Data Visualization - Homework 1 - Report

邓淇升 大数据学院 16307110232

1.1 Implement n-dimensional joint histogram algorithm.

Python script:

1 1 n-dimensional joint histogram.py

• Input:

height vs weight.csv

Output:

height vs weight two-dimensional joint histogram.png

• Description:

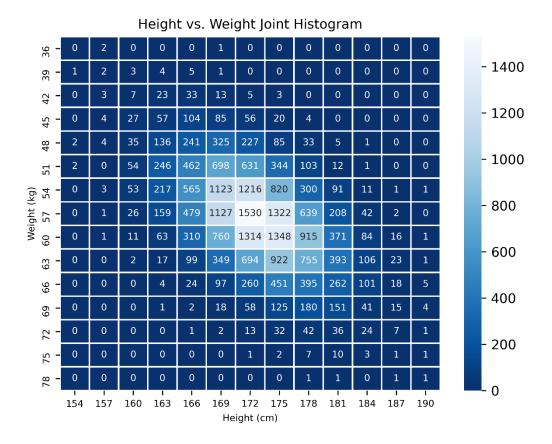
Python 源代码中给出了n维联合直方图的实现算法。假设每一维度的取值个数为 n_i , 则一共有 $\prod_i n_i$ 个直方图区域,包含了所有可能出现的行向量情况。为了计算n维联合直方 图, 需要统计每一个区域在数据中出现的次数, 并将其记录于对应的直方图区域内。算 法实现中需要用到多维数组的数据结构, 储存空间在数据维度增大时会有很大的需求。

实验采用二维数据集,数据集来源为 UCLA 的 SOCR Data,记录了 1993 年香港地区 两万五千名18岁青年的身高体重数据。下载数据集后先把数据单位换算成为厘米和千克, 再将数据取为整数以便统计,清洗后的数据集可见于文件 height vs weight.csv。数据集 的二维联合直方图以热力图形式展示。

```
import os # 文件操作
import time # 时间操作
import numpy as np # 数组操作
import seaborn as sns # 二维联合直方图(熱力图)绘制
from matplotlib import pyplot as plt # 绘图操作
        def __init__(self, dataName):
             self_datAmame | datamame):
self_datAmame = datamame.split('.')[0] # 图像的名称
self_dataPath = os.getcwd() + '\\' + datamame  # 數据集的文件路径
self_dataSet = None # 數据集的法人表籍
self_dataType = np_int # 數据集的數据关望,测试集的數据关型为整型
self_dimension = None # 数据集的雜度,即联合直方图的雜度
                 self.dataSet = np.loadtxt(fname-self.dataPath, dtype-self.dataType, delimiter=',') # 数据集的格式为csv, 分隔符为運号 self.dimension = self.dataSet.shape[1] # 数据形式为n维行向量。shape[1]表示列数,即数据的维度
         def evaluateHistogram(self): # 计算直方图数据并可视化直方图
                   # 計算並及 maxVal = np.amax(self.dataSet, axis=0) # 数据各維度的最大值 minVal = np.amax(self.dataSet, axis=0) # 数据各維度的最大值 binten = 3 * np.ones(shape=self.dimension, dtype=np.int) # 直方图坦距, 不能取过小数值, 否则会导致数据聚集在边界上 binNum = ((maxVal - minVal) / binten + 1).astype(np.int) # 直方图坦距, 不能取过小数值, 否则会导致数据聚集在边界上 binNum = np.zeros([var for var in binNum[::-1]], dtype=np.int) # 储存直方图数据的n维数组, 大小为各维度组数之积, 初始值为 for val in self.dataSet: # 適历发展集的行向量 binNum = 1 # 法传行通量专业等由的能量处据
                               ggram = Прт. cerovit.

al in self. dataset: # 進历要限集的订同量
binho = [] # 该行向量在直方限中的确坐标
for index in range(self. dimension): # 適所存向量的各维度分量。以索引形式访问
loc = int((val[index] - minval[index]) / binten[index]) # 计算变量的组号。区间左闭右开,组号从6到组载或1
loc = max(loc, 0) # 边界检查: 组号不能小于6
loc = min(loc, binthun[index] - 1) # 边界检查: 组号不能大于组载或1
binho, append(loc) # 储存分量的组号,注意组号的顺序
# 在行向量的位置记录数据
```

