数字图像处理

大作业四

手势数字识别

学号　　2017011589

姓名　　吾尔开西

班级　　自 76

目录

一、总述

手势识别是在计算机视觉领域中受到广泛关注的问题，早在深度神经网络崛起前，研究者和程序员们就对该问题进行了一定的探索，但传统的算法比较复杂，而且随着针对目标的变化，算法也会随之改变。

我在本次大作业中用传统的方法实现了手势数字 0 到 5 的识别，不仅能应用于静态图片，而且能在视频流中实现实时的识别。

近年来，深度神经网络得到广泛应用，在计算机视觉领域更是如此，其特点是不需要人工设计算法进行特征提取，且对不同种类的手势有普适性。因此，我训练了一个卷积神经网络进行手势数字的识别。

本项目的传统算法基于 python 的 OpenCV，深度神经网络方法基于 python 的 TensorFlow。

二、静态图片手部提取

静态图片提取手部形状其实是一个图像分割问题，可以使用区域生长算法、分水岭算法等。然而，该问题的难度在于：没有进行区域生长的种子点信息，且图片的背景可能很复杂。因此，我们必须针对特定问题设计特殊的解决方案。

1、彩色空间转换

观察上面的图片，可以看出手部的颜色偏红，于是我们想到可以将图像从 rgb 空间转化到 YUV 空间来进行分割(Liu, You, Jain, & Wang, 2003)。

YUV 空间中，“Y”表示明亮度（Luminance 或 Luma），也就是灰度值；而“U”和“V” 表示的则是色度（Chrominance 或 Chroma），作用是描述影像色彩及饱和度。下面分别是图片的 YUV 分量。

可以看出在图片的 U 分量中，手部的色度比较突出，于是我们选择 U 分量来进行下一步的分割。

为了减小噪声干扰，我们在分割前对 U 分量进行了高斯模糊，卷积核大小为5 × 5

2、OTSU 阀值分割

由于图像没有人工标记的种子点，无法使用区域生长、分水岭算法等。因此我们决定使用简单的阀值分割，阀值的选择使用自适应的 OTSU 算法。

OTSU 算法认为使得目标和前景的方差和最小的的阀值为最佳阀值 (Otsu & cybernetics, 1979)，方差和为：

$$\sigma_w^2(t) = P_b(t)\sigma_b^2(t) + P_f(t)\sigma_f^2(t)$$

图像的总方差：

$$\begin{aligned}\sigma^2 &= P_b(t)\sigma_b^2(t) + P_f(t)\sigma_f^2(t) \\ &\quad + P_b(t)\left[\mu_b(t)-\mu\right]^2 + P_f(t)\left[\mu_f(t)-\mu\right]^2 \\ &= \sigma_w^2(t) + \sigma_0^2(t)\end{aligned}$$

可以看到，图像的总方差等于组内方差和加上组际方差和，且总方差不变。要使组内方差最小，就是要使组际方差最大，即找到使得下式最大的阀值，作为最优阀值。

$$\sigma_0^2 = P_b(t)\left[1-P_b(t)\right]\left[\mu_b(t)-\mu_f(t)\right]^2$$

使用 OTSU 算法得到的分割结果如下：

3、获取手部区域

在分割后，首先使用形态学开运算去除小的噪点，开运算算子为5 × 5大小的矩形，结果如下：

开运算后，除了手部区域还有一些干扰部分也被识别成前景，为了去除这些区域，我们找出其中的最大连通区（具体流程在下一部分详述），将其作为提取结果。

三、视频流手部提取

在视频流中进行手部提取时，第二部分的算法仍然是适用的，但是视频流手部提取要求实时计算，算法的计算速度至少要快于帧率。而且第二部分的算法对每一帧单独进行计算，没有考虑视频中帧与帧之间的相关性，这显然是不合理的，因此，我们对视频流使用特别的算法进行手部提取。

