指导教师：杨涛　　提交时间：2016/3/19

CVPR2015 Paper Translation

No: 01
姓名：冯益民
学号：2013302474
班号：10011301

基于现实世界人脸识别的分级 PEP 模型

Haoxiang Li, Gang Hua
斯蒂文斯理工学院
Hoboken, NJ 07030
{hlil8, ghua}@stevens.edu

摘录

姿态变化仍然是对现实世界人脸识别准确性产生不利影响的主要因素之一。在最近提出的概率弹性部分（PEP）模型和在一系列视觉任务中深层次结构的成功的启发下，我们提出了分层-PEP 模式尝试解决无约束人脸识别难题。我们基于面部表示，分级采用 PEP 模型按照不同层次的细节将面部图像分解为脸的局部，来建立姿态不变的部分。从底向上遵循分层结构，我们在每一层堆叠面部局部表示，分别降低其维数，从而一层接一层地聚集面部局部表示层，建立一个紧凑不变的面部表示。分层-PEP 模型利用了面部局部在不同级别细节下的细粒结构来解决的姿态的变化。它也受在构建面部局部/面部表示的监视信息的指导。我们凭经验在两个公共基准和一个基于图像和视频人脸验证的人脸识别挑战中验证分层 PEP 模型。使用了最先进技术表现证明了我们方法的潜力。

1. 介绍

现实世界的人脸识别，困难源于各种视觉变化，包括表情，光照，姿态等等。姿势的变化是其中的主要挑战之一。同一张脸在不同的姿势下互相看上去截然不同，如图 1 所示，特克等人【47】和贝尔胡米尔等人【4】的早期作品，在这个区域聚焦于识别高准直性的正面人脸。他们凭经验证明人的正脸可以被投影到一个低维的在光照和面部表情上从不变到改变的子空间【4】。这项观察强调了处理姿态变化的重要性，因为它大大地有助于缓解其他视觉变化的不利影响。

一系列的研究通过生成相同视角的按姿态变化配对的面部图像来处理这个难题例如，帕布等人【40】利用三维人脸模型将人脸图像旋转到一个看不见的视野。朱等人【54】利用深层神经网络直接训练来预测从

图 1，姿态变化：相同的人从不同的姿态观察。

多视角的人脸图像到标准视角变换的方法，来恢复标准视角的人脸。使用这种方法，他们试着整体排列人脸来减轻姿态的变化。

另外一套方法凭借定位面部标记来建立姿态不变的人脸表征【10,15】。例如，陈等人【15】连接面部标记周

围密集的容貌特征来建立人脸表征。姿态不变形用这种方法实现，因为它总是从面部标记周围的面部局部抽取容貌特征而不关注它们在图像中的位置。

弹性匹配方法【20,29,30,50】归纳出这个方案。李等人【29,30】提出从面部图像碎片无监督地学习的概率弹性部分（PEP）模型。PEP 模型是局部模型和每一个隐式定义一个面部局部的混合物。PEP 模型从这些不同姿态的人脸的面部局部中寻找图像碎片。然后通过从这些筛选的图像碎片中连结容貌特征来构建人脸表征。

这个过程——定位人脸局部和堆叠容貌特征来建立人脸表征——被陈等人【15】和李等人【29】证明是有效的。在提取面部局部特征方面，陈等人【15】使用高维成堆的 SIFT 特征和李等人【29】简单使用 SIFT 容貌特征。尽管低阶的特征像 SIFT 现在的部分不变到局部变化，我们主张使用这些低阶特征的缺乏说服力的密集提取物直接描述人脸部分可能并不是最理想的。

在这项工作中，我们打算建立一个更好的人脸局部模型来构建一个改进的人脸表征。我们的方法是基于 PEP 模型【29】来构建以人脸局部表征位基础的概率弹性部分。在我们的方法中，我们在一个分层方式中使用不同级别的细节来塑造人脸局部。此外，我们通过定位人脸局部的微妙的结构和堆叠不同级别容貌特征一起来建立人脸局部表征。用这个方法，我们在一个分级结构上构建姿态不变的人脸表征。我们把这种新的模型命名为分级 PEP 模型，这种新的局部的局部的人脸表征称为 POP-PEP。

这种分级结构可能会产生一个维度很高的人脸表征。我们通过构建人脸局部表征中有差别维数的减小来避免这个缺点。此外，这种维数减小的应用来源于底部水平，这种底部水平取决于主要成分分析（PCA）和线性鉴别嵌入（LDE）【16】的简易网络下的完整人脸。这种相似的技术之前已经被西蒙尼扬【42】和张【12】等人采用。张等人【12】提出带有级联 PCA 的一个简易深度网络。西蒙尼扬等人【42】反复进行 PCA 和空间堆叠来形成一个简易深度网络。

在这项工作中，我们使用一个现实的聚集局部表征的结构，进一步完善了监督信息。我们反复堆叠精细人脸局部结构的表征，并且采用有差别的维度减小。我们凭经验验证我们实验中这种设计的可行性（详情请看第四部分）。

在这项工作中我们的贡献可以称得上是一箭三雕：

- 为了改善姿态不变，我们提出一个分级 PEP 模型来利用不同级别细节的精细人脸局部结构。
- 我们提出一个有差别的维度减小的简易网络来将人脸局部表征整合为一个紧凑而又差别的人脸表征。
- 我们在两个公共人脸验证基准和

一个人脸识别挑战上实现了最先进技术的表现。

2. 相关工作

人脸识别在过去数十年里称为一项活跃的研究。最近，设计精良的负载的人脸识别基准【24,48】和不断涌现的人脸技术应用促进了许多现实世界人脸识别方法的发展。【23,3,5,6,9,13,18,28,32,34,35,44,46】。

