

摘要——我们考虑了自动识别有着不同的表情、光照、以及遮挡和伪装的正面人脸存在的问题。我们把这类识别问题描述为一个多线性回归模型中的分类问题，并且认为来自稀疏信号表示的新理论为这类问题提供了解决办法。基于由 l_1 范数最小化计算的稀疏表示，我们提出了一个通用分类算法用于目标识别（基于图像）。这个新的框架为两个人脸识别领域的关键课题（特征提取和对遮挡的鲁棒性）上提供了更好的见解。对于特征提取，我们证明了如果识别问题中的稀疏性能够被合理利用的话，特征的选择也就不再至关重要了。而真正重要的问题是：特征数是否足够大、稀疏表示是否被正确计算。根据稀疏表示理论的预测，只要特征空间维度超过一定阈值，非传统的特征（比如下采样图像（downsample images）和随机投影（random projection））和传统的特征（比如Eigenfaces和Laplacianfaces）表现得一样好。由于遮挡和腐蚀而导致的误差关于标准（像素）正交基通常是稀疏的，因而这个框架可以一致的处理这些误差。稀疏表示的理论能够帮助预测识别算法可以处理多大面积的遮挡以及怎么选择训练图像才能最大化遮挡条件下的鲁棒性。为了证实所提出的算法的有效性和验证以上的结论，我们在公共数据库上进行了大量的实验。

关键字：人脸识别、特征提取、遮挡和腐蚀（corruption）、稀疏表示、压缩感知、 l_1 范数最小化、验证和异常值拒绝

1 引言

简约法（parsimony）作为推理的一个指导原则已经有非常悠久的历史了。一个最有名的例子就是在模型选择[1],[2]中的最小描述长度准则（MDL），它明确提出，在模型类的一个层次中能生成最紧凑表示的模型应该更适用于要做出决策的任务，比如分类。高维数据处理中的一个相关但是更简单的简约法是挑选那些只依赖于少数观察的模型，这些模型只选择一个特征的较小子集来进行分类和可视化（比如稀疏PCA[3],[4]和其他）。这种稀疏特征选择方法，在某种程度上等同于[5]和[6]中的支持向量机方法（SVM）。SYM选择相关训练样本的一个较小子集来描述类之间的决策边界。虽然这些工作仅仅包括一小部分用于推理的简约法文献，但它们确实致力于阐述一个共同的主题：它们都是用简约法作为一个原则来从训练数据中选择一个有限的特征子集或者模型，而不是直接使用训练数据对输入（测试）信号进行表示或者是分类。

简约法在人类感知中的作用也得到了人类视觉研究结果的强烈支持。研究者们最近发现，在中低水平的人类视觉功能中，视觉皮层的许多神经元对于各种具体的刺激，比如说颜色，纹理，方向，尺度甚至是视角不同的物体图像，具有选择性。考虑到这些神经元在每一个视觉阶段能形成一个关于基础信号元素的过完备字典，这些神经元关于给出的输入图像的激励过程是高度稀疏的。

在统计信号处理领域，根据基本元素或者信号原子的一个过完备字典计算稀疏线性表示的算法问题已经成为最近的研究热点[9],[10],[11],[12]。大部分研究热点都围绕着一个发现，即任何时候只要最优表示足够稀疏，它就能够通过凸优化有效的计算[9]，即使在通常情况下此类问题很难解决[13]。由此引发的优化问题和统计学中的Lasso方法[12]相似，[14]惩罚了线性组合中系数的 l_1 范数，而不是直接惩罚非零系数的个数（比如 l_0 范数）。

这些工作的初始目标本质上并不是推理和分类，而是利用可能比香农-奈奎斯特采样定理限制值的更低的采样频率对信号进行表示和压缩[14]。因此算法的效果可以根据表示的稀疏性和原始输入的保真度来衡量。此外，字典中单个的基本元素并不会有任何特别的语义，它们都是从标准基（比如傅里叶、小波、Curvelet、Gabor）中选出的，甚至是从随机矩阵[11],[15]中生成。然而，最稀疏的表示自然具有判别力：在基向量的所有子集中，它选择能够最紧凑表示输入信号的子集，并且拒绝其他所有可能没有那么紧凑的表示子集。

本文中，我们提出了系数表示的自然判别性来进行分类。我们用一个基本元素是训练样本自身的过完备字典来表示测试样本，而不是用上述提到的一般的字典。如果每个类的训练样本足够多，就有可能将测试样本表示为来自同一类的训练样本的线性组合。这个表示自然是稀疏的，因为表示集只包含了总体训练数据的一小部分。在很多有趣的问题上，我们提出，测试样本在过完备字典上的最稀疏线性表示可以有效利用 l_1 范数最小化得到。寻找最稀疏表示，从而在训练集中不同的类找出测试样本所属的类。图1用人脸识别作为一个例子阐述了这个简单的想法。稀疏表示也提供了一个简单但是及其有效的的方法来拒绝不属于训练数据中任何类的无效测试样本：这些样本的最稀疏表示往往包含跨越多个类的大量字典元素。

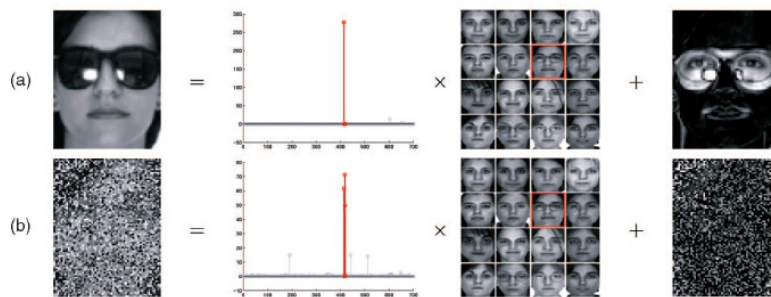


Fig. 1. Overview of our approach. Our method represents a test image (left), which is (a) potentially occluded or (b) corrupted, as a sparse linear combination of all the training images (middle) plus sparse errors (right) due to occlusion or corruption. Red (darker) coefficients correspond to training images of the correct individual. Our algorithm determines the true identity (indicated with a red box at second row and third column) from 700 training images of 100 individuals (7 each) in the standard AR face database.

我们对用来分类的稀疏性的使用和上述提到的各种简约法则有着显著的不同。我们的方法把每个单独测试样本的稀疏表示直接用于分类，并自适应的挑选有着最紧凑表示的训练样本，而不是使用稀疏性来得到一个之后能够用来区分所有测试样本的相关模型或者特征。已经提出的分类器可以被归纳为两种流行的分类器：NN（nearest neighbor）[18]和NS（nearest subspace）[19]（子空间中关于来自每一个目标类的所有训练样本的最小距离）。NN分类器是基于单个训练样本的最好的表示来区分测试样本的，而NS是基于每一类所有训练样本的最好的线性表示来分类。NFL（nearest feature line）算法[20]是NN和NS两个极端的折中办法，基于一部分训练样本的最好的仿射表示来进行分类。我们的方法同样也是折中，但却考虑了所有可能的交集（在每一个类中或者交叉在多个类中）并且自适应的选择了所需的最小数量的训练本来表示每一个测试样本。

我们将在自动人脸识别的环境下激励和研究这个分类的新方法。人脸无疑是基于图像的识别领域中最广泛的研究对象。这在一定程度上是由于人类视觉系统的卓越的人脸识别能力[21]以及人脸识别技术的广泛而又重要的应用[22]。除此之外，与人脸识别相关的技术问题一般而言就是目标识别和数据分类的代表。反过来，稀疏表示和压缩感知理论使得我们对自动人脸识别上有了更深入的理解，这两个关键问题分别是：特征提取的作用以及由于遮挡所导致的困难。

特征提取的作用。“一个目标图像的哪个低维特征对于分类是最有价值的或者能提供最多信息？”这个问题通常是人脸识别以及目标识别的中心问题。已经有大量的文献致力于研究各种依赖数据的特征变换，用于把高维测试数据投射到低维特征空间：比如Eigenfaces[23]，Fisherfaces[24]，Laplacianfaces[25]以及许多变种（variants）[26]，[27]。对于这么多已经提出的特征，孰好孰坏几乎没有一致的看法，使用者们缺乏指导原则来决定使用哪一个特征。然而，在我们提出的框架中，压缩感知理论表明：具体选择哪一个特征空间已经变得不再至关重要。甚至随机的特征也包含了足够的信息来获得稀疏表示，从而能够正确的区分任何测试图像。特征空间的维度是否足够大才是关键的，只有这样稀疏表示才能够被正确计算。

对遮挡的鲁棒性。遮挡对于稳健的真实人脸识别造成了相当大的阻碍[16]，[28]，[29]。这个困难主要是由于遮挡所导致的误差存在不可预测的自然性：它可能影响到图像的任何一部分并且可能是任意量级的。然而，这个误差通常只腐蚀图像像素的一小部分，因此在单个像素给出的标准基下是稀疏的。当误差有这样一个稀疏表示时，在我们的框架中，它能够被一致的处理：这些用于误差稀疏表示的基可以被视为是训练样本中一个特殊的类。随后，在这个扩展的字典上（训练样本+误差基）对一个有遮挡的测试图像进行稀疏表示，自然的就把测试样本遮挡部分的成分和测试对象的主体成分分离（如图1）。在这个背景下，稀疏表示和压缩感知理论描述了什么时候这个主体和误差分离可以发生，从而推断出识别算法可以容忍多大面积的遮挡。