1、混合高斯模型

视频流中的手部提取仍然是一个图像分割问题，我们需要决定每一个像素点属于前景还是背景。我们用概率模型对每个像素点建模，设像素点 x 为前景的概率为$p(x)$,高斯混合模型基于这样的假设(Zivkovic & Van Der Heijden, 2006):

$$p(x) = \sum_{k=1}^{M} \pi_k N(x|u_k, \sigma_k)$$

即整幅图像可由 M 个高斯分布相加拟合，π_k为每个高斯分布的权重，满足$\sum \pi_k = 1$，u_k, σ_k为高斯分布的均值和方差。要计算$p(x)$，我们需要确定合适的 M，u_k, σ_k, π_k。

为了确定模型的参数，我们可以使用最大似然估计，似然函数如下:

$$\ln L(u,\sigma,\pi) = \sum_{n=1}^{N} \ln \sum_{k=1}^{M} \pi_k N(y_n|u_k,\sigma_k)$$

为了使似然函数最大，我们可以求导得到参数的极值点。但由于似然函数的对数符号中有求和号，求极值并不方便。因此我们转而使用 EM 算法。

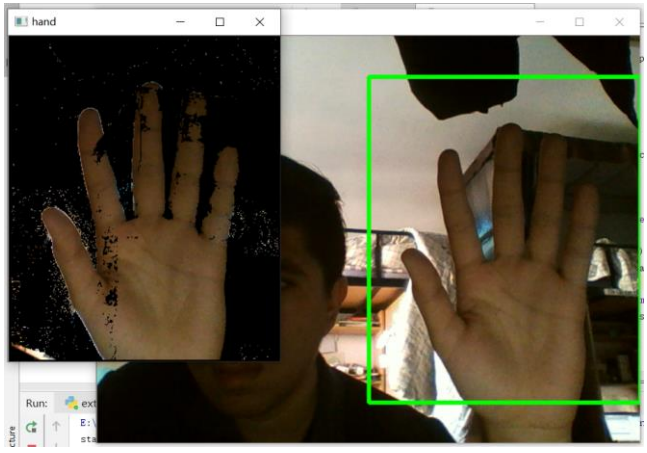
EM 算法适用于不知道采样数据来源于哪一类的情况，首先给一组起始参数(u^0,σ^0,π^0)，每一步对 Q 函数进行优化：

$$u^{i+1},\sigma^{i+1},\pi^{i+1} = \arg\max Q(u,\sigma,\pi,u^i,\sigma^i,\pi^i)$$

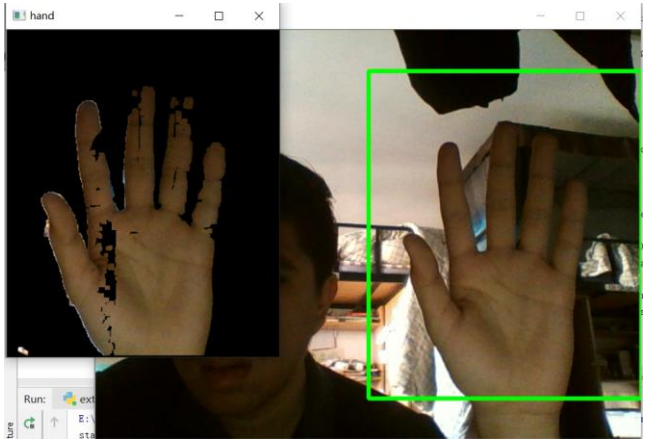
不断迭代，最终得到最优的参数。实际运算时，样本来自于最近多个帧的数据，利用到了视频流帧与帧相关联的特性。这样就可以得到一个分割结果了。