二维联合直方图如下图所示。程序的运行时间为 1.02 秒。



1.2 Implement computation of local histograms using efficient update method.

Python script code:

1_2_local_histogram.py

Input:

lena_std.tif

Output:

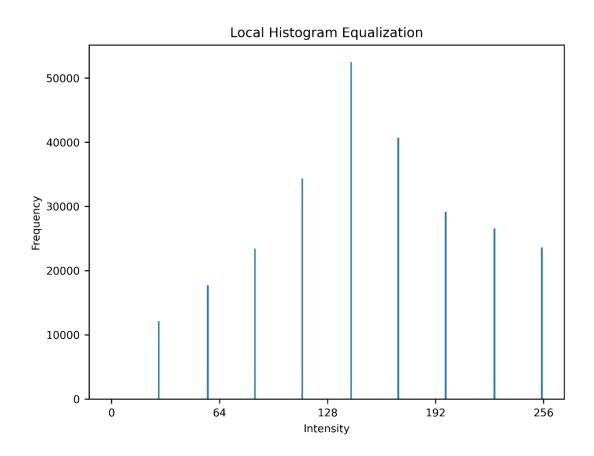
lena std histogram by local histogram algorithm.png

• Description:

Python 源代码中给出了使用高效局部更新方法计算图像直方图的实现算法。局部直方图算法的时间复杂度取决于图像的边长n和感兴趣区域的边长d,普通局部直方图方法需要在每一步遍历感兴趣区域中的所有值,复杂度为 $O(d^2)$,而高效局部更新方法只需要访问感兴趣区域中的其中两列,复杂度降为O(d),能有效提高更新效率。

高效局部更新算法主要用于局部直方图均衡,能有效降低图像整体对细节的影响, 使得图像能够进行自适应均衡化。

使用高效局部更新方法进行局部直方图均衡后的图像直方图如下图所示,原图像为lena std.tif。程序的运行时间为 48.87 秒。



2 Implement piecewise linear transformation function for image contrast stretching.

Python script code:

2_piecewise_linear_transformation.py

• Input:

lena_std.tif

Output:

lena_std_by_image_contrast_stretching.png

Description:

Python 源代码中给出了利用分段线性函数变换进行图像对比度拉伸的实现算法。分段线性函数变换的主要目的是使图像的暗端变得更暗而亮端变得更亮,由于线性函数能够任意定义,相比于其他的函数变换更为自由,故能通过调整线性函数的参数来局部优化图像的灰度。

实验采用三段式的分段线性函数进行处理,其中前后两段线段的斜率相等,两个拐点取在三等分的位置,具体参数在代码中给出。

使用分段线性函数变换进行图像对比度拉伸的图像前后效果如下图所示,其中左图为变换前,右图为变换后。程序的运行时间为 1.94 秒。



3.1 Implement histogram equalization algorithm.

Python script code:

3 1 global histogram equalization.py

• Input:

lena_std.tif

• Output:

lena_std_by_global_histogram_equalization.png
lena std histogram by GHE.png

Description:

Python 源代码中给出了图像全局直方图均衡的实现算法。通过计算图像灰度的累积分布函数,将灰度的分布转换成灰度累积分布的分布,而灰度累计分布的分布近似为均匀分布,可以通过全局均衡化提高图像整体的对比度。由于灰度值为整数,具体转换时需要对计算得到的累积分布值进行四舍五入。

