

基于稀疏表示的人脸识别方法

杨荣根^{1,2} 任明武¹ 杨静宇¹

(南京理工大学计算机学院 南京 210094)¹ (淮阴工学院计算机工程系 淮安 223003)²

摘 要 分析了稀疏表示的数学本质就是稀疏正规化约束下的信号分解,研究了一种正交匹配追踪的稀疏表示算法并利用矩阵 Cholesky 分解简化迭代过程中矩阵求逆计算来快速实现算法,将该算法应用在人脸识别中,利用训练样本构建冗余字典,将测试样本看成冗余字典中训练样本的线性组合,通过在不同人脸库上的实验证明了该方法的有效性。

关键词 稀疏表示,稀疏编码,压缩感知,正交匹配追踪,特征提取

中图分类号 TP391 **文献标识码** A

Sparse Representation Based Face Recognition Algorithm

YANG Rong-gen^{1,2} REN Ming-wu¹ YANG Jing-yu¹

(School of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, China)¹

(Faculty of Computer Engineering, Huaiyin Institute of Technology, Huai'an 223003, China)²

Abstract We analyzed the mathematic essence of sparse representation, sparse regularized signal decomposition. Studied a sparse representation algorithm of orthogonal matching pursuit. Using the matrix Cholesky decomposition, we realized the OMP algorithm a fast version. We cast the recognition problem as one of classifying among multiple linear regression models and developed a new framework from sparse signal representation. We viewed a test sample as the linear combination of training samples. We conducted experiments on face recognition to verify the efficacy of the proposed algorithm.

Keywords Sparse representation, Sparse coding, Compressed sensing, Orthogonal matching pursuit, Feature extraction

1 引言

信号的数字采样过程将数据表示成时间或空间的 delta 函数之和,这种表示给分析带来很大不便。信号处理技术常常需要通过更有效的表示来捕获信号的特征;对识别来说,表示需要突出显著特征,对去噪来说,表示需要有效的分离信号和噪声,对压缩来说,表示需要用少量的系数来描述大部分的信号。有趣的是这些应用中看似目的不同,但都有一个共同的目标就是简化信号表示(稀疏化)。

在冗余字典上的稀疏表示^[1]是一个快速发展的领域,在许多基础的信号和图像处理中取得了目前最好的效果^[2-5]。基本的模型表明自然信号能够被表示(或逼近)成预先定义的原子信号的线性组合,而且这些组合系数是稀疏的,也就是说大部分系数是 0,或接近 0。形式描述如下:设 $y \in R^m$ 是一个列信号,一些原子信号构成字典 $A \in R^{m \times n}$,则稀疏正规化下稀疏编码问题为

$$\hat{x} = \argmin ||y - Ax||_2^2 \text{ subject to } ||x||_0 \leq T \quad (1)$$

误差约束下稀疏编码问题为

$$\hat{x} = \argmin ||x||_0 \text{ subject to } ||y - Ax||_2^2 \leq \epsilon \quad (2)$$

在上面的表达式中 $\hat{x}$ 是 y 的稀疏表示, ϵ 是误差容限, T

是稀疏阈值, $|| \cdot ||_0$ 表示 l^0 范数,列数向量中非 0 元素的个数。

式(1),式(2)在很多文献中都有研究,而且被证明是 NP 难问题,但还是可以通过不同的方法逼近,包括正交匹配追踪、基追踪^[6]、Focal 欠定系统求解(FOCUSS)、梯度追踪^[7]等。

稀疏表示的最初目的是为了用比香农定理更低的采样率来表示和压缩信号^[8],事实上 JPEG2000 压缩编码标准的成功主要就归功于自然图像小波系数的稀疏性^[9]。在去噪方面利用小波方法和平移不变小波也开发出许多有效算法^[10]。稀疏和过完备性已经被成功应用于图像动态范围压缩^[11]、图像卡通和纹理成分的分离^[12]等。在这些应用中算法的性能往往要看表示的稀疏性和是否忠实于原始信号。稀疏表示过程中的字典也没有假定任何特殊的意义,只是从一些标准的基库中选择如 Fourier, Wavelet, Curvelet, Gabor 等,甚至是从随机矩阵中生成的。