为了构建姿态不变人脸表征，之前的工作被建议使用 3D 信息明确地寻址姿态不变。例如，帕布【40】等人使用 3D 人脸模型将图库人脸图像旋转到调查图像的评估视角；易等人【51】使用 3D 可变模型来评估人脸姿态，并且将姿态适应性的过滤器应用到容貌特征提取；李等人【31】提出从 3D 人脸模型中学习形变置换域来合成图库人脸相同视角的探针人脸。

仅使用 2D 信息，尹等人【52】提出关联预测模型转换数据库中一个相似标记的外观来逼近看不见的姿态探针人脸的外观；陈等人【15】在人脸界标密集地提取容貌特征，并且堆叠容貌特征作为高维人脸表征。

和我们工作最相关的工作是 PEP 模型【29】。PEP 模型包含一系列通过无监督学习获得的人脸局部模型。给出一个人脸图像，每一个人脸局部模型选择最相似的图像碎片。PEP 模型接着通过连结从选择图像碎片中提取的容貌特征实现姿态不变来表示人脸图像。

在这项工作中，分级 PEP 模型分级利用人脸局部并且有区别地完善局部表征。除了产生更多有区别的人脸表征，分级 PEP 模型分享的 PEP 模型的诸多优点，如它为人脸图像和人脸视频在统一的结构下建立表征，此外，它并不需要大量的训练数据。

除了基于传统的手工容貌特征的方法，许多包括 DeepID【44】，DeepFace【46】和累计堆叠自动译码器【25】的深度学习模型成功应用于人脸识别问题，意味深远地实现了改进验证的精确度。尽管有着高精度的识别率，这些系统在训练阶段需要极大量的标记数据。

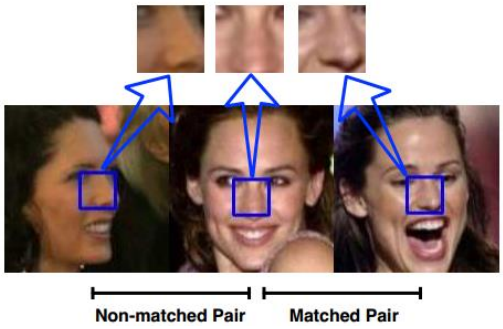
在这项工作中，我专注于理解人脸识别问题而不受大量训练数据的影响。并且我们注意到基于传统的容貌特征描述的人脸识别系统能受益于分级结构。相似的观察被西蒙尼扬等人【42】在常规图像分类工作中报道过，他们提出一种基于 SIFI 特征网的一种 2 层预付矢量编码。

3. 分级 PEP 模型

3.1. PEP 模型的介绍

分级 PEP 模型包含了一个 PEP 模型的分级结构。正式地说，PEP 模型用参数表示为

$$\mathcal{P}(L, \{\mathcal{G}_k\}_{k=1}^{K})$$



图片 2。相同人脸局部的图像碎片在可见的外表上不同。

表达式中的 P 是 PEP 模型中 K 混合物组成。每一个混合物组成是一个人脸局部模型$\mathcal{G}_k$。L 是人脸局部的尺寸。给出一个测试图像，人脸局部模型$\mathcal{G}_k$识别尺寸为 L 乘以 L 的人脸局部。更具体地说，$\mathcal{G}_k$是一个球面高斯模型。给出一个分为 N L 乘以 L 的图像碎片$\{I_n\}_{n=1}^N$的人脸图像，$\mathcal{G}_k$按最高可能性决定$I_{n^\star}$。用表达式表示为，

$$p_n = [\mathbf{a}_n\ \mathbf{l}_n], \quad (2)$$
$$\mathcal{G}_k = \mathcal{N}(p_n|\vec{\mu}_k, \sigma_k^2\mathbf{I}), \quad (3)$$

式中的p_n是图像碎片I_n的表征；$\mathbf{a}_n$是从I_n（例如 SIFT 描述）中提取的外貌特征描述；I_n是在完整人脸图像中图像碎片I_n的空间位置；$\mathbf{I}$是一个标识矩阵；$\vec{\mu}_k$和σ_k^2是分别是高斯模型的均值和方差。$\mathcal{G}_k$选择满足下式的$I_{n_k^\star}$

$$n_k^\star = \arg\max_n \mathcal{N}(p_n|\vec{\mu}_k, \sigma_k^2\mathbf{I}). \quad (4)$$

给出一个人脸图像f，PEP 模型产生人脸表征$\mathcal{F}_\mathcal{P}(f) = \mathcal{B}(\mathcal{P}, f)$，式中的$\mathcal{B}$意思是表示结构进程。更具体地说，第 k 个人脸局部模型$\mathcal{G}_k$产生人脸表征$\mathcal{F}_{\mathcal{G}_k}(f) = \mathcal{B}(\mathcal{G}_k, f)$，

$$f = \{p_n\}_{n=1}^N, \quad (5)$$
$$\mathcal{B}(\mathcal{G}_k, f) = \mathbf{a}_{n_k^\star}, \quad (6)$$
$$\mathcal{B}(\mathcal{P}, f) = [\{\mathcal{B}(\mathcal{G}_k, f)\}_{k=1}^K], \quad (7)$$

式中的$n_k^\star$表示公式 4 人脸局部模型$\mathcal{G}_k$鉴定的图像碎片容貌特征描述。PEP 模型的$\mathcal{P}$然后建立人脸表征$\mathcal{F}_\mathcal{P}(f)$作为$\mathcal{F}_{\mathcal{G}_k}(f), k = 1 \ldots K$的级联。