关于这篇文章的组织。在第2部分，我们提出了一个基本的通用框架用于基于稀疏表示的分类，这个框架可以应用于基于图像的目标识别的各种各样的问题。我们也将讨论为什么稀疏表示能够通过 l_1 范数最小化计算，并且稀疏表示怎么样才能被用于对任何给出的测试样本进行分类和验证。第3部分展示了怎么样运用这个通用的分类框架来研究基于图像的人脸识别领域中两个重要的问题：特征提取和对遮挡的鲁棒性。在第4部分，我们在流行的人脸数据集上用大量的实验证明了我们所提出的方法的有效性，并和其他最先进的人脸识别方法进行比较。我们所提出的方法和NN、NS的进一步关系会在增补的附录中所讨论到，相关内容请访问计算机协会数字图书馆：<http://doi.ieeecomputersociety.org/10.1109/TPAMI.2008.79>。

虽然所提出的方法对于总体意义上的目标识别有着广泛的前景，但是本文的研究和实验结果仅局限于正面的人脸识别。我们将会处理光照和表情，但是我们并没有明确的考虑到目标姿势或者是依赖人脸的3D模型。再比如，由于配准误差，所提出的算法对于姿势和位移的小变化具有鲁棒性。然而，我们需要假设在应用我们的算法之前，人脸的检测、裁剪和标准化已经预先被执行。

2 基于稀疏表示的分类

目标识别中的一个基本问题是，用来自 k 个不同的带有标签的训练样本来正确的判定新的测试样本属于哪一类。我们把给出的来自第 i 类的 n_i 个训练样本表示为矩阵 $A_i = [v_{i,1}, v_{i,2}, \dots, v_{i,n_i}] \in \mathbb{R}^{m \times n_i}$ ， $v_{i,1}, v_{i,2}, \dots, v_{i,n_i} \in \mathbb{R}^m$ 。在人脸识别的背景下，我们会将一张 $w \times h$ 的灰阶图像定义为向量 $v \in \mathbb{R}^m$ ($m = wh$)，这个向量是用图像的各列堆叠出来的；从而 A_i 中的列向量表示就是第 i 个目标类的所有训练人脸图像。

2.1 测试样本表示为训练样本的稀疏线性组合

大量的统计学的模型、生成模型和判别模型被提出，利用 A_i 的结构来进行识别。一个及其简单但是有效的方法就是利用来自单个类的样本在线性空间上建模。子空间模型对于抓取真实数据集中尽可能多的变量足够灵活，尤其在人脸识别的环境中（在这个环境中，不同光照下和表情的人脸图像位于一个特殊的低维子空间（通常叫做人脸子空间）[24]，[30]）能够被很好的激励。尽管所提出的框架和算法能够应用于多模型或者是非线性分布（在增补的附录中了解更多细节，请访问计算机协会数字图书馆 <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2008.79>），为了使表示更轻松，我们首先应该假设来自于单个类的训练样本确实处于同一个子空间上。这将是我们会在结论中用到的唯一的一个关于训练样本的预备知识。

由于测试样本属于哪个样本在初始条件下是未知的，我们为整个训练集定义了一个新的矩阵 $A \in \mathbb{R}^{n \times k}$ 来表示来自 k 个目标类所有的 n 个训练样本： $A = [A_1, A_2, \dots, A_K] \in \mathbb{R}^{n \times k}$ ，其中 $A_i \in \mathbb{R}^{n \times 1}$ 表示第 i 类的所有训练样本。然后 y 关于所有训练样本的线性表示可以写成：

$$y = Ax, \quad x \in \mathbb{R}^m, \quad y \in \mathbb{R}^n, \quad A \in \mathbb{R}^{n \times k}, \quad x \in \mathbb{R}^m, \quad y \in \mathbb{R}^n$$

其中， $x = [x_1, x_2, \dots, x_k]^T \in \mathbb{R}^k$ ， $y = [y_1, y_2, \dots, y_n]^T \in \mathbb{R}^n$ ， $A = [A_1, A_2, \dots, A_K] \in \mathbb{R}^{n \times k}$ ， $A_i \in \mathbb{R}^{n \times 1}$ 表示第 i 类的所有训练样本。向量 x 中除了和 i 类相关的元素是非零元素之外，其他的元素都是 0。

显然, 当 $m \gg n$ 时, 方程组 $y = Ax, y = Ax, y = Ax$ 是超定的, 如果该方程组 $y = Ax, y = Ax, y = Ax$ 存在精确解, 那么它一定是唯一的。然而, 我们将会在第3部分看到, 在稳健的人脸识别中, 方程组 $y = Ax, y = Ax, y = Ax$ 往往是欠定的, 也就是解不唯一。按照惯例, 这个问题可以通过选择最小 $\|x\|_2$ 范数解来解决:

以上结论促使我们通过解决以下优化问题来寻找 $y = Ax$ $y = Ax$ 的最稀疏的解:

$$(5) \min \|x\|_0 \text{ subject to } Ax = y, \quad (10) \min \|x\|_0 \text{ subject to } Ax = y, \quad (10) \min \|x\|_0 \text{ subject to } Ax = y.$$

其中 $\|x\|_0$ 表示向量 x 中非零元素的个数。事实上，如果 A 中的列处于一般的位置， $y = Ax$ 的一些解 x 中，只要 x 中的非零元素小于 m ， x 就是唯一的稀疏解：即 $x^0 = x$ 。然而，求解一个欠定线性方程组从而寻找最稀疏的解是一个 NP 难问题，即使求大约解 [13] 也很难：也就是说，在一般情况下，没有比遍历元素的所有子集更高效的方法来寻找 x 最稀疏的解。

稀疏表示和压缩感知新兴理论[9],[10],[11]最近的发展表明, 如果 $x_0 \in \mathbb{R}^n$ 被认为是足够稀疏的, $\min_{x \in \mathbb{R}^n} \|x\|_1$ 最小化问题的求解可以等同于以下的 $\min_{x \in \mathbb{R}^n} \|x\|_1$ 最小化问题的求解:

(6) $\min_{x \in \mathbb{R}^n} \|x\|_1$ subject to $Ax = y, \|x\|_1 \leq \tau$ (7) $\min_{x \in \mathbb{R}^n} \|x\|_1$ subject to $Ax = y, \|x\|_1 \leq \tau$ (8) $\min_{x \in \mathbb{R}^n} \|x\|_1$ subject to $Ax = y, \|x\|_1 \leq \tau$ (9)

这个问题可以通过线性编程方法在多项式时间内求解[34]。当解非常稠密时, 甚至会有更高效的解决方法。比如说, homotopy算法求得有 t 个非零元素的解时, 时间复杂度为 $O(t^3 + n) O(t^3 + n) O(t^3 + n)$, 在训练集大小上是线性的[35]。

2.3 基于稀疏表示的分类 (SRC)

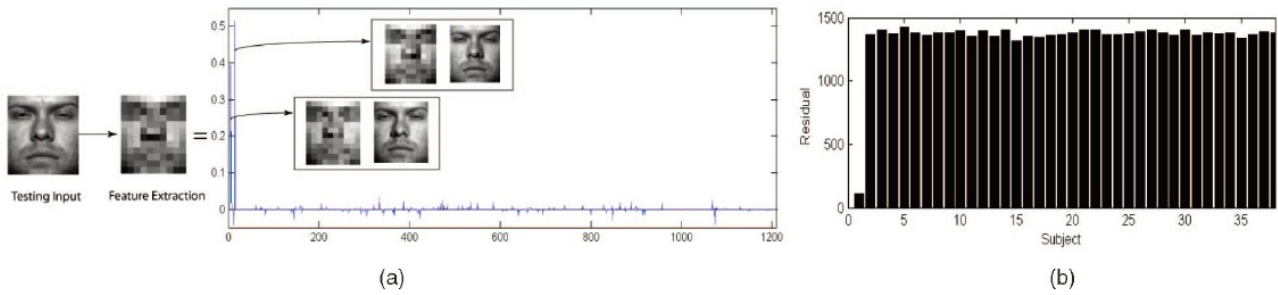


Fig. 3. A valid test image. (a) Recognition with 12×10 downsampled images as features. The test image y belongs to subject 1. The values of the sparse coefficients recovered from Algorithm 1 are plotted on the right together with the two training examples that correspond to the two largest sparse coefficients. (b) The residuals $r_i(y)$ of a test image of subject 1 with respect to the projected sparse coefficients $\delta_i(\hat{x})$ by ℓ^1 -minimization. The ratio between the two smallest residuals is about 1:8.6.