本文主要应用稀疏表示进行模式分类,在分类过程中与图像处理不同的是过完备字典的选择并不是上面提到的从标准的基库中选择而是训练样本本身。测试样本稀疏表示的过程就是找到样本在这些字典元素的,即训练样本上的线性组

到稿日期:2009-10-26 返修日期:2010-01-15 本文受国家自然科学基金项目(60875010)资助。

杨荣根(1979—),男,博士生,讲师,主要研究方向为信号与信息处理、信号稀疏分解等, E-mail: yang_rg@126.com; 任明武(1969—),男,教授,博士生导师,主要研究方向为图像处理与分析; 杨静宇(1941—),男,教授,博士生导师,主要研究方向为模式识别。

合,不过这种组合要满足稀疏约束(或者叫做稀疏正规化)。后面我们会讨论这些有趣的问题,很自然地看到样本确实能通过 OMP-Cholesky 算法找到其在字典上的最稀疏表示,有了这个表示系数就可以将样本归类到系数最大的那个样本所属的类别。这里所用的算法 OMP-Cholesky 简单、快速,不同于文献[13]所用的算法。

2 稀疏表示算法

2.1 正交匹配追踪算法

正交匹配追踪算法是稀疏表示的一种经典的也是最简单的算法之一,它试图求得式(1),式(2)提出的问题逼近解。为了简化描述,假定字典 A 的列已经归一化,正交匹配追踪算法在每一步迭代过程中选择和当前迭代残差最相关的原子,选定原子以后,将信号正交投影到这些原子张成的空间中,重新计算残差,由此循环直到满足约束条件(见算法 1)。注意第 5 行是贪心选择原子步骤,第 7 行是正交投影步骤。

在第 7 行的计算由于其计算成本高通常不会显式进行,一种实际可行的方法就是利用矩阵 Cholesky 或 QR 分解来减少矩阵求逆的计算量。 $x_I = (A_I)^+ y = (A_I^T A_I)^{-1} A_I^T y$ 的计算需要求矩阵 $A_I^T A_I$ 的逆,它在投影过程中仍然保持非奇异性,这保证了线性无关原子的选择。符号 I 表示索引向量 $(i_1, \dots, i_k)$, A_I 是 A 的子矩阵,包含了被 I 索引的列。

算法 1 正交匹配追踪

1. Input: Dictionary A , signal y , target sparsity K or target error ϵ
2. Output: Sparse representation x
3. Init: Set $I := ()$, $r := y$, $x := 0$
4. while(stopping criterion not met) do
5. $\hat{k} := \arg \max_k |d_k^T r|$
6. $I := (I, \hat{k})$
7. $x_I := (A_I)^+ y$
8. $r := y - A_I x_I$
9. end while

2.2 一种更快的改进: OMP-cholesky

在正交匹配追踪算法中矩阵 $A_I^T A_I$ 是一个对称正定阵,在每一步迭代更新中仅仅是追加一行或者一列,因此它的 Cholesky 分解仅仅需要计算它的最后一行。这很容易验证假定未更新之前矩阵的 Cholesky 分解 $\tilde{F} = \tilde{L}\tilde{L}^T \in R^{(n-1) \times (n-1)}$,则在更新追加行之后的 Cholesky 分解

$$F = \begin{pmatrix} \tilde{F} & v \\ v^T & c \end{pmatrix} \in R^{n \times n} \quad (3)$$

由 $F = LL^T$ 和式(4)给定。

$$L = \begin{pmatrix} \tilde{L} & 0 \\ w^T & \sqrt{c - w^T w} \end{pmatrix}, w = \tilde{L}^{-1} v \quad (4)$$

整个 OMP-Cholesky 算法实现见算法 2。

算法 2 OMP-Cholesky

1. Input: Dictionary A , signal y , target sparsity K or target error ϵ
2. Output: Sparse representation x
3. Init: Set $I := ()$, $L := [1]$, $r := y$, $x := 0$, $\alpha^{(0)} := A^T y$, $n := 1$
4. while(stopping criterion not met) do
5. $\hat{k} := \arg \max_k |a_k^T r|$
6. if $n > 1$ then
7. $w := \text{solve for } w \{Lw = A_I^T a_{\hat{k}}\}$