PEP 模型的优点之一是在一个统一的框架中处理图像和视频。给出一个 M 帧的

协方差矩阵被限制为球面来混合源于外表特征和空间位置的约束，以此来平衡两部分的影响，正如提倡的那样【29】。

人脸视频$v = \{f_m\}_{m=1}^M$，PEP 模型建立视频中人脸表征$\mathcal{F}_\mathcal{P}(v) = \mathcal{F}_\mathcal{P}(\cup_{m=1}^M f_m)$。简单地说，我们在接下来的部分仅仅使用人脸图像作为例子。视频中的人脸可以在相同的框架中处理。

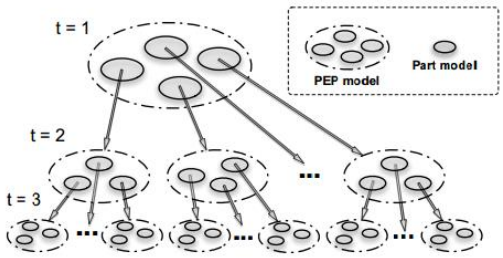


图 3。样例三层分级 PEP 模型：一个 PEP 模型的分级结构。

我们把阅读器归因于【29】PEP 模型的细节训练过程。简而言之，通过 L 和 K 的参数获得一个 PEP 模型，训练人脸图像是第一个 L 乘以 L 密集示例图像碎片的进程；外观描述从图像碎片中提取，并且它和完整人脸图像中的碎片的空间位置相连接；一个 K 要素的高斯混合模型接着通过来自特征描述训练的期望最大化（EM）算法学习称为 PEP 模型。

3.2 分级 PEP 模型

PEP 模型的有效性源于它定位人脸部分的性能。之前的工作【29,30】已经有经验地展现了 PEP 模型为人脸建立姿态不变表征。给一个人脸 f，PEP 模型建立它的表征作为一系列外观描述的连结。然而，使用低阶特征描述（比如 SIFT）描述人脸局部可能对于人脸检验不是最优的。

如图 2 所示的样例，尽管相同的人脸局部从其他三张人脸图像中准确地识别，姿态改变仍然对选择的碎片的匹配有不利影响。在这个观察的激励下，我们提出在更细密度的水平上深入应用另一个 PEP 模型，也就是说，使用更小的尺寸为 L 的图像碎片，来匹配通过人脸局部模型识别的图像碎片。我们在之前的水平上建立另一个姿态不变人脸局部 PEP 表征层来描述每一个人脸局部，而不是提取常规低水平的描述词来描述人脸局部。

一个 T 层的分级 PEP 模型在图 3 中展示（T = 3）。一个在 t 层分级 PEP $\mathcal{H}_t$ 包含以下：

1. K_t 混合要素中的 PEP 模型 $\mathcal{P}_t$ 工作于尺寸为 $L_t \times L_t$ 的人脸局部；
2. 如果 t < T，K_t 分级 PEP 模型 $\{\mathcal{H}_{t+1}^k\}_{k=1}^{K_t}$ 处于 t + 1 层。

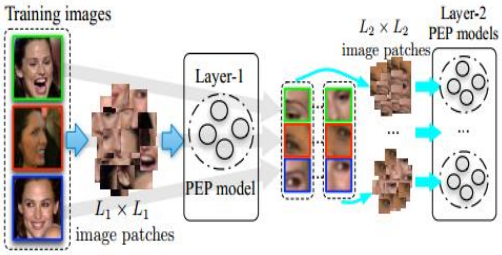


图 4. 两层分级 PEP 模型的训练过程：源于相同人脸的图像碎片颜色一致。

3.2.1 分级 PEP 模型的训练

分级 PEP 模型的训练过程在图 4 中示例。给出一套训练人脸图像 $F = \{f_i\}_{i=1}^{|F|}$，我们重复地训练 T 层分级 PEP 模型。我们首先学习一个来自于 F 的 PEP 模型 P。结合表达式 4，第 k 个人脸局部模型处理全部的 $|F|$ 训练人脸图像，和检验源于 F 的 $|F|$ 图像碎

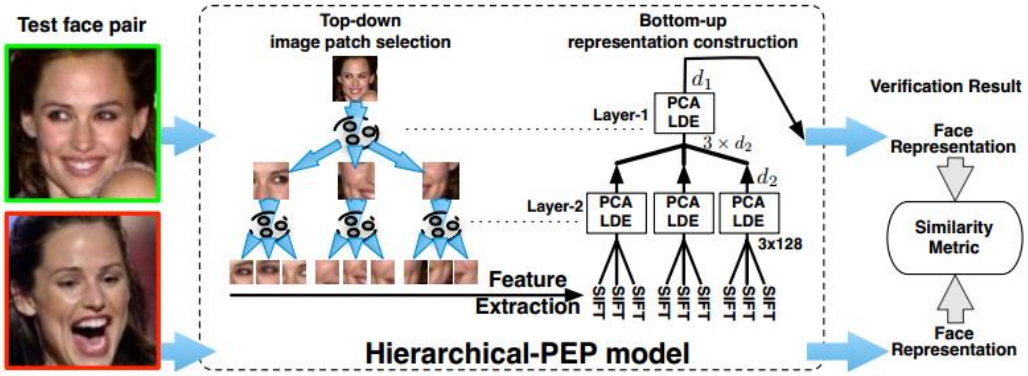


图 5.样例二层分级 PEP 模型下人脸表征结构：t 层的 PCA 保持d_t规模。

片。第 k 个人脸局部模型的这套已鉴定的图像碎片记作 。然后我们按照相同的过程来训练一个来自于 的（T - 1）层分级 PEP 模型。

3.2.2 自上而下的图片碎片选择

如图 5 所示，给一个图像 I，我们按照自上而下的工作流程，使用分级 PEP 模型来定位人脸局部和次局部。

对于分级 PEP 模型$\mathcal{H}_t$，输入图像记作I_t。图像处理为一套$L_t \times L_t$碎片，此外，根据表达式 4，K_t图像碎片被 PEP 模型$\mathcal{P}_t$中的K_t人脸局部模型识别。第 k 个人脸局部模型检验图像碎片I_{t+1}^k。如果 t < T，分级 PEP 模型$\mathcal{H}_{t+1}^k$进一步处理图像I_{t+1}^k。完整分级 PEP 模型$\mathcal{H}_1$的输入图像是完整的人脸图像 I。