从训练集的一个类中给出一个新的测试样本 y ，我们首先通过 (6) 和 (10) 来计算 y 的稀疏表示 $x^1 \hat{x}^1$ 。理想状况下，所预测的稀疏表示 $x^1 \hat{x}^1$ 中的非零元素将仅仅和 $A A^T$ 的部分列关联（即这些非零元素是这些列前面的系数），这些列都来自于同一个目标类 i ，然后我们可以很轻松的把测试样本 y 归为这一类。然而，噪声和模型误差可能导致一小部分非零元素与不同目标类关联（如图3）。基于全局的稀疏表示，我们可以设计很多可能的分类器来解决这个问题。比如，我们可以简单的将 y 归为与 $x^1 \hat{x}^1$ 中最大元素相关联的目标类中。但是，这种方法并没有利用到与人脸识别中的图像相关的子空间结构。为了更好的利用这样的线性结构，我们通过比较 $x^1 \hat{x}^1$ 中与哪个目标类中所有训练样本的相关联的元素能够更好的重构 y ，来对 y 分类。

对于每一类 i ，令 $\delta_i: \mathbb{R}^n \rightarrow \mathbb{R}^n$ 为一个描述函数，作用是选择只和第 i 类相关联的系数。对于 $x \in \mathbb{R}^n$ ， $\delta_i(x) \in \mathbb{R}^n$ ， $\delta_i(x)$ 中所有的非零元素是向量 x 中只与目标类 i 相关联的元素。使用只与第 i 类相关联的系数，我们可以大概重构出给出的测试样本 y ，记为 $y^i = A \delta_i(x^1 \hat{x}^1)$ 。然后我们就可以根据哪个类能够使得 y 和 y^i 之间的残差最小，来对 y 进行分类。

$$(12) \min_i \text{iri}(y) = \|y - A \delta_i(x^1 \hat{x}^1)\|_2 / 2. \quad \text{miniri}(y) = \|y - A \delta_i(x^1 \hat{x}^1)\|_2 / 2. \quad (12)$$

以下的算法1概括了完整的识别步骤。我们的执行通过一个基于[39]和[40]，用于线性编程的对偶算法来最小化 $\|y - A \delta_i(x^1 \hat{x}^1)\|_2$ 范数。

Algorithm 1. Sparse Representation-based Classification (SRC)

- 1: Input: a matrix of training samples $A = [A_1, A_2, \dots, A_k] \in \mathbb{R}^{m \times n}$ for k classes, a test sample $y \in \mathbb{R}^m$, (and an optional error tolerance $\varepsilon > 0$.)
- 2: Normalize the columns of A to have unit ℓ^2 -norm.
- 3: Solve the ℓ^1 -minimization problem:
$$\hat{x}_1 = \arg \min_x \|x\|_1 \text{ subject to } Ax = y. \quad (13)$$
(Or alternatively, solve
$$\hat{x}_1 = \arg \min_x \|x\|_1 \text{ subject to } \|Ax - y\|_2 \leq \varepsilon.$$
)
- 4: Compute the residuals $r_i(y) = \|y - A \delta_i(\hat{x}_1)\|_2$ for $i = 1, \dots, k$.
- 5: Output: identity(y) = $\arg \min_i r_i(y)$.

例1 (ℓ^1 最小化 vs ℓ^2 最小化)

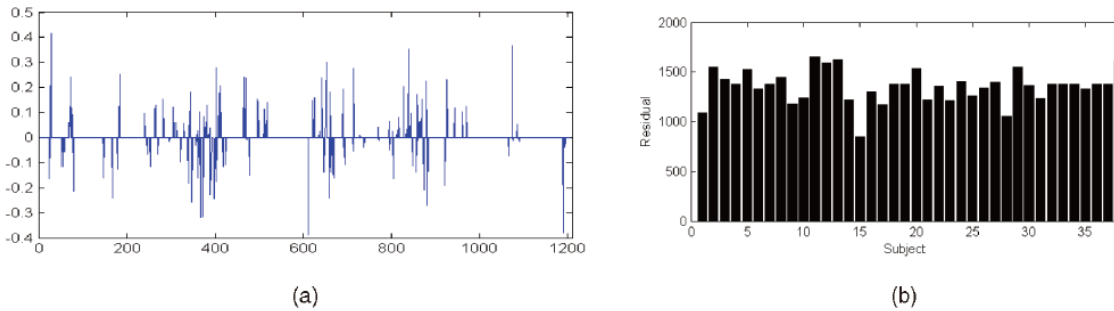


Fig. 4. Nonsparsity of the ℓ^2 -minimizer. (a) Coefficients from ℓ^2 -minimization using the same test image as Fig. 3. The recovered solution is not sparse and, hence, less informative for recognition (large coefficients do not correspond to training images of this test subject). (b) The residuals of the test image from subject 1 with respect to the projection $\delta_i(\hat{x})$ of the coefficients obtained by ℓ^2 -minimization. The ratio between the two smallest residuals is about 1:1.3. The smallest residual is not associated with subject 1.

为了阐述算法1是怎么工作的，我们随机选择了Extended Yale B 数据库中2414张图像的一半作为训练集，剩下的部分用于测试。在这个例子中，我们对所有的图像进行下采样（subsample），使得每一张图片的大小从初始的 192×168 降到了 12×10 。下采样图像的所有像素值构成一个120维的特征，堆叠成算法中 $A A^T$ 的列。因此，矩阵 $A A^T$ 的大小为 120×1207 ，并且方程组 $y = A x$ 是欠定的。图3a展示了一个来自第1类的测试图像通过算法求出的稀疏系数。该图也展示了对于两个最大系数的特征图像和原始图像。这两个最大的系数都来自于目标类1中的训练样本。图3b展示了38个类的投影系数 $\delta_i(x)$ ， $i = 1, 2, \dots, 38$ 的残差。使用 12×10 的下采样图像作为特征，算法1在Extended Yale B 数据库总体上来说能够达到92.1%的识别率。（如果需要了解更多细节，算法使用另外一些特征（如：Eigenfaces和Fisherfaces）的表现以及和其他算法的比较，请见第4部分）。鉴于方程组 $y = A x$ 更加传统的 ℓ^2 范数解往往是非常稠密的，最小化 ℓ^1 范数往往能够找到稀疏解，并且当解是足够稀疏的话，被证明能够找到最稀疏的解。为了阐明最小化 ℓ^2 范数和最小化 ℓ^1 范数的差异，图4a展示了同样的测试样本的系数，通过传统的 ℓ^2 最小化（4）求出，图4b展示了关于38个目标类的残差。很明显，图4中的系数远远没有由 ℓ^1 最小化求出的系数（见图3）稀疏，并且图4中最主要的系数并不与目标1相关联。因此，图4中残差最小的类并不是正确的目标类（目标类1）。

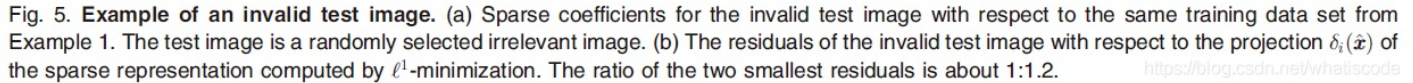
2.4基于稀疏表示的验证

在为给出的测试样本进行分类之前，我们必须确认测试样本是否是一个来自数据集中任意一个类的有效样本。检测然后拒绝无效测试样本（或者是奇异值）的能力对于一个在真实环境下工作的识别系统来说是至关重要的。比如一个人脸识别系统，可能会被给出一个不是数据库中任何一个目标的人脸图像或者一张甚至都不是人脸的图像。

基于传统分类器比如NN或NS的系统，通常使用残差 $r_i(y)$ 进行检验和识别。也就是说，算法基于最小的残差来接受和拒绝一个测试样本。然而每一项残差 $r_i(y)$ 的计算和训练数据集中的其他目标类没有任何关系，并且仅仅测量了测试样本和每一个单独目标类的相似性。

在稀疏表示范式中，系数 $x^1 \hat{x}^1$ 是根据所有类的图像全局计算的。在一定程度上，它能够利用所有类的联合分布来进行验证。我们主张系数 $x^1 \hat{x}^1$ 对于验

例2 (稀疏系数的集中)



为了量化这个发现，我们通过以下定义来衡量系数在数据集中单个类上的集中度：

(14) $\text{SCI}(x) \equiv k \cdot \max_i \|\delta_i(x)\| / \|x\| / 1 - 1 \cdot k - 1 \in [0, 1]$. $\text{SCI}(x) \equiv k - 1 \cdot k \cdot \max_i \|\delta_i(x)\| / \|x\| / 1 - 1 \in [0, 1]$. (14)

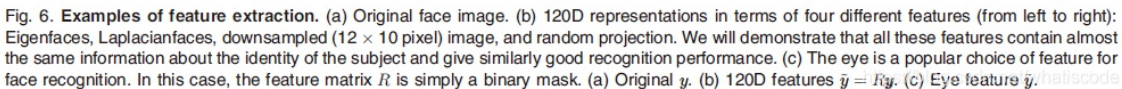
对于一个通过算法1求得的解 $x^* \in \mathcal{X}$, 如果 $\text{SCI}(x^*) = 1$, 测试样本就仅仅通过只来自一个类的图像表示, 如果 $\text{SCI}(x^*) = 0$, 稀疏系数就均匀的分布在所有的类中。我们选择了一个阈值 $\tau \in (0, 1)$ 并且当

(15) $\text{SCI}(\hat{x}) \geq \tau$, $\text{SCI}(\hat{x}) \geq \tau$, (15)
时接受测试样本，否则就当做无效样本拒绝。在算法1的第5步中，只有当测试样本通过了这个检验之后我们才能够输出 yyy 的分类结果。

这个验证方法的一个优点是,在验证一个和很多目标类相似的一般目标时,也有着很好的验证效果。比如,在人脸识别中,一个不属于数据集的普通人脸可能和数据集中的一些目标相当相似,并且关于这些目标类的残差很小。使用残差来验证很有可能导致一个错误的判断。然而,由于一个普通人脸的表示往往需要数据集中大量目标的作用,所以它不可能通过我们的新的验证规则。因此,新的规则能够更好的判定测试样本是一个普通人脸还是数据集中的特定目标。在4.7部分,我们会详细说明新的验证规则比NN和NS方法在给出一个无效样本后的验证率要提高10%到20%。(见第4部分的图14或者增补附录中的图118,见计算机协会数字图书馆: <http://doi.ieeeecomputersociety.org/10.1109/TPAMI.2008.79>)

