LFW人脸数据库说明文档

LFW(Labeled Face in the Wild)[[[1]](#endnote-1)]人脸数据库，数据库有包含来自5749个不同个体的13233张从互联网摘录的照片，其中1680人有两幅及以上的图像，4069人只有一幅图像。图像为250\*250大小的JPEG格式。绝大多数为彩色图，少数为灰度图。

大多数图像都是由Viola-Jones的检测器得到之后，被裁剪为固定大小，有少量的是人为地从false positive中得到。相片中人在不受约束的情况下被抓拍，有不同的姿态、表情、光照度和遮挡度。该数据库主要用于研究不受约束条件下的人脸识别研究。

原数据库中每个个体的图片数一般不同，最少的有一张照片，多的可达500多张。此次试验只选取这个数据库的一个子集：

本次试验我们只在LFW原有人脸数据库中随机选取来自86个个体的1251张人脸图像用做实验[[[2]](#endnote-2)]，每个个体选取有11-20张不同个数的人脸图像，并且已经将彩色图像转换为灰度图像进行试验，每张图像已经将其转换为32\*32大小的人脸图像。原图像如图1.所示，裁剪后图像如图2 所示。







图1. 经裁剪变换前的Bill\_Gates人脸图像

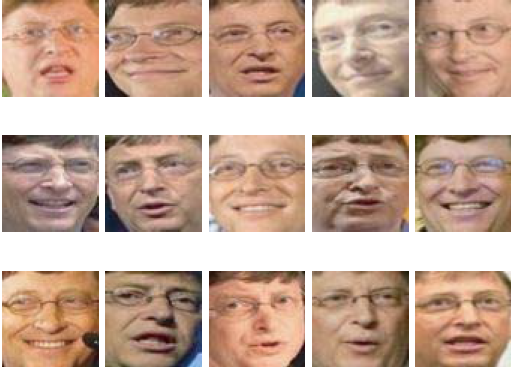


图1. 经过裁剪后的Bill\_Gates人脸图像

读取图像的matlab程序见**readData.m**文件。

readData.m **文件中有两种读取数据的方法：**

1.使用在每一个体随机取 每类的样本总数 一半向下取整个数据进行训练，剩下的数据进行测试（随机训练样本个数已给定）

2.使用每一个体的图像的前train\_num个数据进行训练，剩下的作为测试。

**文件数据说明：**

LFW.mat 人脸数据库数据，包括以下几个小子集：

（1）LFW

大小：1024\*1251

说明：每一个列向量是一个样本，大小为1024\*1；总共有1251的样本。

（2）a

大小：32\*32\*3

说明：数据库中的一个样本，大小为32\*32的三通道图像

（3）b

大小：32\*32

说明：数据库中被转化灰度图像的数据

（4）fea

说明：所有的数据总集（此次试验可以忽略）

（5）gnd

大小：1251\*1

说明：每一个样本对应的类别

（6）i 表示样本总数

（7）num 表示每一个体所包含的图像数，如:第1个个体到第16个个体包含有11个图像样本

（8）testIdx 给定随机取测试样本的标号

（9）trainIdx给定随机取训练样本的标号

**matlab程序数据说明：**

（1）train\_data 训练样本数据

train\_label 训练数据类别编号

（2）test\_data 测试样本数据

test\_label 测试数据类别编号

（3）train\_num 用于表示每个个体前train\_num个图像用于训练（随机取没有此项）

1. [] [Online]. Available: <http://vis-www.cs.umass.edu/lfw/> [↑](#endnote-ref-1)
2. [] S-J Wang, J. Yang, el al., Sparse Tensor Discriminant Color Space for Face Verification, IEEE Transactions on Neural Networks and Learning Systems, 23(6), 2012 [↑](#endnote-ref-2)