如图 6 所示，一个样例自上而下的图像碎片选择产生一个人脸成对地沿着分层中的一条路径。我们可以观察到不同细节级别的弹性匹配效果。

3.2.3 自底而上的表征结构

我们按照自上而下图像碎片选择的工作流程获得所有分级 PEP 模型的输入图像，然后开始自底而上地聚集表征。

按照之前人脸表征的计数法，一个 t 层的分级 PEP 在给入输入图像I_t时，可以建立表征$\mathcal{B}(\mathcal{H}_t, I_t)$，

$$\mathcal{B}(\mathcal{H}_t, I_t) = \begin{cases} [\,\{\mathcal{B}(\mathcal{H}_{t+1}^k, I_{t+1}^k)\}_{k=1}^{K_t}\,], \text{ if } t < T, \\ \mathcal{B}(\mathcal{P}_t, I_t), \text{ if } t = T. \end{cases} \quad (8)$$

如图 5 所示（忽略之后的 PCA/LDE 图例介绍），在第二次分级 PEP 模型建立 PEP 表征（堆叠 SIFT 描述词描述人脸局部），表征被堆叠为上层表征来表示完整人脸图像。

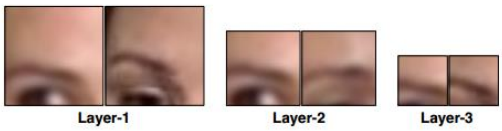


图 6. 在自上而下的图像碎片选择人脸局部的筛选过程：最左边的一对是从完整人脸图像中通过一个分级PEP第一层的人脸局部模

型选出的；中间的这对是从这个人脸局部的子结构中选出的；最右边的这对描述一个更精细的结构。我们可以观察到匹配变得更加细密并且在后面的层中更精确。

3.2.4 有区别的维数衰减

按照上述的自底而上的表征构建过程，我们建立一个人脸表征的形式 $\bar{\Pi}_{t=1}^{T} K_t \times D$ 维矢量，式中的 D 是被选中的外观特征描述其的维数，也就是说，D 等于 SIFT 描述词的 128。这种表示可以在练习中有很高的维数。因此它极大地利于减小它的高效存贮和计算的维度。

给出一套匹配的和无匹配的训练成双的人脸，李等人【30】提出使用 PCA 减小 PEP 表示法的维数，并且利用联合贝叶斯分类器【14】做检验。相同的过程也适用于分级 PEP 表示法。然而，当联合贝叶斯分类器产生一个有区别地相似度量，我们更乐意有一个有区别地人脸表征。我们采用线性判别式嵌入（LDE）【16】方法，再借助于更小的节点内的类（匹配的人脸）变化和更大的节点内的类（不匹配的人脸）变化来寻找一个子空间。

我们通过 PCA 第一次减少分级 PEP 表征的维度。然后我们寻找扩大不匹配人脸对间距离和缩小匹配人脸对间距离的子空间，

$$\bar{\mathbf{w}} = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{A} \mathbf{w}}{\mathbf{w}^T \mathbf{B} \mathbf{w}}, \tag{9}$$

$$\begin{aligned} \mathcal{F}(I_i) &= \mathbf{PCA}(\mathcal{B}(\mathcal{H}, I_i)), \\ \mathcal{F}(I_j) &= \mathbf{PCA}(\mathcal{B}(\mathcal{H}, I_j)), \\ \mathbf{A} &= \sum_{l_{ij}=0} (\mathcal{F}(I_i) - \mathcal{F}(I_j))(\mathcal{F}(I_i) - \mathcal{F}(I_j))^T, \\ \mathbf{B} &= \sum_{l_{ij}=1} (\mathcal{F}(I_i) - \mathcal{F}(I_j))(\mathcal{F}(I_i) - \mathcal{F}(I_j))^T, \end{aligned}$$

3.2.5 监督信息的整合

PCA 和 LDE 投影有区别地减小人脸表征的维度。相同的进程也适用于在人脸局部层面上构建表征。

自底而上地递归，我们堆叠局部表征，使用 PCA 和 LDE 来构建上层表征。如图 5 所示的 PCA/LDE 样例网络，监督信息分级整合到人脸表征中。

在如图 5 所示的整合过程中，我们把 PCA 和 LDE 应用到在分级 PEP 模型底层建立的 PEP 表示法。然后我们聚集低维有区别地表征到上层，而不是聚集 PEP 表征。用这种方法，所有在分级结构中聚集的表征不仅是姿态不变而且也是对人脸检验有区别的。

表达式 8 可以升级为这种过程：

$$\mathcal{B}(\mathcal{H}_t, I_t) = \begin{cases} \mathbf{DR}([\{\mathcal{B}(\mathcal{H}_{t+1}^k, I_{t+1}^k)\}_{k=1}^{K_t}]), \text{ if } t < T, \\ \mathbf{DR}(\mathcal{B}(\mathcal{P}_t, I_t)), \text{ if } t = T, \end{cases} \tag{10}$$

式中的 $\mathbf{DR}(X) = \mathbf{LDE}(\mathbf{PCA}(X))$。

3.3. 人脸验证的分级 PEP 模型

分级 PEP 模型建立了有区别的低维 POP-PEP 人脸表征。在人脸检验任务中，给出两个人脸图像 I1 和 I2，T 层分级 PEP 模型根据表达式 10 构建人脸表征 $\mathcal{B}(\mathcal{H}_1, I_1)$ 和 $\mathcal{B}(\mathcal{H}_1, I_2)$。两