在全局直方图均衡过程中,图像的每部分都会受到来自图像整体分布的影响。

Source Code:

```
import numpy as np # 數組操作
from matplotlib import pyplot as plt # 绘图操作
from PIL import Image # 图像操作
       def __init__(self, imageName):
    self.imageName = imageName.split('.')[0] # 图像的名称
    self.imageName = imageName.split('.')[0] # 图像的名称
    self.imageName # 图像的路径
    self.imageName # None # 图像的宽
    self.imageNidth = None # 图像的宽
    self.imageSplice = None # 至自拼接图像

      self.imageMatrix = None
      # 图像的灰度矩阵

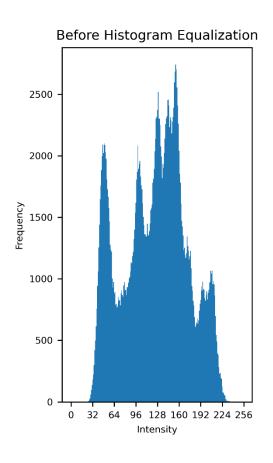
      self.globalMatrix = None
      # 全局直方图均衡的灰度矩阵

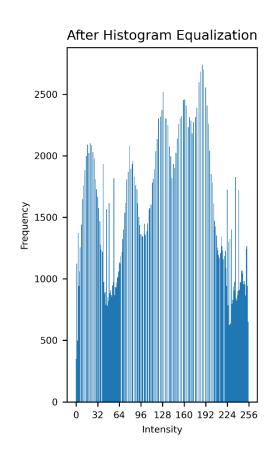
                     self.histogram = None # 均衡前的图像直方图
self.globalHistogram = None # 全局直方图均衡后的图像直方图
           def inputImage(self):
    image = Image.open(self.imagePath) # 輸入图像
    self.imageMatrix = np.array(image.convert('L')) # 将图像转换成灰度矩阵
    self.imageHeight, self.imageWidth = self.imageMatrix.shape # 图像大小
    self.histogram = np.histogram(self.imageWatrix.reshape(1, -1), range(257))[0] # 计算均衡前的图像直方图
          def histogramEqualization(self): # 计算全局直方图均离 self.globalMatrix = copy.deepcopy(self.imageMatrix) # 创建液度矩阵的深拷贝 mapping = {val: np.round(255 * np.sum(self.histogram[:val + 1]) / (self.imageHeight * self.imageWidth)) for val in range(256)} # 全局直方图均离 for h in range(self.imageHeight): # 通历离方向 for w in range(self.imageWidth): # 通历离方向 self.globalMatrix[h][w] # 與計算物產化后的檔案值 self.globalMatrix[h][w] # 與計算物產化后的檔案值 self.globalMatrix[squam(self.globalMatrix.reshape(1, -1), range(257))[0] # 计算均离后的图像直方图
         def outputHistogram(self): # 可视化直力图 plt.figure() # 创建空白图像 plt.subplot(1, 2, 1) # 五子图 plt.bar(range(256), self.histogram, width=1) # 根据均离前的直为图数据绘制直方图 plt.title('Before Histogram Equalization', fontsize=10) # 标注标题 plt.ylabel('Intensity', fontsize=2) # 标注本标签 plt.ylabel('Frequency', fontsize=7) # 标注本标签 plt.ylabel('Frequency', fontsize=7) # 设定水槽字号 plt.xticks(range(26, 257, 32), fontsize=7) # 设定水槽字号 plt.subplot(1, 2, 2) # 五子图 plt.bar(range(256), self-globalHistogram, width=1) # 根据均衡后的直方图数据绘制直方图 plt.title('After Histogram Equalization', fontsize=10) plt.xdabel('Intensity', fontsize=7) # 标注本标签 plt.ylabel('Intensity', fontsize=7) # 标注本标签 plt.xticks(range(26, 257, 32), fontsize=7) # 标注本标签 plt.xticks(range(26, 257, 32), fontsize=7) # 设定水槽字号 plt.xticks(ronsize=7) # 设定水槽字号 plt.subplots_adjust(wspace=0.5) # 课整两子图向的距离 plt.savefig(self.imageName + 'histogram by_GHE.png', dpi=480) # 导出两幅直方图为png文件,分辨率为480dpi # plt.show() # 发行程序时显示正力图
           def outputHistogram(self): # 可视化直方图 plt.figure() # 创建空白图像
          def outputImage(self):
    self.imageSplice = Image.new(mode='L', size=(self.imageWidth * 2, self.imageHeight)) # 生成空白拼接图像
    self.imageSplice.paste(im=Image.fromarray(self.imageWatrix)) # 生成天度化图像
    self.imageSplice.paste(im=Image.fromarray(self.globalMatrix), box=(self.imageWidth, 0)) # 生成直方图均衡图像
    self.imageSplice.save(self.imageWame + *_by_global_histogram_equalization.png') # 保存图像
    # self.imageSplice.show() # 运行程序时显示图像
          def histogramProcessing(self): # 计算直方图均衡的程序接口
self.inputImage() # 輸入图像
self.histogramEqualization() # 全局直方图均衡
self.outputHistogram() # 輸出直方图
self.outputImage() # 輸出图像
          imageName = 'lena_std.tif' # 图像的名称
imageData = ImageData(imageName=imageName) # 实例化Image对象
imageData.histogramProcessing() # 護用函數计算直方图均衡
```

