疏。

到稿日期:2013-05-04 返修日期:2013-08-11 本文受国家自然科学基金项目(60875029)资助。

姜 伟(1969—),男,博士生,副教授,主要研究方向为数据挖掘、流形学习、模式识别,E-mail:swxxjw@yahoo.com.cn;陈 耀(1987—),女,硕士生,主要研究方向为机器学习、模式识别;杨炳儒(1943—),男,教授,博士生导师,主要研究方向为知识工程、数据挖掘和人工智能。

许多现实世界中的数据是嵌入在高维欧氏空间的低维流形上[8]。nsNMF 是建立在“全局线性”假设下的算法，没有考虑数据的内蕴几何结构，当数据分布于高度非线性流形上时，它将不能胜任，这就大大地限制了当数据位于非线性流形时 nsNMF 算法的使用。针对此问题，近些年来提出的流形学习(Manifold Learning，ML)算法能够揭示数据内在的几何结构，寻找高维数据在低维空间中紧致嵌入。蔡登等[9]提出了图正则非负矩阵分解(Graph Regularized Nonneagtive Matrix Factorization，GRNMF)算法，该方法在矩阵分解过程中明确考虑了数据集携带的几何信息：数据点如果在原空间是邻近点，那么对应到新的基下也是邻近点。受此方法的启发，本文提出了一种新颖的基于流形正则化的非光滑非负矩阵分解(Manifold Regularized-based Nonsmooth Nonnegative Matrix Factorization，MRnsNMF)算法用于人脸图像的特征提取，通过构建所有样本的近邻图来估计数据空间的几何结构，然后将其作为正则化项添加到 nsNMF 的目标函数。提出的算法具有以下优点：1)充分利用了数据内蕴几何结构，刻画了数据非线性结构；2) 稀疏性表达方式在一定程度上克服了噪声给算法造成的影响；3)构造了一个有效的乘积更新规则，并在理论上证明了算法的收敛性。

2 非光滑非负矩阵分解

非光滑非负矩阵分解的基本思想是通过引入一个光滑矩阵，将数据矩阵分解为 3 个矩阵的乘积，其相应的目标函数定义如下：

$$X=USV \tag{1}$$

式中，X 表示数据矩阵，U 为基矩阵，V 是编码矩阵，S 是引入的正对称光滑矩阵，即

$$S=(1-\theta)I+\frac{\theta}{r}11^{\mathrm{T}}$$

式中，I 是单位矩阵，1 是全一矢量，r 是基的个数。参数 θ 满足 $0\leqslant\theta\leqslant 1$，当 $\theta=0$ 时，退化为经典非负矩阵分解模型，当 $\theta=1$ 时，任一矩阵与 S 相乘所得结果矩阵中的元素为该矩阵的均值。矩阵 S 的光滑性导致基矩阵 U 和编码矩阵 V 的稀疏性。进一步，更新规则为：

$$u_{ik}\leftarrow u_{ik}\frac{(X(SV))_{ik}}{(U(SV)^{\mathrm{T}}(SV))_{ik}} \tag{2}$$

$$v_{jk}\leftarrow v_{jk}\frac{(X^{\mathrm{T}}(US))_{jk}}{(V(US)^{\mathrm{T}}(US))_{jk}} \tag{3}$$

3 基于流形正则化的非光滑的非负矩阵分解

尽管非光滑非负矩阵分解(nsNMF)具有稀疏性和局部化特征，但是该方法并没有明确考虑对分类问题尤为重要的内蕴几何结构。鉴于此，本文提出一种新的基于流形正则化的非光滑非负矩阵分解(Manifold Regularized-based Nonsmooth Nonnegative Matrix Factorization，MRnsNMF)算法，通过包括一个正则化项，更好地保持和优化数据点之间的几何结构。