这一部分，我们研究以上通用的识别框架如何应用到两个人脸识别领域的关键问题上：1) 特征变换的选择；2) 对于腐蚀、遮挡和伪装的鲁棒性。

在计算机视觉的文献中,大量的特征提取的方法被研究,以寻找能够在低维空间(通常指人脸空间)更好的分离不同类的投影。一类方法提取全局的人脸特征,如: Eigenfaces [23]、Fisherfaces [24]和Laplacianfaces [25]。另一类方法试图提取一部分有用的人脸特征(比如眼睛和鼻子周围的一小块) [21], [41] (一些例子见图6)。传统意义上来说,当特征提取被用于连接像NN和NS这样简单的分类器,特征变换的框架对算法的成功运行来说就很关键。这可能导致一大部分逐渐复杂的特征提取方法的发展,如非线性特征和核特征[42], [43]。这一部分,我们在新的基于稀疏表示的人脸识别框架中重新检查了特征提取的作用。



由于大部分特征变换都只包括了线性运算（或者接近于线性运算），从图像空间到特征空间的投影可以表示为一个矩阵 $R \in \mathbb{R}^{d \times m}$ ，其中 $d \ll m$ 。对方程组（3）两边同时使用 R 进行投影，得到：

$$(16) \quad y \sim \doteq R y = R A x_0 \in \mathcal{R} d \backslash \tilde{\text{tildel}}\{y\} \backslash \text{doteq} R y = R A x_0 \in \mathbb{R}^d \backslash \text{tag}\{16\} \quad y \sim \doteq R y = R A x_0 \in \mathcal{R} d(16)$$

实际应用中, 特征空间的维度 d 通常要远小于 n 。在这种情况下, 方程组 $\tilde{y} = R A x \in \mathbb{R}^d, \tilde{y} = R A x \in \mathbb{R}^d$ 对于未知的 $x \in \mathbb{R}^n, x \in \mathbb{R}^n$

$x \in \mathbb{R}^n$ 是欠定的。尽管如此, 由于期望解 $x_0^* x_0^* x_0^*$ 是稀疏的, 我们可以通过对以下简化的 $\|x\|_1$ 最小化问题进行求解从而得到 $x_0^* x_0^* x_0^*$:

$$(17) \text{ (lr1) } : x^* = \arg\min \|x\| \text{ subject to } \|R_A x - y\| \leq \varepsilon. \quad (\text{lr1}) \quad \hat{x} = \arg\min_{\|x\| \leq 1} \|R_A x - \tilde{y}\|$$

其中 ε 为误差容忍度 $\varepsilon>0$ 。因此，在算法1中，训练图像的矩阵 $A \in \mathbb{R}^{d \times n}$ 被 d 维特征矩阵 $RA \in \mathbb{R}^{d \times n}$ 代替；测试样本 x 的特征 \tilde{x} 所代替。

对于我们用于识别的稀疏表示方法,我们将会说明特征提取 RRR 的选择是怎样影响 $\|1111\|$ 最小化 (17) 求得正确稀疏解 x_0 的能力的。从 $\|1111\|$ 最小化的几何解释 (见 2.2.1 部分) 来看,以上问题的答案依赖于相关的新轴体 $P = R A (P_1) P = R A (P_1) P = R A (P_1)$ 是否保持足够的邻度。很容易看出,轴体 $P = R A (P_1) P = R A (P_1) P = R A (P_1)$ 的邻度随着 d 的增加而增加 [11], [15]。在第 4 部分,我们的实验结果将会证实 $\|1111\|$ 最小化适用到各种特征求得人脸识别稀疏表示的能力,尤其是对于稳定版本 (17)。这也意味着大部分人脸识别领域中流行的基于数据的特征 (如: eigenfaces 和 Laplacianfaces) 确实能够得到高度邻近的轴体 P 。

对高维几何轴体的进一步的分析揭示了一些惊人的发现: 如果解 $x_0 x_0 x_0$ 足够稀疏, 并且 d 是一个足够大的数, 那么有极大的概率可以通过 l_1 最小化正确的求出线性方程组 $y \sim R A x_0$ 的解 $y_i = R A x_0$ 的解。更精确来说, 如果 $x_0 x_0 x_0$ 有着 t 个非零元素 ($t \ll n$) ($t \ll n$) ($t \ll n$), 那么只要:

随机线性衡量就足够让 L1/L2 最小化求出正确的稀疏解 $x_0 \times 0 \times 0$ [44]。这个惊人的发现被称为“维度的祝福”[15], [46]。随机的特征可以被看做是经典人脸特征（如：Eigenfaces

2017年12月15日 星期五 15:00

和Fisherfaces) 没那么结构化的等同体。据此，我们可以将根据高斯 (Gaussian) 随机矩阵生成的线性投影称为Randomfaces。

定义2 (randomfaces) 考虑到一个变换矩阵 $R \in \mathbb{R}^{d \times m}$ 在 $\mathbb{R}^{d \times m}$ 中的元素是从一个均值为0的正态分布中独立采样的，并且每一行都被标准化为单位长度。 R 的行向量可以被看做是 m 张随机人脸。

一个Randomfaces的主要优点是它们能够很高效的被生成，因为变换 R 是独立于训练数据集的。这个优点对于人脸识别系统来说是非常重要的，因为我们可能并不能获得所有目标类一个完整的数据集来预先计算依赖数据的变换 (如Eigenfaces)，或者数据集的目标可能随时会发生改变。在这种情况下，我们就不需要再次计算随机变换 R 。

只要能够求出正确的稀疏解 x_0 ，不管使用什么样的特征，算法1将会始终给出相同的识别结果。因此，当特征的维数 d 满足以上的限制 (18)，使用不同特征的算法1都会快速聚集，并且是否选择一个“最优的”特征变换变得不再至关重要：即使是随机的投影或者是下采样图像都应该和其他精心设计的特征表现得一样好。这一点将会在第4部分的实验结果中被证实。

3.2对于遮挡和腐蚀的鲁棒性

在很多可行的人脸识别方案中，测试图像 y 可以可能被部分的腐蚀或者遮挡。在这种情况下，上面的线性模型 (3) 可以被改进为：

$$(19) y = y_0 + e_0 = Ax_0 + e_0, y = y_0 + e_0 \tag{19}$$

其中 $e_0 \in \mathbb{R}^m$ 是一个误差向量——其一小部分 (ρ) 元素为非零向量。 e_0 模型中的非零向量是 y 中被腐蚀或者是被遮挡的像素。不同测试图像有着不同的腐蚀位置并且计算机对这些位置是未知的。这些误差可能有着任意量级因此不能够被忽略或者当做2.2.2部分的小噪声来对待。

编码理论[52]中一个重要的原则是测量结果中的冗余度对于检测和更正过失误差很重要。冗余提高了目标识别的效果，原因是图像像素的数量通常要比形成图像的目标多得多。在这种情况下，即使一小部分像素完全被遮挡，基于剩下像素进行识别仍然是可能实现的。另一方面，在之前部分中讨论的特征提取方法将会去掉那些可能帮助消除遮挡的信息。这种情况下，没有任何一种表示比原始的图像更加冗余、稳健并且能够提供更多的有用信息。因此，当处理遮挡和腐蚀时，我们应该使用尽可能高的分辨率，只有当原始图像的分辨率太高以至于不能够处理时，我们才使用下采样或者是特征提取。

当然，如果冗余没有有效的计算工具来利用冗余数据中的有效编码信息，那么它也就没有用处。直接处理原始腐蚀图像的困难已经导致了研究者们转向利用空间局部性作为稳健识别的一个指导原则。局部特征通过一小部分图像像素进行计算，这些像素上由于遮挡所造成的腐蚀显然要比全局特征少得多。在人脸识别领域，像ICA[53]和LNMF[54]这类方法利用以上的发现来自适应的选择局部集中的过滤器基准。由于局部二值模式 (LBP) [55]和Gabor小波[56]都是在局部图像区域上计算的，因此它们有着相似的特性。一个相关的方法把图像分成固定的区域并且分别计算每一个区域的特征[16], [57]。值得注意的是，局部集中的基上的投影只是转化了遮挡问题的计算域，但并没有消除遮挡。原始像素上的误差变成了转化后的计算域上的误差并且可能变得没有那么具有局部性。由于没有基或者特征要比原始图像像素本身更加具有空间局部性，因而特征提取在利用空间局部性中的作用因此让人质疑。事实上，使得基于空间的方法变得稳健的最流行的办法就是在单个像素上进行随机采样[28]，有时也利用多元切尾法[29]来协助。

现在，让我们展示怎么扩展我们提出的稀疏表示分类框架以处理遮挡问题。首先假设被腐蚀的像素是图像中相当小的一部分。误差向量 e_0 就像向量 x_0 一样，有着稀疏的非零向量。由于 $y_0 = Ax_0$ ，我们可以把 (19) 改进为：

$$(20) y = [A, I] \begin{bmatrix} x_0 \\ e_0 \end{bmatrix} = Bw_0, y = [A, I] \begin{bmatrix} x_0 \\ e_0 \end{bmatrix} = Bw_0 \tag{20}$$

因此， $B = [A, I] \in \mathbb{R}^{m \times (n+m)}$ ，所以方程组 $y = Bw$ 通常是欠定的并且没有唯一的解 w 。然而，从之前关于 x_0 和 e_0 的稀疏性的讨论，正确的 $w_0 = [x_0, e_0]$ 应该最多有 $n + \rho$ 个非零元素。我们因此希望可能求出方程组 $y = Bw$ 最稀疏的解 w_0 。事实上，对于一般情况下的矩阵 B ，只要 $y = Bw$ 中的 w 有着少于 $m/2$ 的非零元素，那么 w 就是唯一的稀疏解。因此，如果遮挡 e_0 遮挡了少于 $m - n/2$ 的像素， $\approx 50\%$ 的图像，稀疏解 w_0 对于方程组 $y = Bw$ 就是一个有效的值， $w_0 = [x_0, e_0]$ 。