8. $L := \begin{bmatrix} L & 0 \\ w^T & \sqrt{c - w^T w} \end{bmatrix}$
9. end if
10. $I := (I, \hat{k})$
11. $x_I := \text{solve for } c \{LL^T c = \alpha_I\}$
12. $r := y - A_I x_I$
13. $n := n + 1$
14. end while

与原来正交匹配追踪方法不同的是在每次迭代原子选择步骤,不再需要显式地知道残差 r 或 x ,只要知道 $A^T r$ 。因此,我们不需要显式地计算 r 和它的乘数 A^T ,取而代之去计算低计算开销的 $A^T r$ 。用 α 表示 $\alpha := A^T r$, $\alpha^{(0)} = A^T y$ 和 $G = A^T A$,可以写成

$$\alpha = A^T (y - A_I (A_I)^+ y) = \alpha^{(0)} - G_I (A_I)^+ y = \alpha^{(0)} - G_I (G_{I,I})^{-1} \alpha_I^{(0)} \quad (5)$$

这意味着给定预先计算的 $\alpha^{(0)}$ (α 的初始值也就是迭代之前的值)和 G ,就可以在每次迭代中只要计算 α 不去显式计算 r 。更新步骤中只要矩阵 G_I 而不是应用整个的字典 A^T 。注意,矩阵 $G_{I,I}$ 的逆矩阵应用上面讨论的 Cholesky 分解进行求逆。

3 OMP-cholesky 稀疏分解实现分类

模式识别的基本问题就是利用标记过类别信息的训练样本将新的测试样本正确归类到所属类别中。我们将给定的第 i 类的 n_i 个训练样本作为以上稀疏表示模型中过完备字典中的列集 $A_i = [v_{i,1}, v_{i,2}, \dots, v_{i,n_i}] \in R^{m \times n_i}$ 。具体在人脸识别中就是将一幅 $w \times h$ 的灰度人脸图像作为列向量 $v \in R^m$ ($m = wh$),由这些列向量构成过完备字典。

3.1 基于稀疏表示的分类原理

假定任意一类 i 都有足够多训练样本, $A_i = [v_{i,1}, v_{i,2}, \dots, v_{i,n_i}] \in R^{m \times n_i}$,来自同一类别的测试样本 $y \in R^m$ 可以被该类训练样本的线性组合逼近

$$y = \alpha_{i,1} v_{i,1} + \alpha_{i,2} v_{i,2} + \dots + \alpha_{i,n_i} v_{i,n_i} \quad (6)$$

但在识别过程中测试样本类别 i 未知,我们定义一个新的矩阵 A ,其列由 k 个类别的所有训练样本构成。

$$A = [A_1, A_2, \dots, A_k] = [v_{1,1}, v_{1,2}, \dots, v_{k,n_k}] \quad (7)$$

y 的线性表示重写为 $y = Ax$,其中 $x = [0, \dots, 0, \alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,n_i}, 0, \dots, 0]^T \in R^n$ 是一个系数向量,理想情况下除了该测试样本所属类别的索引值非 0,其它的系数都应该为 0。

对于一个新的测试样本 y ,我们首先通过 OMP-cholesky 稀疏表示算法计算它的稀疏表示 $\hat{x}$,该估计中的非 0 系数应该和 A 中属于某一类 i 的原子相关,根据这些非 0 系数就可以很快判断测试样本所属类别。但是由于噪声和模型误差的存在使得许多小的非 0 项和多个类别相关。基于全局的稀疏表示,我们可以设计许多可能的分类器来解决这个问题,在这里只是简单地将测试样本 y 归类到系数最大也就是最相关的那个原子所属的类别中。下面的算法 3 是整个识别过程。

算法 3 基于 OMP-cholesky 稀疏分解算法的分类

1. Input: a matrix of training samples
- $A = [A_1, A_2, \dots, A_k] \in R^{m \times n}$ for k classes, a test sample
- $y \in R^m$, sparsity threshold T .