3.1 目标函数

基于流形的正则化算法的关键之处是一致性先验假设。也就是说近邻图像集合在投影低维流形时有相似的嵌入。具体而言，已知 n 个数据点 $X=\{x_1,x_2,\cdots,x_n\}$ 分布在高维欧氏空间的低维流形上，如果数据点 x_i 属于 x_j 的 k-近邻，或者 x_j 属于 x_i 的 k-近邻，则称 x_i 和 x_j 为近邻点，并在图上以边相连，并附上相应的权重 W_{ij}，其定义如下：

$$W_{ij}=\begin{cases}\exp(-\|x_i-x_j\|^2/2\sigma^2), & x_i\in N_k(x_j)\vee x_j\in N_k(x_i)\\ 0, & \text{otherwise}\end{cases}$$

式中，$N_k(x_i)$ 和 $N_k(x_j)$ 分别表示 x_i 和 x_j 的 k-最近邻，σ 和 k 为自由参数。依据流形假设，如果数据点之间有边连接，则被投影到低维流形时，数据点依然保持近邻。由此，定义局部保持正则化项如下：

$$\begin{aligned}R&=\frac{1}{2}\sum_{i,j}(y_i-y_j)^2W_{ij}\\&=\sum_i y_i^{\mathrm{T}}y_iD_{ii}-\sum_{ij}y_i^{\mathrm{T}}y_jW_{ij}\\&=Tr(VDV^{\mathrm{T}})-Tr(VWV^{\mathrm{T}})=Tr(VLV^{\mathrm{T}})\end{aligned} \tag{4}$$

式中，矩阵 D 为一对角矩阵，且 $D_{ii}=\sum_j W_{ij}$，$L=D-W$ 为图拉普拉斯矩阵。

在非光滑非负矩阵分解的基础上，考虑当数据取样于嵌入在高维空间低维流形数据的几何结构和判别信息时，最小化下面的目标函数：

$$\begin{aligned}&O=\|X-USV\|^2+\lambda Tr[VLV^{\mathrm{T}}]\\&\text{s. t.}\ U\geqslant 0,V\geqslant 0\end{aligned} \tag{5}$$

式中，λ 为正则化参数，若 $\lambda=0$，则 MRnsNMF 就退化为 nsNMF。

3.2 优化算法

目标函数 O 是 U 和 V 非凸函数，因此，它没有闭式解(Close-form Solution)，本文构建了一个迭代算法优化这个目标函数。其核心思想是固定 U，求出一个最优的矩阵 V；然后固定 V，求出一个最优的矩阵 U；不停地迭代，可以得到一个局部最优解。

依据矩阵的性质，$Tr(AB)=Tr(BA)$，则目标函数(5)可表示为：

$$\begin{aligned}O_F&=Tr((X-USV)(X-USV)^{\mathrm{T}})+\lambda Tr(VLV^{\mathrm{T}})\\&=Tr(XX^{\mathrm{T}})-2Tr(XV^{\mathrm{T}}S^{\mathrm{T}}U^{\mathrm{T}})+Tr(USVV^{\mathrm{T}}S^{\mathrm{T}}U^{\mathrm{T}})+\\&\quad\lambda Tr(VLV^{\mathrm{T}})\end{aligned}$$

由 Lagrange 定理，可得：

$$J=O_F+Tr(\Psi U^{\mathrm{T}})+Tr(\Phi V^{\mathrm{T}})$$

则 J 关于 U 和 V 的偏导为：

$$\frac{\partial J}{\partial V}=-2S^{\mathrm{T}}U^{\mathrm{T}}X+2S^{\mathrm{T}}U^{\mathrm{T}}USV+2\lambda VL+\Phi$$

$$\frac{\partial J}{\partial U}=-2XV^{\mathrm{T}}S^{\mathrm{T}}+2USVV^{\mathrm{T}}S^{\mathrm{T}}+\Psi$$