更一般的情况下，我们可以假设腐蚀误差 e_0 关于一些基 $A_e \in \mathbb{R}^{m \times n_e}$ 有着稀疏表示。也就是说，对于一些稀疏向量 u_0 $e_0 = A_e u_0$ 。在此，我们选择了一些特殊的例子：假设 e_0 关于自然像素坐标稀疏，则 $A_e = I \in \mathbb{R}^{m \times m}$ 。如果误差 e_0 关于其他一些基 (比如傅里叶和Haar) 更加稀疏，我们可以简单的把矩阵 B 重新定义为在 A 后面加上 A_e (而不是直接使用 I)，因而我们可以转而通过以下等式寻找最稀疏的解 w_0 ：

$$(21) y = Bw \text{ with } B = [A, A_e] \in \mathbb{R}^{m \times (n+n_e)}, y = Bw \tag{21}$$

这样，相同的公式就可以处理更加普遍的腐蚀 (稀疏) 类。

像之前那样，我们试着去寻找最稀疏的解 w_0 来解决以下的扩展的 ℓ_1 最小化问题：

$$(22) (e_1) : w^1 = \arg \min \|w\|_1 \text{ subject to } Bw = y, (e_1) : w^1 = \arg \min \|w\|_1 \text{ subject to } Bw = y \tag{22}$$

$$(e_1) : w^1 = \arg \min \|w\|_1 \text{ subject to } Bw = y, (e_1) : w^1 = \arg \min \|w\|_1 \text{ subject to } Bw = y \tag{22}$$

也就是在算法1中把图像矩阵 A 换成了扩展的矩阵 $B = [A, I]$ 并且把 x 换成了 $w = [x, e]$ 。

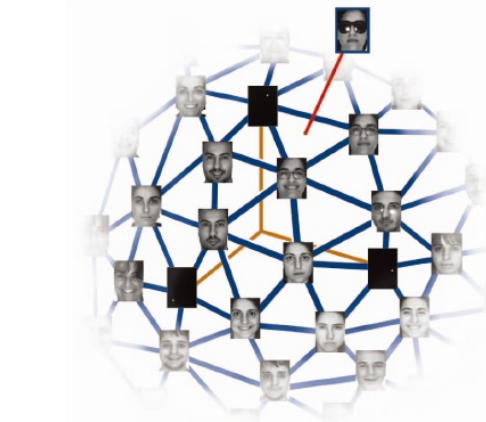


Fig. 7. Face recognition with occlusion. The columns of $\pm B = \pm [A, I]$ span a high-dimensional polytope $P = B(P_1)$ in \mathbb{R}^m . Each vertex of this polytope is either a training image or an image with just a single pixel illuminated (corresponding to the identity submatrix I). Given a test image, solving the ℓ_1 -minimization problem essentially locates which facet of the polytope the test image falls on. The ℓ_1 -minimization finds the facet with the fewest possible vertices. Only vertices of that facet contribute to the representation; all other vertices have no contribution.

显然稀疏解 w_0 是否能够从以上的 ℓ_1 最小化中求出依赖于新多面体 $P = B(P_1) = [A, I](P_1)$ 的邻度。如图7所示，这个新多面体包括了来自训练图像 A 和单位矩阵 I 的顶点。 (8) 中给出的限制条件意味着如果 y 是一个目标类 i 的图像，那么 ℓ_1 最小化并不能保证正确的求解出 $w_0 = [x_0, e_0]$ ，除非：

$$n + |\text{support}(e_0)| \geq d/3, n + |\text{support}(e_0)| \geq d/3$$

通常情况下， $d \geq n$ ，因此 (8) 意味着我们希望达到完美重构效果的前提下，最大的遮挡分量为33%。这个限制条件也将会在我们的实验结果中被证明 (见图12)。

为了精确求出多大的遮挡能够被容忍，除了 (8)，我们需要关于多面体 P 邻度更多的精确信息。比如，我们想要知道对于一个给出的训练图像，能够处理的最大数量 (最坏的可能下) 的遮挡是多少。已知用来精确计算一个多面体的邻度的最好算法是自然组合，通过将零空间 B 和 ℓ_1 -球交集的搜寻限制在 ℓ_1 -球的一个随机子集上，我们可以得到更紧的上限 (更多细节见[37])。我们将使用这个方法来自估计所有将在我们实验中用到的训练数据集的邻度。

根据经验，我们发现稳定的版本 (10) 只有当我们不考虑模型中的遮挡和腐蚀 e_0 时 (比如之前部分所提到的特征提取实例)，才有用。当我们通过使用 $B = [A, I]$ 和 $B = [A, I]$ 显性的考虑粗差时，使用精确限制 $Bw = y$ 范数最小化 (22) 在温和的误差下是稳定的。

一旦稀疏解 $w^1 = [x^1, e^1]$ 被求出， $y = Ax^1 + Ae^1$ 被求出， $y = Ax^1 + Ae^1$ 就能够表示一张消除了遮挡和噪声的干净目标图像。为了识别这个目标，我们对算法1中的残差 $r_i(y)$ 进行了如下的轻微改进：

$$(23) r_i(y) = \|y - A\delta_i(x^1)\|_2 = \|y - Ae^1 - A\delta_i(x^1)\|_2 = \|y - Ae^1 - A\delta_i(x^1)\|_2 \tag{23}$$

4 实验验证

在这一部分，我们在公共数据库上进行了人脸识别实验，以证明所提出的识别算法的有效性和验证之前部分的结论。我们首先将会检查在我们的框架中特征提取的作用，并

比较不同种类的特征空间和特征维数的效果，以及比较不同的比较流行的分类器。然后我们将会证明所提出的算法对于腐蚀和遮挡的鲁棒性。最后，我们会证明（使用ROC曲线）稀疏性的作为一个验证测试图像的方法的有效性并且检验怎么样选择训练集能够最大化对遮挡的鲁棒性。

4.1特征提取和识别方法

我们使用了几个传统的全局人脸特征，即Eigenfaces、Laplacianfaces和Fisherfaces测试了我们的SRC方法并且和两个非传统的特征：Randomfaces和下采样图像比较了它们的效果。我们将我们的算法和3个经典的算法，即之前部分中提到的NN、NS以及线性的SVM进行了比较。在这一部分，我们使用了SRC的稳定版本在各种低维的特征空间上，使用了误差容忍度 $\epsilon = 0.05$ \(\epsilon = 0.05\) 来解决简单优化问题（17）。算法1的简单（特征空间）版本在经典的3-GHz电脑上通过MATLAB实现，在每个测试图像上只花费了几秒钟。

4.1.1 Extended Yale B 数据库

Extended Yale B数据库由38个人的2414张正面人脸图像组成。裁剪和标准化后的 192×168 人脸图像是在各种实验室控制的灯光条件下拍摄的[59]。对于每一个目标，我们随机的选择了一半的图像用于训练（大概每个人32张图像），另一半用于测试。随机选择训练集以保证我们的结果和结论将不会依赖于任何训练数据的特殊选择。

我们计算了特征空间维度分别为30、56、120、和504时的识别率。这些数字分别对应下采样率为1/32、1/24、1/16和1/8。值得注意的是，由于最大数量的有效Fisherfaces特征数比类数 k （在本例中为38）少1，因此Fisherfaces和其他特征不同。因此，在我们的实验中，只有当维度为30时，Fisherfaces的识别结果才有效。

NS算法的子空间维度是9，这已经在只有光照变化的人脸图像处理文献中被广泛接受。图8展示了四种不同分类器：SRC、NN、NS和SVM在不同特征下的识别效果。

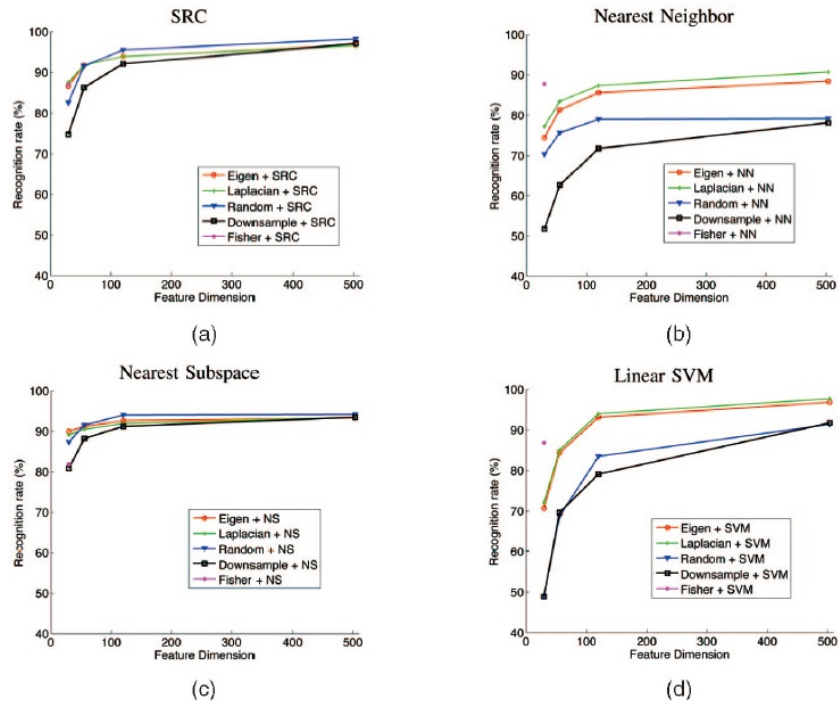


Fig. 8. Recognition rates on Extended Yale B database, for various feature transformations and classifiers. (a) SRC (our approach). (b) NN. (c) NS. (d) SVM (linear kernel).