2. Normalize the columns of A to have unit l^2 -norm.
3. Solve the problem below via OMP-cholesky algorithm

$$\hat{x} = \operatorname{argmin} ||y - Ax||_2^2 \text{ subject to } ||x||_0 \leq T$$

4. Find the index index of $\hat{x}$ corresponding to $\max(\hat{x})$
5. Output: identity(y) = $\lceil \text{index}/n_i \rceil$

3.2 应用在 ORL 数据库的人脸识别

我们应用这种方法在人脸识别中,人脸数据以其数据的高维性和复杂性成为模式识别公认的最广泛的研究数据。我们主要关注在两种情况下的人脸识别,分别是不同光照和不同表情条件。在识别过程中假定人脸从图像中已经经过如检测、分割、正规化等预处理。

我们在两个不同的数据库上做识别实验,分别是 ORL 和扩展的 Yale B,前者有 40 个不同人脸,每个人脸有 10 张人脸图像,总共 400 张图像,这些图像有不同的表情变化如有的带了眼镜有的则没有,有的睁开眼睛有的闭着眼睛,有的则有脸部侧向移位等(见图 1),所有图像都有相同的背景。扩展的 Yale B 数据库是一个大的人脸库,包含 16128 张人脸图像,38 个不同人脸在 9 种姿势 64 种不同光照条件下的数据,我们用其中的一个子集,不过这已经足够演示 OMP-cholesky 稀疏表示算法对光照条件的敏感性。

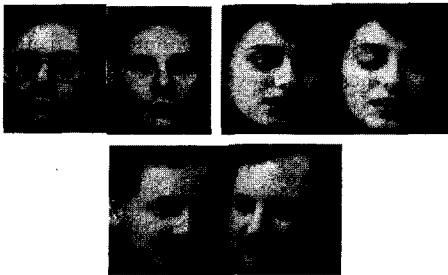


图 1 ORL 数据库不同表情变化

实验 1 有表情变化的识别

我们随机选择 ORL 每个脸类的不同数量的图像作为训练集,余下的做测试。为了保证实验过程中字典的过完备性,也就是字典的列的维数大于行的维数,需要将原来的图像下采样,原始图像的大小为 112×92 ,下采样至 14×12 ,形成以 168 维的特征向量。

在实验中分别选择每个类别中 5,6,7,8 张图像作为训练集构成过完备字典,因此字典矩阵 A 的大小分别为 $168 \times 200, 168 \times 240, 168 \times 280, 168 \times 320$,这样 $y = Ax$ 方程组都是欠定的,其稀疏阈值设置成和训练样本的个数相同。表 1 显示的是随着样本数量的不同识别率的情况,当然每一类中训练样本的数量越多,识别的效果也会更好。总之在训练样本足够多的情况下测试样本都能表示成字典中所属类别的原子的线性组合。

表 1 OMP-Cholesky 应用在 ORL 数据库

训练图像	测试图像	字典大小	稀疏阈值	识别率(%)
5	5	168×200	5	81.0
6	4	168×240	6	87.5
7	3	168×280	7	91.7
8	2	168×320	8	92.5

实验 2 OMP-cholesky 对不同光照的敏感性

扩展的 Yale B 数据库提供了不同光照条件下的人脸数据,将 OMP-cholesky 应用在 Yale B 可以测试该算法对光照

条件的敏感性。我们只用其中一个子集共 2432 张图像,38 个人在 64 种不同的光照条件下的数据(见图 2)。Yale B 数据库中图像大小为 192×168 ,同样需要下采样,至 12×11 。实验中的稀疏阈值也设置成训练样本的个数。表 2 列出了实验中其他参数和不同训练样本数量下的识别率。每个类别的训练样本数量从 16 到 48 变化。

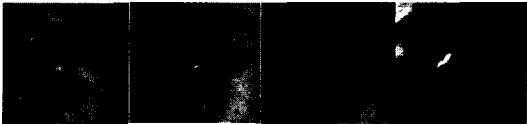
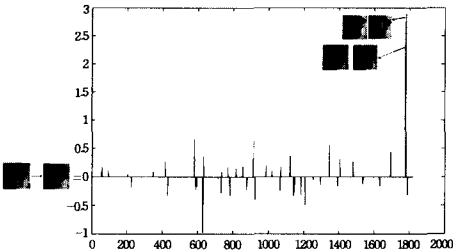