个人脸相似的得分只不过是两个人脸表征的余弦相似度（或者标准化的点乘）。

$$s(I_1, I_2) = \frac{\mathcal{B}(\mathcal{H}_1, I_1)\mathcal{B}(\mathcal{H}_1, I_2)}{|\mathcal{B}(\mathcal{H}_1, I_1)||\mathcal{B}(\mathcal{H}_1, I_2)|}.$$

3.3.1 复合层合成

给一个 T 层的分级 PEP 模型我们可以裁去所有叶子 PEP 模型来得到一个（T - 1）层的分级 PEP 模型。给一个（T - 1）层的分级 PEP 模型，它构建不同细节程度的人脸表征。在之前工作的观察采用一个由粗到精的结构【17,27】启发我们可以获益于融合表征或者贯穿由粗到精结构的得分。

那就是如图 5 所示的，在建立人脸表征，我们可以按照自上而下的工作流程到上述的倒数第二个层并且从那里自底而上地聚集的获得人脸表征。更明确地说，给出人脸配对 I1 和 I2，我们可以把表达式 10 中的 T 设置为 $t' < T$，使用分级 PEP 模型来获得 t' 层的可靠得分 $s_{t'}$。人脸配对的最终可靠得分是平均得分 $s(I_1, I_2) = \frac{1}{T}\sum_{t=1}^{T} s_t$。在我们的实验中，我么你在复合层合成中观察到一致的进步。

4. 实验评估

我们从基于图像的人脸验证和基于视频的人脸验证评估了分级 PEP 模型。分享 PEP 模型的相同优点，

表 1 显示了和基准方法比较下的效果。

Algorithm	Accuracy ± Error(%)
a) 1-layer, 4096-component	89.30 ± 1.33
b) 3-layer, w/o LDE	88.00 ± 1.80

分级 PEP 模型在一个统一的框架中为人脸图像和人脸视频建立表示法。

4.1 自然环境中标记的人脸

自然环境中标记的人脸（LFW）被设计为不可控的基于图像的人脸验证的一个基准。这个资料组包含源于 5749 个人的 13233 张图像。LFW 为了公平对比定义了六条协议。在无法得到外界的训练数据的情况下，我们在有限的环境中训练我们的人脸识别系统。具体地说，我们使用没有外部数据协议的有限的图像，报告了 10 倍的平均准确度。潜在的改进可能会获得更具攻击性的环境。

4.1.1 环境

按照在 LFW 中预定义的协议，我们通过漏斗算法使用图像粗略地排列【21】。我们在图像中心裁剪尺寸为 150*150，排除大多数背景来专注于识别人脸。

我们训练一个三层的分级 PEP 模型（T = 3）。第一层包含一个 PEP 模型，这个模型里有 256 个作用在尺寸为 32*32（L1 = 32）的图像碎片的人脸局部模型（K1 = 256）。第二层包含 PEP 模型，这个模型有四个（K2 = 4）作用在尺寸为 24*24（L2 = 24）的图像碎片的人脸局部模型。最后一层包含 PEP 模型，这个模型有四个（K3 = 4）

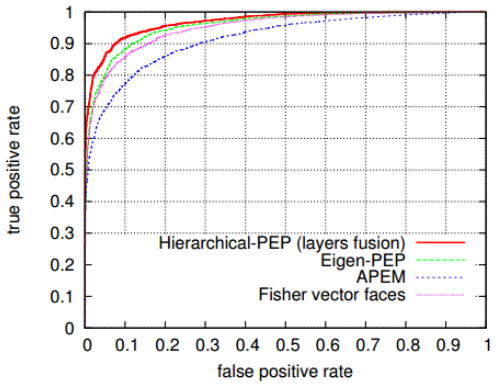


图 7.有限图像和没有外部数据协议的 LFW 的效果对比

作用在尺寸为 16*16（L3 = 16）的图像碎片的人脸局部模型。我设定 d1 = 200，d2 = 100, d3 = 50。最终的人脸表征是 200 维。我们为了公平对比，保持其他参数和本征 PEP【30】模型一致。

出于保持计算代价在典型 CPU 工作站是可接受的考虑，参数将被选择。在一个有 12 块 CPU 核心的 PC 上训练包括 PCA 和 LDE 计划的分级 PEP 模型将花费 41 小时。

4.1.2 成果

如图 7 和表 2 所示，我们观察到分级 PEP 模型实现了非常有竞争力的精确度。当李等人【30】结合 SIFT 特征和局部二进制模式（LBP）【1】来获得平局 88.97%的精确度，我们仅通过 SIFT 特征就实现 91.10%的精确度。在表 1 中，我们进一步提出一些基准结果来探究在提出的方案中每一步是怎样致力于效果改进的。

在基准实验 a）中，我们比较含有

表 3.YTF 和不同数目的二层分级 PEP 模型视频帧的效果比较。

# frames	Accuracy ± Error(%)
10	85.40 ± 1.36
50	86.84 ± 1.35
all (181 on average)	87.00 ± 1.50

4096 组件的一层分级 PEP 模型（L1 = 16, d1 = 200）。没有分级结构但是保持高斯组件的总体数量和三层分级 PEP 模型一样时，10 倍平均精确度退化。这就证明了分级的体系结构帮助提升了效果。

在基准实验 b）中，我们从自底而上的聚集中移除 LDE 但是仍然保留使维度减小的 PCA。我们仅是把 LDE 应用到最后的人脸表征。我们观察到在低层不使用 LDE 的话，它之前的方案表现不佳。这证明了在构建人脸表征时简易有区别的维度减小网络的有效性。