<https://blog.csdn.net/whaliscode>

对于所有的120维特征空间，SRC的识别率在92.1%和95.6%之间，并且在504维的Randomfaces特征上有着最高的识别率——98.1%。NN、NS以及SVM最大的识别率分别为90.7%、94.1%以及97.7%。所有识别率的表格能够在增补附录中（详见计算机协会数字图书馆 <http://doi.ieeeecomputersociety.org/10.1109/TPAMI.2008.79>）进行查询。图8中所展示的识别率对于文献中已经被报道出来的方法是连续的，尽管一些报道出来的方法是基于不同的数据库或者是不同的训练集。比如，He等人[25]提出了在Yale人脸数据库中，当维度为33时使用Eigenfaces的识别率为75%，当维度为28时使用Laplacianfaces的识别效率为89%。在[32]中，Lee等人提出了在Yale B数据库上使用NS方法的精确识别率为95.4%。

4.1.2 AR数据库

AR数据库由126个人的4000张正面人脸图像组成。对于每一个人，26张图像被分为两部分[60]。这些图像比Extended Yale B数据库包含更多的人脸变量，比如光照变化、表情变化和人脸遮挡。在这个实验中，我们选择了由50个男人和50个女人组成的数据集。对于每一个对象，只选择了有光照变化和表情变化的14张图像：7张图像作为第一部分用于训练，剩下的7张作为第二部分用于测试。这些图像被裁剪成维度为 165×120 并且被转化为灰度图像。我们选择了4个特征空间维度：30、54、130以及504，分别对应下采样率为1/24、1/18、1/12和1/6。因为对象的个数为100，Fisherfaces的结果分别在维度为30和54时给出。

由于目标数变成了100，在这个数据库的实验本质上比Yale数据库上的实验要更具挑战性，但是训练图像被缩减为每个目标7张图像：4张有着不同光照条件的不带表情的脸以及3张不同表情的脸。对于NS，由于每个目标训练样本的数量为7，任何人脸上空间的估计都不能大于7。在本例中，我们选择了保持所有的7维。

图9展示了该实验的识别率。当使用504维特征时，SRC的识别率在92.0%和94.7%之间。另一方面，NN和NS的最好的识别率分别为89.7%和90.3%。SVM在这个数据集中表现得稍微比SRC要好，最高的识别率达到了95.7%。然而，SVM的效果在选择不同的特征空间时差别很大——使用随机特征时的识别率只有88.8%。增补附录中包含了一个数值结果的表格（详见计算机协会数字图书馆 <http://doi.ieeeecomputersociety.org/10.1109/TPAMI.2008.79>）。

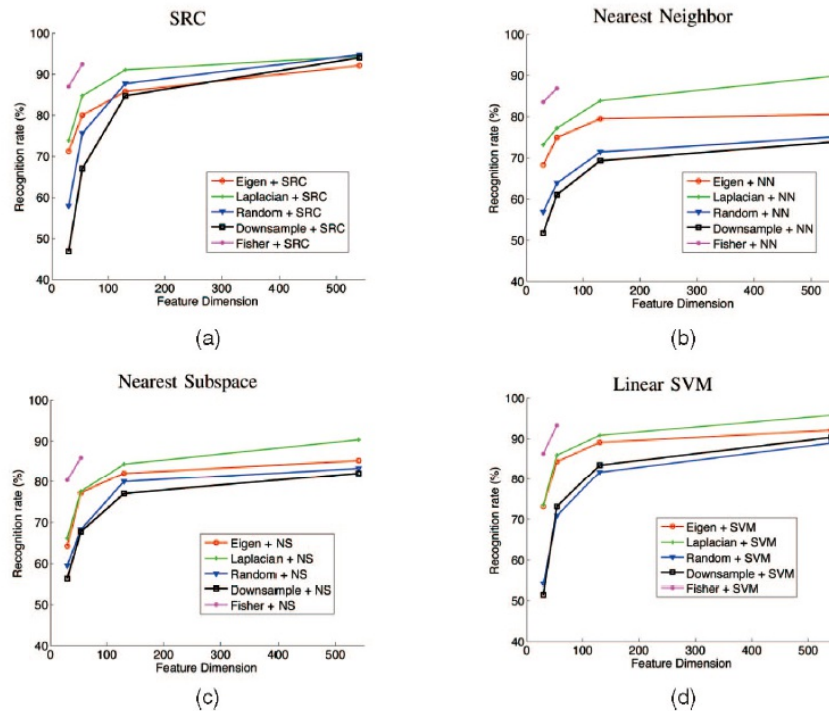


Fig. 9. Recognition rates on AR database, for various feature transformations and classifiers. (a) SRC (our approach). (b) NN. (c) NS. (d) SVM (linear kernel).

<https://blog.csdn.net/whafiscode>

基于Extended Yale B数据库和AR数据库的实验结果，我们可以得出以下结论：

1. 在Extended Yale B数据库和AR数据库上，任何特征维度下，SRC和SVM的最好效果都要优于NN和NS这两种经典方法的最好效果。更具体的说，SRC在Yale数据库上的最高识别率为98.1%，SVM为97.7%，NS为94.0%，NN为90.7%；SRC在AR数据库上的最高识别率为94.7%，SVM为95.7%，NS为90.3%，NN为89.7%。
2. 其他三个分类器的效果非常依赖于最优特征的选择——Fisherfaces用于低维空间；Laplacianfaces用于高维空间。NN和SVM中，不同特征的效果并不随着特征空间维度的增加而聚集。
3. 实验结果证明了压缩感知理论：（18）意味着在Yale数据库上，当 $d \approx 128$ 时，随机线性衡量对于求出稀疏解已经足够；在AR数据库上，当 $d \approx 88$ 时，随机线性衡量对于求出稀疏解已经足够[44]。超过这些维度，使用 ℓ_1 最小化的不同特征的效果就会聚集，传统的特征和非传统的特征（如：Randomfaces和下采样图像）有着相似的表现。当特征维度很大时，单个的随机投影表现最好（Yale数据库上的识别率为98.1%，AR数据库上的识别率为94.7%）。

4.2 局部人脸特征

在人类和计算机视觉文献中，已经有大量关于验证局部人脸识别的有效性的研究，如[21]和[41]。作为实验的第二部分，我们在三个局部人脸特征上测试了我们的算法：鼻子、右眼和包含嘴巴和下巴的部分。我们在Extended Yale B数据库上使用了和4.1.1部分相同的测试集进行了实验。从图10中可以看到一个典型的特征提取的例子。

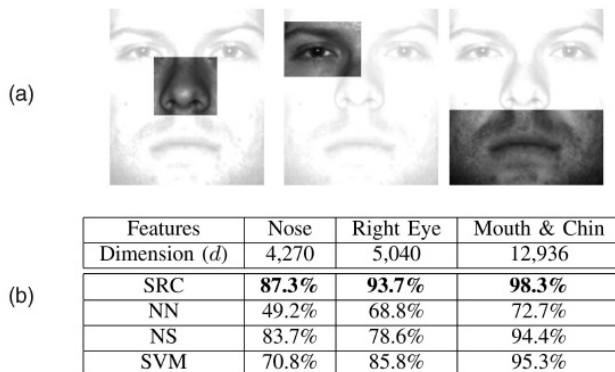


Fig. 10. Recognition with partial face features. (a) Example features. (b) Recognition rates of SRC, NN, NS, and SVM on the Extended Yale B database.

<https://blog.csdn.net/whafiscode>

对于三个特征中的任意一个特征，维度 d 要比训练样本的数量要大 ($n=1207$)，因而需要求解的线性方程组 (16) 变成了过定的。然而，稀疏近似解 x 仍然能够通过解决 ϵ -相关的 ℓ_1 最小化问题 (17) 得到 (在此， $\epsilon = 0.05$)。图10的结果再一次正确的证明了所提出的SRC算法的识别率比NN、NS和SVM更高。这些实验也证明了所提出的算法在超过 10^4 维特征上的可扩展性。

4.3 在随机像素腐蚀上的识别

对于这个实验，我们在Extended Yale B数据库上测试了SRC的稳定版本，这个稳定版本解决了扩展的 ℓ_1 最小化问题 (22)。我们使用子集1和子集2 (717张图像，正常到适度的光照条件) 用于训练，子集3 (453张图像，极端的光照条件) 用于测试。在没有遮挡的情况下，这是一个相当容易的识别问题。为了隔离出遮挡的作用，选择必须经过谨慎考虑。图像被裁剪为 96×84 ，因此在这个例子中， $B = [A, I]$ 是一个 8064×8761 的矩阵。对于这个数据集，我们估计多面体 $P = \text{conv}(\pm B)$ 。

我们从每一个测试样本中随机选择了一部分像素进行腐蚀，把这些像素的值换成了均匀分布中的独立同分布样本。腐蚀的像素是从每一个测试样本中随机选择的，因此这些像素的位置对于算法是未知的。我们腐蚀像素的比例从0%到90%。图11a、11b、11c和11d展示了一些测试图像的样本。当超过50%的腐蚀比例，腐蚀的人脸图像就已经不能够被人眼所识别 (如图11a的第二幅和第三幅图像)，识别它们似乎就成为了不可能的问题。然而，即使是在这种极端情况下，SRC仍然能够正确的识别目标。