图 2 Yale B 数据库中不同光照变化

表 2 OMP-Cholesky 在应用 Yale B 数据库

训练图像	测试图像	字典大小	稀疏阈值	识别率(%)
16	48	132×608	16	88.3
32	32	132×1216	32	92.8
48	16	132×1824	48	96.2

图 3 显示的是数据库中第 38 类第 64 幅图像和它下采样后的特征图像,其稀疏表示系数对应着 48 个训练样本的线性组合。最大两个组合系数 2.87, 2.73 其对应的索引值为 1779, 1778, 正是第 38 类中的样本。



第 38 类中测试样本及其下采样后 12×11 的特征图像经过 OMP-Cholesky 稀疏分解后的稀疏系数。其最大的两个系数值 2.87, 2.73 的索引分别对应着 1779 和 1778

图 3

表 2 的数据可以看出 OMP-Cholesky 稀疏表示算法在大的人脸库,特别是在不同光照条件下同样表现出好的性能。总的来说,基于 OMP-Cholesky 的分类对表情变化比光照条件更敏感。这也是情理中事,因为我们选的 Yale B 数据库中一个子集只有光照变化也就是像素灰度值变化,而 ORL 数据库中的图像由于表情变化导致像素本身就发生了位移。

结束语 本文研究了稀疏表示理论并应用矩阵 Cholesky 分解快速实现算法,该算法简化迭代过程中逆矩阵的计算,简单快速易于实现,并在两个分别具有表情变化和不同光照的人脸数据上实验,取得很好的识别率。这种算法适合于人脸这种类似的高维数据的分类,在人脸识别过程中算法对表情的变化比对光照的变化敏感。

人脸识别的困难主要在于解决遮挡之类的问题,将稀疏表示应用在有这些限制的识别中以及在模式识别图像处理方面潜在的鲁棒应用还有待进一步研究,这也是一个活跃的研究方向。

参考文献

[1] Chen S, Donoho D, Saunders M. Atomic Decomposition by Basis Pursuit[J]. SIAM Review, 2001, 43(1): 129-159

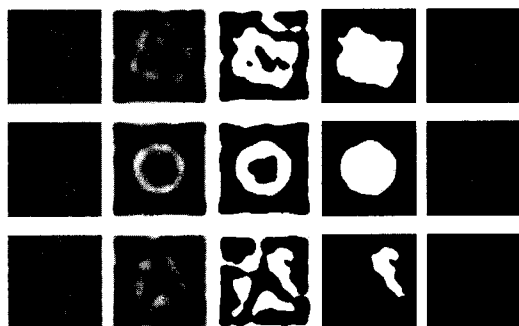


图2 3个候选颗粒图像采用DoG掩模的处理结果

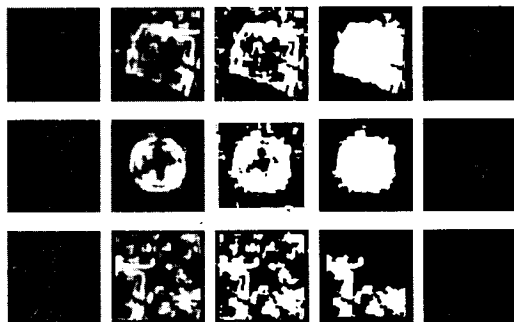


图3 3个候选颗粒图像采用差分图像掩模的处理结果

表1 DoG和差分图像掩模特征识别率对比

特征	DoG	差分图像
面积	68%	36.96%
中心偏移	85%	44.74%
形状复杂性	41.46%	36.96%
周长	43.59%	45.95%
Hu矩M1	70.83%	33.33%
Hu矩M2	51.52%	30.91%
Hu矩M3	56.67%	29.82%
Hu矩M4	44.74%	32.69%
Hu矩M5	43.59%	31.48%
Hu矩M6	45.95%	33.33%
Hu矩M7	47.22%	31.48%
扁度	53.13%	33.33%
方差	50%	43.59%
圆形度	41.46%	36.96%