阿瑞斯鲁斯等人【2】通过熔接三种描述词 MLBP【11】，LPQ【45】，BSIF【26】，实现一个更高的精确度。仅使用 MLBP 描述词，它们的精确度是 90.68%，而我们的成果是仅使用 SIFT 就达到了 91.1%。此外，他们的方案依赖于马尔科夫随机字段来寻址姿态不变，这项优化过程在计算方式上非常昂贵。除了 LFW，我们也评估了在其他两个资料组（章节 4.2 和章节 4.3）的方案。

我们也评估了 POP-PEP 人脸表征的联合贝叶斯分类器来和结果【30】比较。在有限图像协议内，我们跟随李等人【30】使用标记配对人脸来训

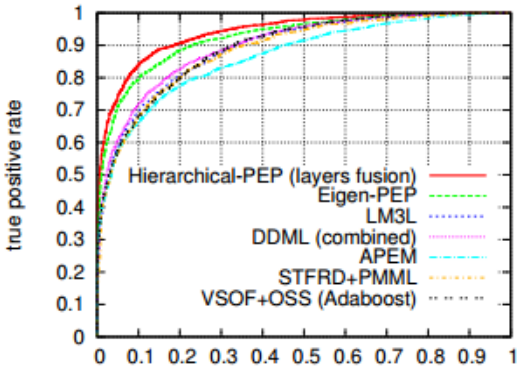


图 8. 有限和没有外部数据协议的 YTF 的效果对比。

练联合贝叶斯分类器，学习人脸配对的相似得分。然而，和简易余弦相似度比较之下我们观察不到任何改进。在章节 4.3.3，我们在 PaSC 上进行了更多的实验，来探究联合贝叶斯分类器是否能提高分级 PEP 人脸表征的识别精确度。

4.2 YouTube 人脸

YouTube 人脸（YTF）资料组【48】按照 LFW 设计为一个不可控的基于视频的人脸验证的基准。这个资料组含有 1595 人的 3425 段视频。每一个视频包含同一人的人脸。平均而言，一个视频有 181 帧。我们在有限图像和没有外部数据协议的情况下报告结构。在这个资料组中，分级 PEP 模型进一步改进了最先进技术的精确度。

4.2.1 设置

我们在视频帧中心裁剪 100*100 来排除大多数背景和减少计算量。考虑到人脸视频的低分辨率，我们训练

表 4。有限和没有外部数据协议的 YTF 的效果对比

Algorithm	Accuracy ± Error(%)
MBGS [48]	76.4 ± 1.8
MBGS+SVM- [49]	78.9 ± 1.9
STFRD+PMML [53]	79.5 ± 2.5
VSOF+OSS(Adaboost) [38]	79.7 ± 1.8
APEM [29])	79.10 ± 1.50
VF^2 [39]	84.7 ± 1.4
DDML (combined) [18]	82.3 ± 1.5
Eigen-PEP [30]	84.8 ± 1.4
LM3L [19]	81.3 ± 1.2
Hierarchical-PEP (1-layer)	86.16 ± 1.11
Hierarchical-PEP (2-layer)	86.72 ± 1.51
Hierarchical-PEP (layers fusion)	**87.00 ± 1.50**

一个两层分级 PEP 模型（T = 2）。第一层包含一个有作用于尺寸为 32*32（L1 = 32）的图像碎片的 256 个人脸局部模型（K1 =256）的 PEP 模型。第二层包含一个有作用于尺寸为 24*24（L2 = 24）的图像碎片的 256 个人脸局部模型（K2 = 16）的 PEP 模型。我们设置 d1 = 400， d2 = 200.最终人脸表征是 400 维度。其他设置和章节 4.1.1 中一样。

4.2.2 成果

如图 8 和表 4 所示，我们观察到分级 PEP 模型极大地改进了最先进技术的精确度。两层分级 PEP 模型一贯地改进一层模型的精确度，并且多层合成可以进一步改进精确度。

在表 3 中，我们展示了效果是如何提升的。通过添加更多框架在构建人脸视频表示法。我们观察到从每一个人脸视频中随机选择 10 帧，分级 PEP 模型实现了最先进技术效果。

表 5.余弦相似度和用标记的配对训练的联合贝叶斯分类器的 PaSC 的评估

Experiment	layer-1	layer-2	fusion
exp1 (cosine)	0.199	0.206	0.212
exp1 (Joint Bayesian)	0.195	0.251	0.258
exp2 (cosine)	0.259	0.284	0.299
exp2 (Joint Bayesian)	0.226	0.264	0.288

4.3 点射人脸识别挑战

贝弗里奇等人【7】提出点射人脸识别挑战（PaSC）来推进不可控视频人脸识别算法的发展。PaSC 包括 293 人的 9376 张图像和 265 人的 2802 段视频，这些与不同因素例如相机的距离，视点，传感器类型灯并重。我们提供读者更多的细节详见贝弗里奇等人【7】的报告。

PaSC 定义了两个实验，视频到视频的实验和视频到静止的实验。在视频到视频的实验中，给出目标并且分别查询视频的设置，参加者被要求报告两组成对视频到视频的相似度，并且报告在 0.01 概率错误报警信号下的验证精确度。在视频到静止的实验中，设置是相同的除了目标设置为包含静止图像而不是视频。我们评估我们在 PaSC 下的方案并且和报告的结果【8】对比。

4.3.1 设置

我们使用 PaSC 建立者提供的视力协调配比人脸，并且为了一个成对比较裁剪出 150*150 的图像。考虑到低分辨率的视频，我们在 LFW 资料组训练一个两层分级 PEP 模型（T = 2）。