由 KKT 条件 $\varphi_{ab}v_{ab}=0$ 和 $\psi_{ab}u_{ab}=0$，得到以下乘积更新规则：

$$v_{ab}^{t+1}\leftarrow v_{ab}^{t}\frac{(S^{\mathrm{T}}U^{\mathrm{T}}X+\lambda VW)_{ab}}{(S^{\mathrm{T}}U^{\mathrm{T}}USV+\lambda VD)_{ab}} \tag{6}$$

$$u_{ab}^{t+1}\leftarrow u_{ab}^{t}\frac{(XV^{\mathrm{T}}S^{\mathrm{T}})_{ab}}{(USVV^{\mathrm{T}}S^{\mathrm{T}})_{ab}} \tag{7}$$

3.3 收敛性证明

下面证明算法的收敛性，首先引入相关的定义。

定义 1 设 $G(v,v')$ 为 $F(v)$ 的辅助函数，满足条件：$G(v,v')\geqslant F(v)$，$G(v,v)=F(v)$。

引理 1　如果 $G(v,v^t)$ 是 $F(v)$ 的辅助函数，则 $F(v)$ 在下面迭代规则下是单调递减函数：

$$v^{t+1}=\arg\min G(v,v^t) \tag{8}$$

证明：由式(8)可知，当 $v=v^{t+1}$ 时，函数 $G(v,v^t)$ 取得最小值，由定义 1 可知，$G(v^{t+1},v^t)\geqslant F(v^{t+1})$，用不等式可概括为：

$$F(v^{t+1})\overset{def}{\leqslant}G(v^{t+1},v^t)\overset{def}{\leqslant}G(v^t,v^t)\overset{def}{=}F(v^t)$$

只有当 v^t 为 $G(v,v^t)$ 的局部极小点时 $F(v^{t+1})=F(v^t)$ 才成立。如果函数 F 的导数存在且在 v^t 的一个极小领域内连续，则微分 $\nabla F(v^t)=0$。因此每次取式(8)进行迭代，最终会收敛到一个局部极小点 $v_{\min}=\arg\min_h F(v)$ 的一个序列：$F(v_{\min})\leqslant\cdots\leqslant F(v^{t+1})\leqslant F(v^t)\cdots\leqslant F(v^1)\leqslant F(v^0)$。

我们将看到通过定义这样的辅助函数 $G(v,v^t)$，对于目标函数 O，其相应的迭代规则能满足 $v^{t+1}=\arg\min G(v,v^t)$。

对于 V 中任何元素 v_{ab}，我们用 F_{ab} 表示 O 中仅与 v_{ab} 有关的部分，易得：

$$F'_{ab}=\left(\frac{\partial O}{\partial v}\right)_{ab}=(-2S^TU^TX+2S^TU^TUSV+2\lambda VL)_{ab} \tag{9}$$

$$F''_{ab}=2(S^TU^TUS)_{aa}+2\lambda L_{bb} \tag{10}$$

引理 2　函数

$$G(v,v^t_{ab})=F_{ab}(v^t_{ab})+F'_{ab}(v^t_{ab})(v-v^t_{ab})+\frac{(S^TU^TUSV)_{ab}+\lambda(VD)_{ab}}{v^t_{ab}}(v-v^t_{ab})^2 \tag{11}$$

是 F_{ab} 的辅助函数。

证明：显然有 $G(v,v)=F_{ab}(v)$。

下面证明 $G(v,v^t_{ab})\geqslant F_{ab}(v)$。

由 $F_{ab}(v)$ 的泰勒展式：

$$F_{ab}(v)=F_{ab}(v^t_{ab})+F'_{ab}(v)(v-v^t_{ab})+[(S^TU^TUS)_{aa}+\lambda L_{bb}](v-v^t_{ab})^2$$

并且

$$(S^TU^TUSV)_{ab}=\sum_{l=1}^{k}(S^TU^TUS)_{al}v^t_{al}\geqslant(S^TU^TUS)_{aa}v^t_{ab}$$

$$\lambda(VD)_{ab}=\lambda\sum_{j=1}^{m}v^t_{aj}D_{jb}\geqslant\lambda v^t_{ab}D_{bb}\geqslant\lambda v^t_{ab}(D-W)_{bb}=\lambda v^t_{ab}L_{bb}$$