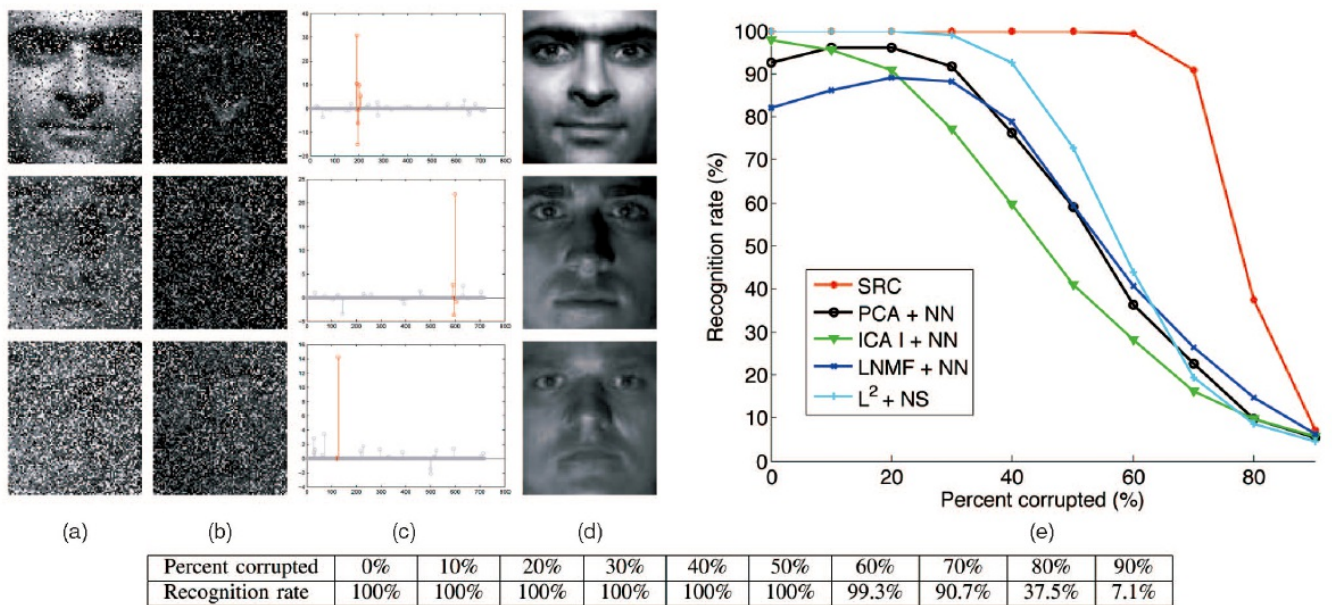


Fig. 11. **Recognition under random corruption.** (a) Test images y from Extended Yale B, with random corruption. Top row: 30 percent of pixels are corrupted. Middle row: 50 percent corrupted. Bottom row: 70 percent corrupted. (b) Estimated errors \hat{e}_1 . (c) Estimated sparse coefficients \hat{x}_1 . (d) Reconstructed images \hat{y} . SRC correctly identifies all three corrupted face images. (e) The recognition rate across the entire range of corruption for various algorithms. SRC (red curve) significantly outperforms others, performing almost perfectly upto 60 percent random corruption (see table below).

<https://blog.csdn.net/whatiscode>

我们定量的将我们的方法和视觉文献中的其他4种人脸识别方法进行了比较。[23]中的主成分分析法（PCA）对于遮挡并没有鲁棒性。有很多改进使得PCA对于腐蚀或者是不完整数据具有了鲁棒性，并且有的改进已经被应用到了稳健的人脸识别，如[29]。我们稍后将会讨论我们的算法在实际应用条件下的效果。在此，我们使用基本的PCA来提供一个比较的标准。剩下的3中方法则对于遮挡更具有鲁棒性。独立成分分析（ICA）构架I[53]试图去用一个统计学的独立基图像的线性组合来表示训练集。局部非负矩阵分解（LNMF）使用基图像的加法组合来近似表示训练集，并偏向于使用稀疏基进行运算。最后，为了说明所提高的鲁棒性确实是来自于 l_1 范数的使用，我们和最小二乘法（首先把测试样本投影到由所有人脸图像构成的子空间，然后再执行NS）进行了比较。图11e给出了SRC和它的4个竞争者的识别效果在不同腐蚀程度下的函数图。我们发现SRC比其他4中算法表现得都要好。从0到50%的遮挡率，SRC正确识别了所有的目标。在50%的腐蚀下，其他的算法中没有有一个算法的识别率是高于73%的，而SRC的识别率是100%。甚至在70%的遮挡下，SRC的识别率仍然能够达到90.7%。这远远超过了理论上最差情况下算法能够容忍的腐蚀值——13.3%。显然，最差情况下的分析对于随机腐蚀过于保守。

4.4 在随机遮挡块上的识别

我们接下来模拟不同程度的邻近遮挡，从0到50%，通过将每个测试样本中随机位置的方块换成一个不相关的图像，如图12a所示。相类似的，遮挡的位置随机选择因此对于计算机是未知的。选择人脸特征或者是图像块的方法（如[16]和[57]），由于遮挡位置的不可预测性变得没有那么有效。图12a、12b、12c和12d的前两行展示了算法1在30%遮挡下的两个表示结果。图12a是遮挡图像。在第二行中，人脸的整个中心部分都被遮挡；这对于人类都是一个很难完成的识别任务。图12b展示了估计误差 \hat{e}_1 的向量。值得注意的是， \hat{e}_1 不仅由遮挡图像构成，还包含了由于噪声阴影所造成的线性子空间模型的异常。图12c画出了系数向量 \hat{x}_1 的预测值。红色的元素是对应于正确测试样本的系数。在所有的例子中，所估计的系数是稀疏的并且对于同一个人的测试图像有着很大的量级。在所有例子中，SRC算法正确的识别了遮挡图像。对于整个数据集，在PowerMac G5电脑通过MATLAB进行的实验对于每个测试图像需要90秒的测试时间。

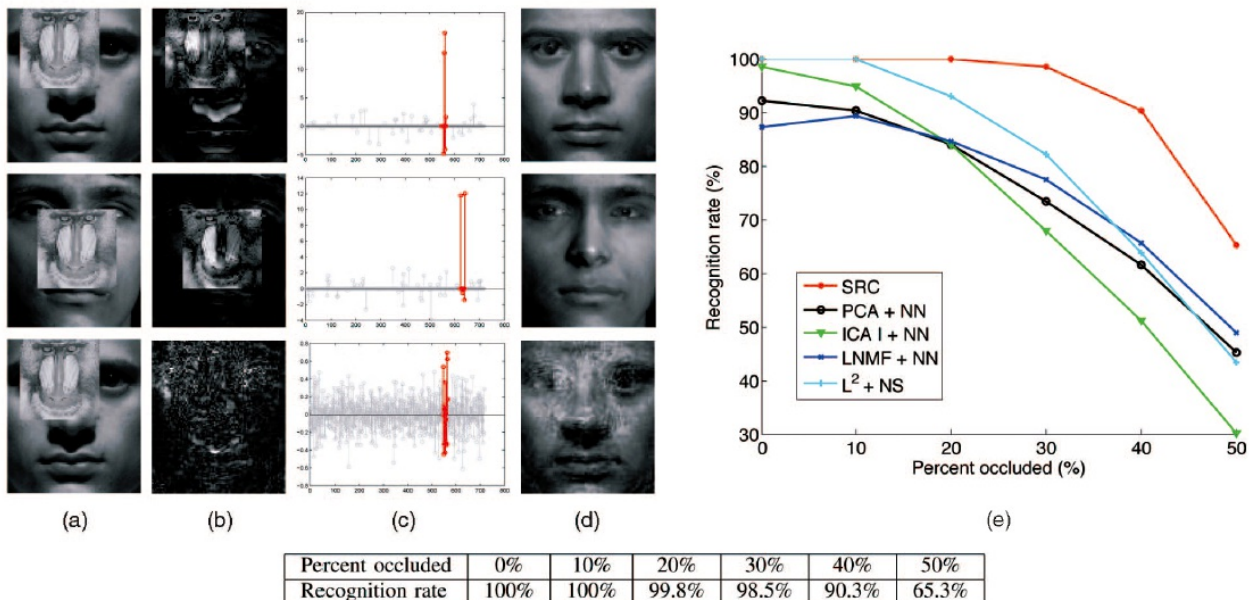


Fig. 12. **Recognition under varying level of contiguous occlusion.** Left, top two rows: (a) 30 percent occluded test face images y from Extended Yale B. (b) Estimated sparse errors \hat{e}_1 . (c) Estimated sparse coefficients \hat{x}_1 , red (darker) entries correspond to training images of the same person. (d) Reconstructed images \hat{y} . SRC correctly identifies both occluded faces. For comparison, the bottom row shows the same test case, with the result given by least squares (overdetermined ℓ^2 -minimization). (e) The recognition rate across the entire range of corruption for various algorithms. SRC (red curve) significantly outperforms others, performing almost perfectly up to 30 percent contiguous occlusion (see table below).