表 6.PaSC 的效果对比

Algorithm	exp1	exp2
LPB-SIFT-WPCA-SILD [37]	0.09	0.23
ISV-GMM [36]	0.05	0.11
PLDA-WPCA-LLR [43]	0.19	0.26
LRPCA Baseline [7]	0.08	0.10
Eigen-PEP [30]	0.26	0.24
Hierarchical-PEP(1-layer)	0.261	0.275
Hierarchical-PEP(2-layer)	0.287	0.289
Hierarchical-PEP(2 layers fusion)	**0.307**	**0.320**

第一层包含一个有作用于尺寸为 32*32（L1 = 32）的图像碎片的 256 个人脸局部模型（K1 =256）的 PEP 模型。第二层包含一个有作用于尺寸为 24*24（L2 = 24）的图像碎片的 256 个人脸局部模型（K2 = 16）的 PEP 模型。我们设置 d1 = 100， d2 = 50. 最终人脸表征是 100 维度。其他设置和章节 4.1.1 中一样。

我们使用漏斗算法【21】粗略排列的 LFW 中的 6000 对人脸图像，来训练这个分级 PEP 模型。然后我们为 LFW 中所有 13233 张人脸图像构建人脸表示法，并且按照陈等人【14】使用它们特性标注训练一个联合贝叶斯分类器。

4.3.2 成果

我们报告了在视频到视频实验（exp1）和视频到静止实验（exp2）PaSC 中 0.01 概率错误警报信号下的验证精确度。尽管两个资料组非常不同，我们的方案展现出非常漂亮的普适性。当把我们在 LFW 训练的系统直接采用到 PaSC，我们的系统远远表现得比如表 6 所示的两组实验结果还好。

4.3.3 联合贝叶斯分类器

联合贝叶斯分类器使用两个不同的协方差矩阵塑造额外的人和内在的人的变化为零均值高斯。凭经验而论，它在人脸识别方面胜过线性判别式分析（LDA）【4】。我们将给读者提供更多细节【14】。

在 LFW 我们观察到源于标记人脸配对的联合贝叶斯分类器比得上分级 PEP 人脸表示法的余弦相似度。在 PaSC，实验结果进一步支持了这个观察。我们使用 LFW 中 6000 个标记人脸配对来训练和使用联合贝叶斯分类器的余弦相似度作比较，见表 5。联合贝叶斯分类器在视频到视频实验中胜于余弦相似度，但是在视屏到静止的实验中劣于余弦相似度。

5. 结论

我们为现实人脸识别提出一个分级 PEP 模型。 自底而上，分级 PEP 模型分级在一个统一的框架中为人脸图像和人脸视频建立姿态不变的人脸表示法。分级 PEP 模型为人脸局部和它的细粒度的结构建立姿态不变表示法。 基于局部的表示法自底而上聚集来构建人脸表示法。有监督的信息在集合的过程中整合，通过一个简易有区别的维度减小网络。分级 PEP 模型最终为人脸验证建立低维有区别的完整人脸表示法。 我们观察到一个简易复合层合成方案一贯地改进精确度。在 LFW，YTF 和 PaSC 的图像到图像，视频到视频，视频到图像人脸验证的资料组中，我们评估分级 PEP 模型。最先进技术的效果证明分级 PEP 模型的有效性。如何提升计算法来有效地采用一个更具侵略性和潜力设置仍然是一个归于我们未来工作的问题。

鸣谢

在此出版物上报告的研究得到国立卫生研究院国家护理研究所的部分支持如奖号 RO1NR015371。内容仅是作业的责任，不代表国际护理研究所的官方观点。这项工作也得到美国国家科学基金课题 IIS 1350763 和 GH 的来自于斯蒂文斯理工学院的启动资金。

引用

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. In Proc. European Conference on Computer Vision, 2004. 6

[2] S. Arashloo and J. Kittler. Class-specific kernel fusion of multiple descriptors for face verification using multiscale binarised statistical image features. Information Forensics and Security, IEEE Transactions on, 2014. 6

[3] O. Barkan, J. Weill, L. Wolf,

and H. Aronowitz. Fast high dimensional vector multiplication face recognition. In Proc. IEEE International Conference on Computer Vision, 2013. 2

[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997. 1, 8

[5] T. Berg and P. Belhumeur. Tom-vs-pete classifiers and identity-preserving alignment for face verification. In British Machine Vision Conference, 2012. 2

[6] T. Berg and P. N. Belhumeur. Poof: Part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation. In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, 2013. 2

[7] J. Beveridge, P. Phillips, D. Bolme, B. Draper, G. Givens, Y. M. Lui, M. Teli, H. Zhang, W. Scruggs, K. Bowyer, P. Flynn, and S. Cheng. The challenge of face recognition from digital point-and-shoot cameras. In BTAS, 2013. 7, 8

[8] J. R. Beveridge, H. Zhang, P. J. Flynn, Y. Lee, V. E. Liong, J. Lu, M. de Assis Angeloni, T. de Freitas Pereira, H. Li, G. Hua, V. Struc, J. Krizaj, and P. J. Phillips. The ijcb 2014 pasc video face and person recognition competition. International Joint Conference on Biometrics (IJCB)., 2014. 8

[9] X. Cao, D. Wipf, F. Wen, G. Duan, and J. Sun. A practical transfer learning algorithm for face verification. In Computer Vision (ICCV), 2013 IEEE International Conference on, 2013. 2

[10] Z. Cao, Q. Yin, X. Tang, and J. Sun. Face recognition with learning-based descriptor. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2010. 1

[11] C.-H. Chan, J. Kittler, and K. Messer. Multi-scale local binary pattern histograms for face recognition. In Proceedings of the 2007 International Conference on Advances in Biometrics, 2007. 6

[12] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma. Pcanet: A simple deep learning baseline for image classification?