则

$$\frac{(S^TU^TUSV)_{ab}+\lambda(VD)_{ab}}{v^t_{ab}}\geqslant(S^TU^TUS)_{aa}+\lambda L_{bb}$$

因此，$G(v,v^t_{ab})\geqslant F_{ab}(v)$。

综上所述，$G(v,v^t_{ab})$ 为 F_{ab} 的辅助函数。

引理 3　通过最小化辅助函数 $G(v,v^t_{ab})$ 可以得到式(6)和式(7)的解。

证明：为了得到最优解，辅助函数 $G(v,v^t_{ab})$ 对 v 进行微分有：

$$\frac{\partial G(v,v^t_{ab})}{\partial v}=F'_{ab}(v^t_{ab})+\frac{2(S^TU^TUSV)_{ab}+2\lambda(VD)_{ab}}{v^t_{ab}}(v-v^t_{ab})=0 \tag{12}$$

将式(9)代入式(12)，化简可得 v_{ab} 的更新规则：

$$(-2S^TU^TX+2S^TU^TUSV+2\lambda VL)_{ab}+\frac{2(S^TU^TUSV)_{ab}+2\lambda(VD)_{ab}}{v^t_{ab}}(v-v^t_{ab})=0$$

$$(2(S^TU^TUSV)_{ab}+2\lambda(VD)_{ab})(v-v^t_{ab})=(2S^TU^TX-2S^TU^TUSV-2\lambda VL)_{ab}v^t_{ab}$$

$$v-v^t_{ab}=\left(\frac{2S^TU^TX-2S^TU^TUSV-2\lambda VL}{2S^TU^TUSV+2\lambda(VD)}\right)_{ab}v^t_{ab}$$

$$v=v^t_{ab}\left(1+\frac{2S^TU^TX-2S^TU^TUSV-2\lambda VL}{2S^TU^TUSV+2\lambda VD}\right)_{ab}$$

所以更新规则为：

$$v^{t+1}_{ab}\leftarrow v^t_{ab}\frac{(S^TU^TX+\lambda VW)_{ab}}{(S^TU^TUSV+\lambda VD)_{ab}}$$

同理可得：

$$u^{t+1}_{ab}\leftarrow u^t_{ab}\frac{(XV^TS^T)_{ab}}{(USVV^TS^T)_{ab}}$$

4　实验和结果分析

为了全面验证基于流形的(MRnsNMF)有效性，针对本文所研究的算法与其它相关算法的关系，做了一系列对比实验。对比的算法包括非负矩阵分解(NMF)、非平滑非负矩阵分解(nsNMF)、图正则非负矩阵分解(GNMF)。为了保证对比实验的充分性与广泛性，选用了 3 个具有不同特点的数据集，包括 3 个基准人脸数据集 ORL 人脸数据集、Extended YaleB 和 M1T-CBCL 人脸集。在实验中，所有的人脸图像都经过了标准化处理，即首先标定出人脸图像中两眼的位置并在该位置将图像对齐，然后剪切出图像中的面部区域并将其缩放成统一的 32×32 大小，图像最终按列堆叠成 1024 维的向量。

4.1　数据集

ORL 数据集：此数据集包含了 40 个人的 400 幅 112×92 像素的人脸灰度图像；人脸遮挡、表情和姿态都有一定程度的不同；深度旋转和平面旋转可达 20 度，人脸的尺度也有多达 10%的变化。本文在 ORL 人脸库上选取 200 幅作为训练集(即每人 5 幅图)，其它的作为测试集。

Extended YaleB 数据集：此数据集包括 38 个人的 2414 幅正面人脸图像，其中每个人大约 64 张图像。这些图像均在实验室控制光照的条件下拍摄。本文的实验中，随机选取 Extended YaleB 人脸图像库中的 30 个人，每个人 30 幅图像，共构成 900 幅图作为训练集，剩下的样本作为测试集。