图12e中的函数图展示了所有5个算法的识别率。SRC再一次比其他4个算法在所有不同遮挡程度上都要表现得更好。在30%的遮挡上，算法1表现得几乎完美，正确识别了98%的测试目标。即使是在40%的遮挡上，只有9.7%的目标未能识别成功。与随机像素的腐蚀相比，相邻的遮挡对于算法肯定是一种更坏的误差。尽管如此，要注意的是，我们的算法并没有假设有任何的自然腐蚀或遮挡的先验知识。在4.6部分，我们将会看到邻近遮挡的先验知识是如何被用于改进算法并且显著提高识别效果的。

这个实验结果对于在人脸识别中应该使用全局特征还是局部特征的讨论[22]有着有趣的影响。由于他们的基是局部集中的，遮挡只腐蚀了一小部分系数，所以这意味着ICA和LNMF对于遮挡都具有鲁棒性。相反，如果使用 l_2 最小化（垂直投影）来根据像训练样本本身的全局基表示遮挡图像，那么所有的系数都可能被腐蚀（如图12第3行）。这个后果所带来的问题将会是怎样被计算，而不是选择全局还是局部基表示测试样本。合适的利用冗余和稀疏性是误差更正和鲁棒性的关键。提取局部或者是不相交的特征

4.5 在伪装上的识别

我们把200张测试图像分成了两个测试集。第一个测试集包含了带墨镜的目标图像，墨镜大概遮挡了图像20%。图1a展示了这个测试集中的一个成功实例。注意到对于由墨镜所导致的遮挡， $\hat{e}^1 \nabla \text{hat}[c]_1 e^1$ 由图像边缘很小的误差构成。第二个测试集包含了带着围巾的目标图像，围巾大概遮挡了图像的40%。由于这个遮挡率大概是能够由邻域 $\text{conv}(\pm B) \text{conv}(\text{pmB}) \text{conv}(\pm B)$ 所给出的最坏情况下最大遮挡率的3倍，我们的算法可能在这个范围内并不能成功。图13a展示了一个失败的案例。图中最大的系数对应的胡莱男的图像中，他的嘴巴部分和围巾相似。

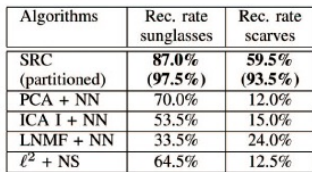


Fig. 13. (a)-(d) Partition scheme to tackle contiguous disguise. The top row visualizes an example for which SRC failed with the whole image (holistic). The two largest coefficients correspond to a bearded man and a screaming woman, two images whose mouth region resembles the occluding scarf. If the occlusion is known to be contiguous, one can partition the image into multiple smaller blocks, apply the SRC algorithm to each of the blocks and then aggregate the results by voting. The second row visualizes how this partition-based scheme works on the same test image but leads to a correct identification. (a) The test image, occluded by scarf. (b) Estimated sparse error \hat{e}_1 . (c) Estimated sparse coefficients \hat{x}_1 . (d) Reconstructed image. **Table of recognition rates on the AR database.** The table shows the performance of all the algorithms for both types of occlusion. SRC, its holistic version (right top) and partitioned version (right bottom), achieves the highest recognition rate.

4.6 通过分块来提高识别效果

[illegible]

4.7 拒绝无效测试图像

我们接下来将会说明稀疏性与拒绝无效（带遮挡或者无遮挡的）测试样本的相关性。我们在Extended Yale B数据库上根据稀疏集中度（SCI）（14）测试了奇异拒绝规则（15），像之前那样，我们使用子集1和2来训练，用子集3来测试。我们再次通过用一个不相关的图像替换测试样本的随机选择的块，模拟了不同程度的遮挡（10%，30%和50%）。然而，在这个实验中，我们仅仅包括了训练集的一半目标。因此，训练集的另的一半的目标对于算法来说是新的。我们通过在上[0,1]范围内变化阈值 τ ，并生成如图14所示的受试者工作特征（ROC）曲线，来测试系统判断一个给出的测试目标是否是训练集的新目标的能力。我们也通过在PCA、ICA和LNMF特征空间中为测试图像（的特征）与最近邻训练图像（的特征）的欧几里得距离设置了阈值，以考虑异常值拒绝项。这些曲线同样也在图14中展示。值得注意的是简单的拒绝规则（15）在10%到30%的遮挡上表现得近乎完美。在50%的遮挡上，它的表现显著的要比其他3个算法要好，并且是4个算法中唯一一个比预期要好的算法。增补附录中包含了更多在AR数据库上使用Eigenfaces的验证结果，再一次说明了ROC的显著改进。

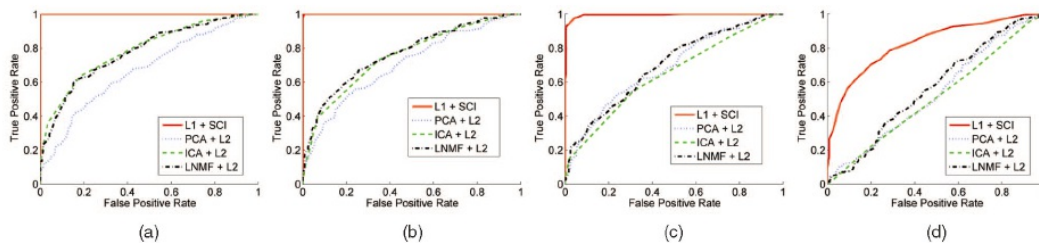


Fig. 14. ROC curves for outlier rejection. Vertical axis: true positive rate. Horizontal axis: false positive rate. The solid red curve is generated by SRC with outliers rejected based on (15). The SCI-based validation and SRC classification together perform almost perfectly for upto 30 percent occlusion. (a) No occlusion. (b) Ten percent occlusion. (c) Thirty percent. (d) Fifty percent.

4.8 设计训练样本以达到鲁棒性

设计识别系统的一个重要考虑就是选择训练样本的数目以及拍摄的条件（光照、表情、视角等等）。训练图像应该要足够大以能够包含所有可能发生在测试集中的条件：从模式识别的观点来说，它们应该是充分的。举个例子，Lee等人[59]阐述了怎么样选择最小的表示图像来大约解释每一个人脸的圆锥。在第2部分所讨论的邻度的概念为怎样才算是一个鲁棒的训练集提供了一个不同的衡量方式：算法所能够容忍的最大数量的遮挡比例直接通过相对应的多面体的邻度确定。由于遮挡目标可能和例外一个训练类的图像非常相似，所以视觉识别中最差的情况是有意义的。然而，如果遮挡是随机的并和训练图像无关（就像4.3部分），算法的平均效果也可能会有意义。

事实上，这两个重要的概念，充分性和鲁棒性是互补的。图15a展示了在Extended Yale B数据库上四个子集的邻度估计值。注意到子集4（最极端的光照条件）有着最高的邻度，大约为1330。图15b展示了AR数据库上有着不同面部表情的子集的临界点。数据集包含了如图15b所示的四种面部表情：无表情、开心、生气和尖叫。我们从各种表情图像中生成了训练集，并计算了每个训练集相对应的多面体邻度。最具鲁棒性的训练集是分别为无表情+开心的组合，接下来是开心+尖叫的组合；鲁棒性最差的是无表情+生气的组合。值得注意的是，无表情图像和生气图像表现得非常相似，而开心和尖叫非常不相似。

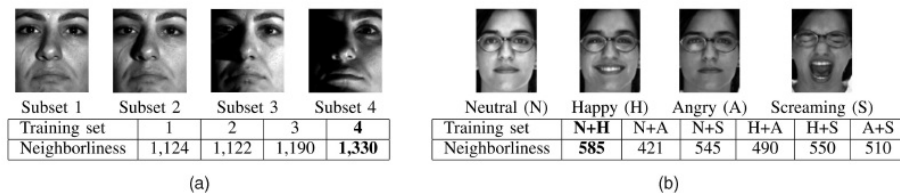


Fig. 15. Robust training set design. (a) Varying illumination. Top left: four subsets of Extended Yale B, containing increasingly extreme lighting conditions. Bottom left: estimated neighborliness of the polytope $\text{conv}(\pm B)$ for each subset. (b) Varying expression. Top right: four facial expressions in the AR database. Bottom right: estimated neighborliness of $\text{conv}(\pm B)$ when taking the training set from different pairs of expressions.

因此，对于不同的光照（图15a）和不同的表情（图15b），在图像中有着更多变量的训练集对遮挡有着更好的鲁棒性。设计一个允许在各种广泛的条件下识别的训练集并不会限制我们的算法，反而 would 帮助提高算法的效果。然而，训练集不应该包含太多相似的图像，就像在图15b的无表情+生气的例子。在信号表示的术语中，训练图像应该被称为不相关字典（incoherent dictionary）[9]。

5 结论和讨论

在这篇论文中，我们通过理论分析和实验验证说明了稀疏性的利用对如人脸图像的高维数据进行高效的分类至关重要。当稀疏性被合理的利用时，特征的选择就没有使用的特征数目重要（在我们的人脸识别例子中，大约100就足够让这些不同特征的差异变得微不足道）。除此之外，遮挡和腐蚀在相同的识别框架中能够被一致并且稳健的处理。在没有特殊工程处理的情况下，通过一个简单的算法就能够对一些遮挡或者腐蚀的图像有着显著的识别效果。

关于未来工作的一个有趣的问题是，这个框架除了识别之外，是否能够用于目标检测。稀疏性在检测的用处已经在文献[61]和最新发表的[62]中得到了注意。我们相信稀疏性在稳健的目标检测和识别中所有潜在的可能性还没有被完全发掘。从一个实际应用的立场来说，把这个算法应用到限制更少的条件中可能会有用，尤其是不同的目标姿势。对于遮挡的鲁棒性允许算法容忍小的姿势变化和误差。在增补附录中（详见计算机协会数字图书馆 <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2008.79>）我们进一步讨论了我们的算法在适应非线性训练分布的能力。然而，在不同姿势下直接表示人脸图像分布所需的训练样本的数量可能过大。在姿势中使用外推法（比如仅仅使用正面训练图像），将会需要整合特征匹配方法或者是非线性变形模型来用于测试样本的稀疏表示计算。原则上，想要实现以上的方法，仍然需要对未来工作的重要导向。