arXiv preprint arXiv:1404.3606, 2014. 2

[13] B.-C. Chen, C.-S. Chen, and W. H. Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In Computer Vision–ECCV 2014. 2014. 2

[14] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun. Bayesian face revisited: A joint formulation. In Proc. European Conference on Computer Vision, 2012. 4, 6, 8

[15] D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of dimensionality: High dimensional feature and its efficient compression for face verification. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2013. 1, 2

[16] J. Duchene and S. Leclercq. An optimal transformation for discriminant and principal component analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1988. 2, 4

[17] K. Grauman and T. Darrell. The pyramid match kernel: Discriminative classification with sets of image features. In Proc. IEEE International Conference on Computer Vision, 2005. 5

[18] J. Hu, J. Lu, and Y.-P. Tan. Discriminative deep metric learning for face verification in the wild. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2014. 2, 7

[19] J. Hu, J. Lu, J. Yuan, and Y.-P. Tan. Large margin multimetric learning for face and kinship verification in the wild. In ACCV, 2014. 7

[20] G. Hua and A. Akbarzadeh. A robust elastic and partial matching metric for face recognition. In Proc. IEEE International Conference on Computer Vision, 2009. 1

[21] G. Huang, V. Jain, and E. Learned-Miller. Unsupervised joint alignment of complex images. In Proc. IEEE International Conference on Computer Vision, 2007. 6, 8

[22] G. B. Huang and E. Learned-Miller. Labeled faces in the wild: Updates and new reporting procedures. 6

[23] G. B. Huang, H. Lee, and E. Learned-Miller. Learning

hierarchical representations for face verification with convolutional deep belief networks. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012. 2

[24] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, University of Massachusetts, Amherst, 2007. 2, 6

[25] M. Kan, S. Shan, H. Chang, and X. Chen. Stacked progressive auto-encoders (spae) for face recognition across poses. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2014. 2

[26] J. Kannala and E. Rahtu. Bsif: Binarized statistical image features. In Pattern Recognition (ICPR), 21st International Conference on, 2012. 6

[27] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In Computer Vision and Pattern Recognition, 2006. 5

[28] Z. Lei, M. Pietikainen, and S. Z. Li. Learning discriminant face descriptor. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014. 2

[29] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang. Probabilistic elastic matching for pose variant face verification. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2013. 1, 2, 3, 6, 7

[30] H. Li, G. Hua, X. Shen, Z. Lin, and J. Brandt. Eigen-pep for video face recognition. In ACCV, 2014. 1, 3, 4, 6, 7, 8

[31] S. Li, X. Liu, X. Chai, H. Zhang, S. Lao, and S. Shan. Morphable displacement field based image matching for face recognition across pose. In Proc. European Conference on Computer Vision, 2012. 2

[32] S. Liao, A. Jain, and S. Li. Partial face recognition: Alignment-free approach. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013. 2

[33] D. G. Lowe. Distinctive image features from scale-invariant keypoints. International journal of computer vision, 2004. 1

[34] C. Lu and X. Tang. Learning the face prior for bayesian face

recognition. In Computer Vision－ECCV 2014. 2014. 2

[35] J. Lu, G. Wang, W. Deng, and P. Moulin. Simultaneous feature and dictionary learning for image set based face recognition. In Computer Vision－ECCV 2014. 2014. 2

[36] C. McCool, R. Wallace, M. McLaren, L. El Shafey, and S. Marcel. Session variability modelling for face authentication. Biometrics, IET, 2013. 8

[37] D. X. Meina Kan, Shiguang Shan and X. Chen. Sideinformation based linear discriminant analysis for face recognition. In British Machine Vision Conference, 2011. 8

[38] H. Mendez-Vazquez, Y. Martinez-Diaz, and Z. Chai. Volume structured ordinal features with background similarity measure for video face recognition, 2013. 7

[39] O. M. Parkhi, K. Simonyan, A. Vedaldi, and A. Zisserman. A compact and discriminative face track descriptor. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2014. 7

[40] U. Prabhu, J. Heo, and M. Savvides. Unconstrained poseinvariant face recognition using 3d generic elastic models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011. 1, 2

[41] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman. Fisher Vector Faces in the Wild. In British Machine Vision Conference, 2013. 6

[42] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep fisher networks for large-scale image classification. Advances in neural information processing systems, 2013. 2

[43] V. Struc, J. ganec Gros, S. Dobriek, and N. Pavei. Exploiting representation plurality for robust and efcient face recognition. In ERK, 2013. 8

[44] Y. Sun, X. Wang, and X. Tang. Deep learning face representation from predicting 10,000 classes. In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, 2014. 2

[45] M. Tahir, C. Chan, J. Kittler, and A. Bouridane. Face recognition using multi-scale local phase quantisation and linear regression classifier. In Image Processing (ICIP), 18th IEEE

International Conference on, 2011. 6

[46] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2014. 2

[47] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 1991. 1

[48] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2011. 2, 7

[49] L. Wolf and N. Levy. The svm-minus similarity score for video face recognition. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2013. 7

[50] J. Wright and G. Hua. Implicit elastic matching with randomized projections for pose-variant face recognition. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2009. 1

[51] D. Yi, Z. Lei, and S. Z. Li. Towards pose robust face recognition. In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, 2013. 2

[52] Q. Yin, X. Tang, and J. Sun. An associate-predict model for face recognition. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2011. 2

[53] C. Zhen, W. Li, D. Xu, S. Shan, and X. Chen. Fusing robust face region descriptors via multiple metric learning for face recognition in the wild. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2013. 7

[54] Z. Zhu, P. Luo, X. Wang, and X. Tang. Recover canonicalview faces in the wild with deep neural networks. arXiv preprint arXiv:1404.3543, 2014. 1