• Result:

进行全局直方图均衡的图像及其直方图前后效果如下图所示,其中左图为均衡前, 右图为均衡后。程序的运行时间为 1.27 秒。







3.2 Implement local histogram equalization algorithm using efficient update method.

Python script code:

3 2 local histogram equalization.py

• Input:

```
lena_std.tif
square_noisy.tif
```

Output:

```
lena_std_by_local_histogram_equalization.png
lena_std_histogram_by_LHE.png
square_noisy_by_local_histogram_equalization.png
square_noisy_histogram_by_LHE.png
```

Description:

Python 源代码中给出了使用高效局部更新方法进行图像局部直方图均衡的实现算法。 局部直方图均衡能有效降低图像整体对细节的影响,使得图像能够进行自适应均衡化。

实验采用两幅图像进行局部直方图均衡处理测试,其中对课本中的例图进行了结果重现。不同于全局直方图均衡,局部直方图均衡对图像细节的调整更为有效。

```
f localHistogramEqualization(self): # 计算局部直方图均衡
self.localMatrix = copy.deepcopy(self.imageMatrix) # 创建效度矩阵的深拷贝
roiHeight, roiWidth = 3, 3 # 應兴趣区域和T的高和度,規定为3*3邻域、能忽略边界造成的影响
halfRoiHeight, halfRoiWidth = roiHeight // 2, roiWidth // 2 # 應兴趣区域和T的卡斯和中意
roiHistogram = np.histogram(self.imageMatrix(:roiHeight,:roiWidth), reshape(1, -1), range(257))[0] # 初始應兴趣区域和T的直方图
for hin range(self.imageWaifth - roiWidth) +  # 總历意方向, 代表意光整区域和T的立上角线坐标
for w in range(self.imageWaifth - roiWidth + 1): # 總历意方向, 代表意光整区域和T的主上角线坐标
currentPixel = (halfRoiHeight + h, halfRoiWidth + w) # 当前邻域中心健素位
currentValue = self.imageWaifth - roiWidth # 当前邻域中心健素位
self.localMatrix[currentPixel] = np.round(255 * np.sum(roiHistogram(:currentValue + 1]) / (roiHeight * roiWidth))
if w = self.imageWaifth - roiWidth: # 感兴趣区域和ToiMple的场面子来》,是后一次调用对处于健身状态,但不会对健素值装值
roiHistogram = np.histogram(self.imageWatrix[h + 1:h + roiHeight + 1, :roiWidth].reshape(1, -1), range(257))[0] # 计算次行首个感兴趣区域和ToiMple的方面
roiHistogram = np.histogram(self.imageWatrix[h + roiWinighth + ng)
roiHistogram = np.histogram(self.imageWatrix[h + ng)
roiHistogram = np.histogram(self.imageWatrix[h + ng)
roiHistogram = ng.histogram(self.imageWatrix[h + ng)
roiHistogram = ng.histogram = ng.histogram(self.imageWatr
                  else: # 麼呆壓逐級の1両石移动一个條業
roiHistogram -= np.histogram(self.imageMatrix[h:h + roiHeight, w], range(257))[0] # 删除最左边一行
roiHistogram += np.histogram(self.imageMatrix[h:h + roiHeight, w + roiWidth], range(257))[0] # 淺加最右边一行
self.localHistogram = np.histogram(self.localMatrix.reshape(1, -1), range(257))[0] # 计算局部均衡后的图像直方图
 def outputHistogram(self): # 可視化直方图 plt.figure(figsize=(15, 5)) # 创建空白图像 plt.subplot(1, 3, 1) # 左子图 plt.bar(range(256), self.histogram, width=1) # 根摺均實前的直方图数据绘制直方图 plt.title('Before HE', fontsize=10) # 标註标题 plt.ylabel('Intensity', fontsize=8) # 标註本标形象 plt.ylabel('Intensity', fontsize=8) # 标註本标序象 plt.ylabel('Frequency', fontsize=8) # 被注外标签 plt.xticks(range(6, 257, 64), fontsize=8) # 改定x结子号 plt.xticks(range(6, 257, 64), fontsize=8) # 改定x结子号 plt.xticks(fontsize=8) # 设定x结子号