表1列出了本文定义的14个特征在两种方法下的单一识别率。特征的单一识别率是仅依据该项特征进行分类正确的百分比。从结果可以看出,本文提出的方法计算得到的特

征值的单一识别率是令人满意的。除了周长以外,本文方法得到的特征值的识别能力均高于差分图像掩模方法。在此基础上,如能选用合适的分类器,将取得良好的识别效果。

结束语 特征是模式识别的关键。形状特征是识别生物大分子颗粒的有力依据。冷冻电镜图像因为低对比度和低信噪比的特点,很难提取其形状特征。DoG能提取不同尺度的边缘信号。本文利用DoG这一特点提取生物大分子颗粒的边缘信号,尽量抑制颗粒内部不明显的纹理细节信息,从而得到生物大分子颗粒的近似区域,并以此计算其近似形状特征。实验表明该方法得到的形状特征值对识别大分子颗粒能起到很好的效果。但本文方法中得到的近似颗粒区域还是会受到相邻物体的边缘影响而产生毛刺,因此,在今后的研究中还应逐步改进,以得到更接近实际颗粒的近似颗粒区域。

参考文献

- [1] Chen J Z, Grigorieff N. SIGNATURE: a single-particle selection system for molecular electron microscopy[J]. J. Struct. Biol., 2007, 157(1): 168-173
- [2] David W, Ben H, Ericksson G. The Laplacian of Gaussian and arbitrary z-crossings approach applied to automated single particle reconstruction[J]. Journal of Structural Biology, 2007, 159(1): 122-134
- [3] Ogura T, Sato C. Automatic particle pickup method using a neural network has high accuracy by applying an initial weight derived from eigenimages; a new reference free method for single-particle analysis[J]. J. Structure Biology, 2004, 145: 63-75
- [4] Sorzano C O S, Recarte E, Alcorlo M, et al. Automatic particle selection from electron micrographs using machine learning techniques[J]. Journal of Structural Biology, 2009
- [5] Satya P M, Zhu Yuanxin, Kriegman D. Detecting particles in cryo-EM micrographs using learned features[J]. J. structure Biology, 2004, 145: 52-62
- [6] Hall R J, Ardan P. A two step approach for semi-automated particle selection from low contrast cryo-electron micrographs[J]. J. Structure Biology, 2004, 145: 19-28
- [7] Pizer S M, Koenderink J J, Lifshitz L M, et al. An image description for object definition based on external regions in the stack [C]//Proc. of 9th Conf. Information Processing in Medical Imaging. Boston, MA: Martinus Nijhoff, 1986

(上接第269页)

- [2] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries[J]. IEEE Transactions on Image Processing, 2006, 15(12): 3736-3745
- [3] Protter M, Elad M. Image Sequence Denoising via Sparse and Redundant Representations[J]. IEEE Transactions on Image Processing, 2008, 18: 27-35
- [4] Bryt O, Elad M. Compression of Facial Images Using the K-SVD Algorithm[J]. Journal of Visual Communication and Image Representation, 2008, 19(4): 270-283
- [5] Liao H, Sapiro G. Sparse image representation for limited data tomography[J]. IEEE International Symposium on Biomedical Imaging, 2008: 1375-1378
- [6] Donoho D, Elad M. Optimal Sparse Representation in General Dictionaries via L1 Minimization[J]. Proceedings of the National Academy of Sciences, 2003, 100: 2197-2202
- [7] Blumensath T, Davies M. Gradient pursuits [J]. IEEE Transactions on Signal Processing, 2008, 56(6): 2370-2382
- [8] Cande E. Compressive Sampling [C]//Proc. Int'l Congress of Mathematicians. 2006
- [9] Marcellin M W, Gormish M J, Bilgin A, et al. An overview of JPEG-2000[C]//Proc. Data Compression Conf. 2000: 523-541
- [10] Starck J L, Candes E J, Donoho D L. The curvelet transform for image denoising[J]. IEEE Trans. Image Process., 2002, 11: 670-684
- [11] Gstaad R, Starck J L. Dynamic range compression: A new method based on wavelet transform[C]//Astron. Data Anal. Software Systems Conf. Strasbourg, 2003
- [12] Starck J L, Elad M, Donoho D L. Image decomposition: Separation of texture from piece-wise smooth content[C]//SPIE Conf. Signal Image Process. Wavelet Applicat. Signal Image Process. 2003
- [13] Wright J, Yang Y, Ganesh A, et al. Robust face recognition via sparse representation[J]. pattern analysis and machine intelligence, 2009, 31(2): 210-227