M1T-CBCL 数据集：此数据集由 10 个人的 2000 幅人脸图像构成，人脸姿态有很大的变化，其平面旋转最多可达 90 度。本文的实验中，随机选取 MIT-CBCL 人脸图像库中的 10 个人，每个人 10 幅图像，共构成 100 幅图像作为训练集，另外 10 幅图像作为测试样本。

4.2　识别率

图 1－图 3 分别给出了 MRnsNMF、NMF、nsNMF 和 GNMF 算法在 ORL 数据集、Extended YaleB 数据集和 CBCL 数据集上对应于不同子空间维数的识别率曲线图。重复 10 轮，取平均识别率作为最终性能指标。由图 1 可知，在 ORL 数据集上，MRnsNMF 与其他 3 个算法相比得到最好的识别率，并在特征子空间维数 $r=49$ 时，达到最高识别率85.35%；从图 2 可以看出，在 Extended YaleB 数据库上，该算法在 $r=49$ 时达到最高识别率 85.2%；从图 3 可以看出，在 CBCL 数据库上，该算法在 $r=100$ 时达到最高识别率 95.39%。

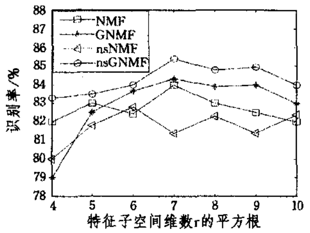

图 1　在 ORL 上 r 取不同值时的识别率

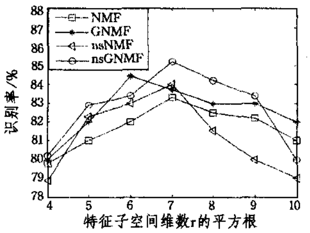

图 2　在 Extended YaleB 上 r 取不同值时的识别率

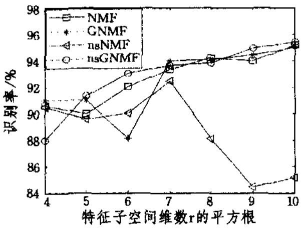

图 3 在 CBCL r 取不同值时的识别率

4.3 参数 r 的选择

在子空间的学习中，参数的选择是非常重要的，同样 MRnsNMF 作为一种子空间算法，基矩阵的维数 r 的取值影响着算法的收敛性能，理论上讲 $r\ll\min(m,n)$，但 r 的取值过大或过小都会影响算法的收敛性速度和最终分解的效果。图 4、图 5 和图 6 分别为在 ORL、Extended YaleB 和 CBCL 数据集上 r 取不同值时 MRnsNMF 算法对应的目标函数误差曲线图（其中参数 $\lambda=1,\theta=0.5$）。由图可知，目标函数呈单调下降趋势，最终收敛。

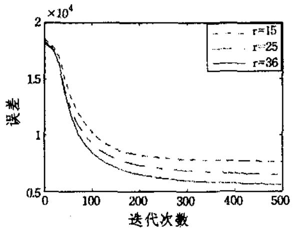

图 4 r 取不同值时的误差曲线

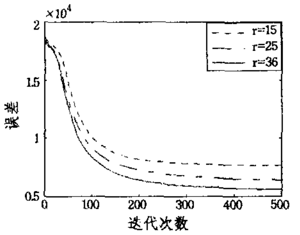

图 5 r 取不同值时的误差曲线

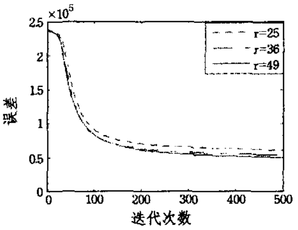

图 6 r 取不同值时的误差曲线

4.4 基图像与编码图像的稀疏度

首先定义度量稀疏度的函数[10]为：

$$sp(x)=\frac{\sqrt{n}-(\sum|x_i|)/\sqrt{\sum x_i^2}}{\sqrt{n}-1} \tag{13}$$

其中，n 是向量 x 的维度，且 $0\leqslant sp(x)\leqslant 1$，当且仅当 x 的所有元素都相等时，函数值为 0；当且仅当 x 仅有一个非零元素时，函数值为 1。