                  plt.xticks(range(e, 22), 64), rontsize=8) # 皮皮科維于罗
plt.yticks(fontsize=8) # 皮皮科維于罗
plt.subplot(1, 3, 2) # 中子图
plt.bar(range(256), self-globalHistogram, width=1) # 根据全局均衡后的直方图数据绘制直方图
plt.title('After GHE', fontsize=10)
plt.xdabel('Intensity', fontsize=8) # 标注文轴标签
plt.ytlabel('Irequency', fontsize=8) # 标注文轴标签
plt.ytlabel('Frequency', fontsize=8) # 标注文轴标签
                 plt.ylabel('Frequency', fontsize-e8) # 按注/結标签
plt.xticks(range(0, 257, 64), fontsize-e8) # 设定x袖字号
plt.yticks(fontsize-8) # 设定y轴字号
plt.subplots_adjust(wspace-0.5) # 调整两子图闸的距离
plt.subplot(1, 3, 3) # 石子图
plt.subplot(1, 3, 3) # 石子图
plt.bar(range(256), self.localHistogram, width=1) # 根据全局均衡后的直方图数据绘制直方图
plt.title('After LHE', fontsize-e8)
plt.xtlabel('Intensity', fontsize-e8) # 标注x轴标签
plt.ylabel('Frequency', fontsize-e8) # 标注x抽样签
plt.xtlcks(range(0, 257, 64), fontsize-e8) # 设定x袖字号
plt.yticks(fontsize-e8) # 设定x袖字号
plt.subplots_adjust(wspace-e0.3) # 调整两子图闸的距离
                    plt.yticks(fontsize=8) # 设定y轴字号
plt.subplots_adjust(wspace=0.3) # 调整两子图同的距离
plt.savefig(self.imageName + 'histogram_by_LHE.png', dpi=480) # 导出两幅直方图为png文件, 分辨率为480dpi
# plt.show() # 运行程序对显示直方图
    def outputImage(self):
               f outputtmage(self):
self.imageSplice = Image.new(mode='L', size=(self.imageWidth * 3, self.imageHeight)) # 生成空白拼接图像
self.imageSplice.paste(im=Image.fromarray(self.imageWatrix)) # 生成灰度化图像
self.imageSplice.paste(im=Image.fromarray(self.globalMatrix), box=(self.imageWidth, 0)) # 生成全局直方图均衡图像
self.imageSplice.paste(im=Image.fromarray(self.localMatrix), box=(self.imageWidth * 2, 0)) # 生成局萨直方图均衡图像
self.imageSplice.save(self.imageWidth * 2, 0)) # 生成局萨直方图均衡图像
self.imageSplice.show() # 运行程序对显示图像
# self.imageSplice.show() # 运行程序对显示图像
def histogramProcessing(self): # 计算直方图均衡的程序接口
self.inputImage() # 输入图像
self.globalHistogramEqualization() # 全局直方图均衡
self.localHistogramEqualization() # 局部直方图均衡
self.outputHistogram() # 輸出直方图
self.outputImage() # 輸出直角
 start = time.time() # 程序开始运行时刻
  # 四條-- imageName2 = 'lena_std.tif' # 图像的名称
imageName2 = 'lena_std.tif' # 图像的名称
imageNata2 = ImageData(imageName=imageName2) # 实例化Image对象
imageData2.histogramProcessing() # 调用函数计算直方图均衡
```

使用高效局部更新方法进行局部直方图均衡的图像及其直方图前后效果如下图所示, 其中左图为均衡前,中图为全局均衡后,右图为局部均衡后。程序的运行时间为99.88秒。