图 7 和图 8 表示了在数据库 ORL 上特征维数 r 取值 36 时，分别用 NMF、nsNMF、MRnsNMF 算法对由训练图像构成的非负矩阵进行矩阵分解得到的基图像和编码图像，并依据式(13)度量稀疏度。由图 7 可知，NMF 的稀疏度最差，MRnsNMF 的稀疏度最高，超过了 nsNMF 基图像的稀疏度。由图 8 可知，MRnsNMF 编码图像的稀疏度也相对较高。也就是说，与其它两个算法相比，MRnsNMF 算法得到了最佳的局部表示，并且在基矩阵和编码矩阵上同时实现了稀疏。

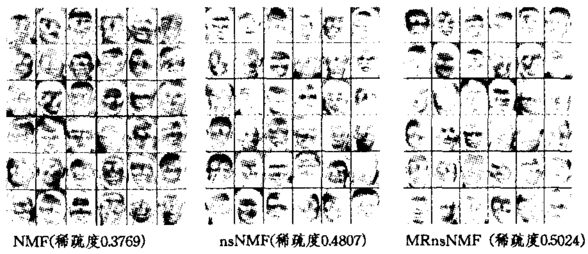

图 7 3 种算法在特征维数为 36 时提取得到的基图像

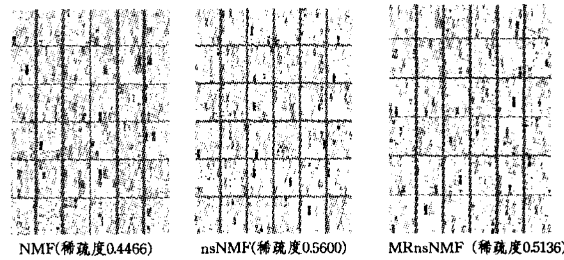

图 8 3 种算法在特征维数为 36 时提取得到的编码图像

结束语 提出了基于流形正则化的非光滑非负矩阵分解，通过在原目标函数上附加正则化项，定义算法的目标函数，给出了迭代规则及其证明过程，并与原始 nsNMF 算法做比较。以 ORL 人脸库、Yale 人脸库和 CBCL 人脸库的数据为实验对象对算法进行了验证，结果表明新算法不仅可行而且有效并能得到更好的识别率，由算法得到的解具有充分的稀疏性，收敛速度较快。

参 考 文 献

[1] 姜伟，杨炳儒. 局部敏感非负矩阵分解[J]. 计算机科学，2010，37(12)：211-214

[2] Lee D D, Seung H S. Learning the parts of objects by non-negativematrix factorization[J]. Nature, 1999, 401(6755): 788-791

[3] Li Stan Z, Hou Xin-wen, Zhang Hong-jiang, et al. Learning spatially localized, parts-based representation[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2001, 1: 207-212

[4] Hoyer P O. Non-negative sparse coding[C]// IEEE Workshop on Neural Networks for Signal Processing. 2002: 557-565

[5] Liu W X, Zheng N N, Lu X F. Non-negative matrix factorization for visual coding[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. 2003, 3: 293-296

[6] Hoyer P O. Non-negative matrix factorization with sparseness constraints [J]. Journal of Machine Learning Research, 2004, 5(9): 1457-1469

[7] Alberto P M, Carazo J M, Kochi K, et al. Nonsmooth Nonnegative Matrix Factorization [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(3): 403-414

[8] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computations, 2003, 15(6): 1373-1396

[9] Cai Deng, He Xiao-fei, Han Jia-wei, et al. Graph Regularized Non-negative Matrix Factorization for Data Representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 8(33): 1548-1560

[10] Hoyer P O. Non-negative matrix factorization with sparseness constraints [J]. Journal of Machine Learning Research, 2004, 5(9): 1457-1469