2、形态学运算

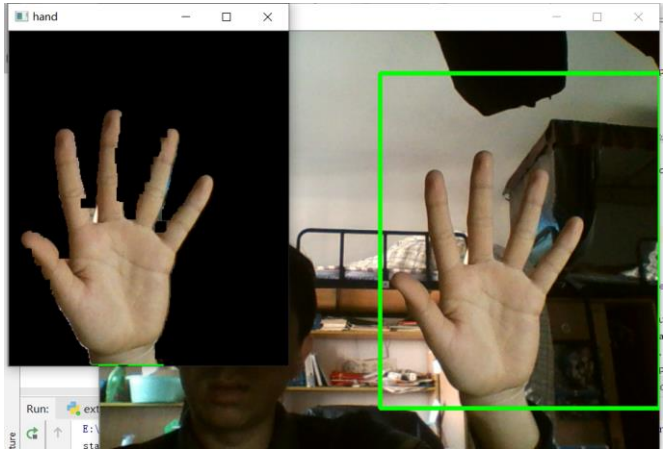
用 GMM 算法得到的手部 mask 比较粗糙，如下图



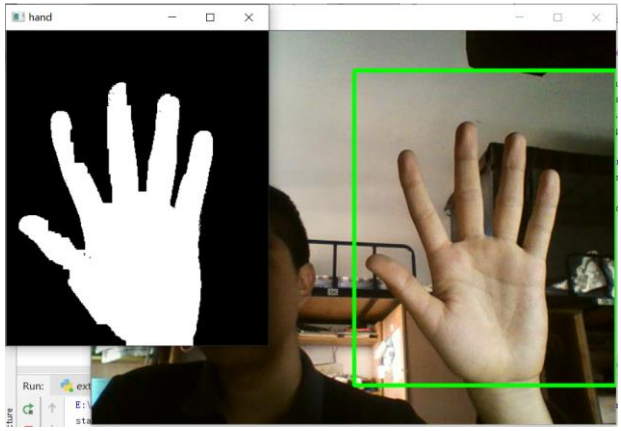
可以看出手部之外的一些噪点也被识别为前景，为了去除这些噪点，我们使用开运算：



开运算后，手部内有一些小洞，可以使用闭运算填补:

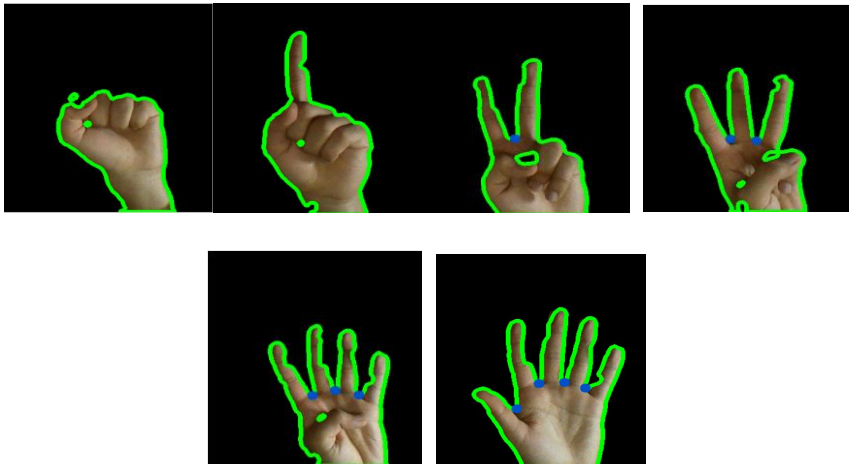


经过两步运算后，手部区域已经比较完整了，噪声点也比较少，于是我们对灰度图进行二值化，得到结果:



四、手势识别

在提取出手部形状的二值图后，接下来就是对手势的识别。我们需要识别的手势包括如下 6 种。



1、利用最小凸集算法的识别

从上面的图片中我们可以看出，数字手势的一个明显特征是两指间指缝的数量：5 的手势有 4 个指缝，4 的手势有 3 个指缝，以此类推。

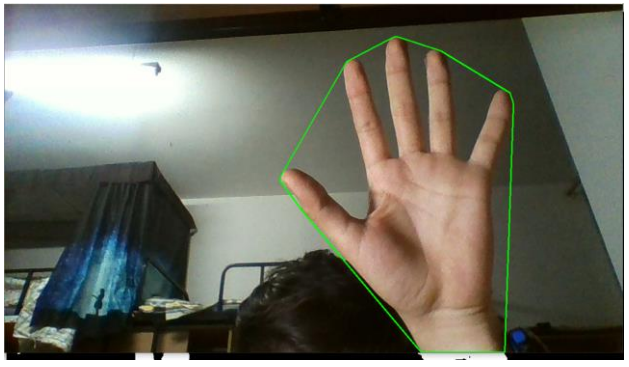
0 和 1 的手势都没有指缝，这两者的区分我们会用另外的算法。首先讨论如何识别出指缝的数量。

上面的图片中，手的外部有一圈绿色的线，这是手部形状的外轮廓。求外轮廓可以使用形态学的膨胀运算，使用八领域矩形算子进行膨胀，之后减去原图，就可以得到轮廓。

可到轮廓后可计算轮廓内的面积，找出最大面积的轮廓。这样可以去除一些干扰背景的影响。

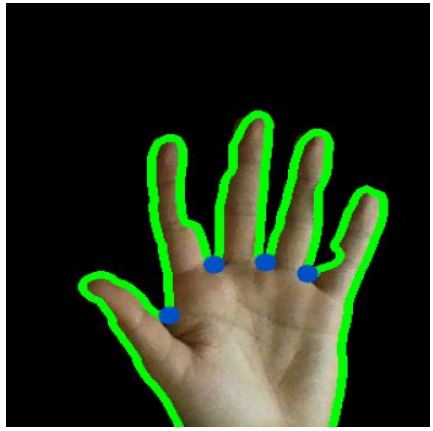
要识别指缝的数量，我们要用到最小凸集算法，给出最小闭包的定义：在一个向量空间 V 中，对于给定集合 X，所有包含 X 的凸集的交集 S 被称为 X 的最小凸集(Eddy, 1977)

计算最小凸集的算法比较简单，按扫描方向碰到的第一个顶点，例如最左、最下之点，为起点 A。沿逆时针方向找到边界上的第二点 B。看看是否所有顶点都在 AB 的一侧？如果不是，取下一点 C 为顶点，再看是否多边形的所有顶点都在 AC 的一侧。如果是则保留 AC 并将分析移至 C 点；继续同样的分析，直到又回到起点。由这些保留点构成了最小凸集(Eddy, 1977)。



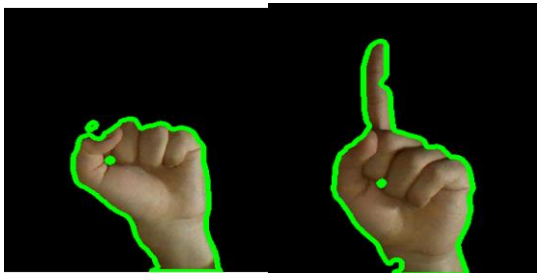
如图，是对手的轮廓求取最小凸集的结果。

得到最小凸集后，求凸缺陷，即原图形不在最小凸集上的最大角。指缝也在这些角中，其特点是夹角较小，于是我们挑出其中夹角小于某个阀值的缺陷，其个数加一就是伸出的手指个数。例如下图中检测到 4 个缺陷角（蓝色点为角顶点），检测结果为伸出 5 个手指。



2、利用区域紧性特征区分零一

使用 1 中的方法可以区分手势 2，3，4，5.然而手势 0 和手势 1 的缺陷角个数都为 0，要如何区分二者呢？



我们注意到手势 0 与手势 1 相比更接近圆，于是想到用区域紧性（Compactness）来区分，紧性的定义如下

$$C = \frac{P^2}{A}$$

其中，P 代表周长，A 代表面积。C 越小，表示图形越接近圆。于是我们用紧性的一个阀值来区分手势 0 和 1.

五、基于深度神经网络的手势识别

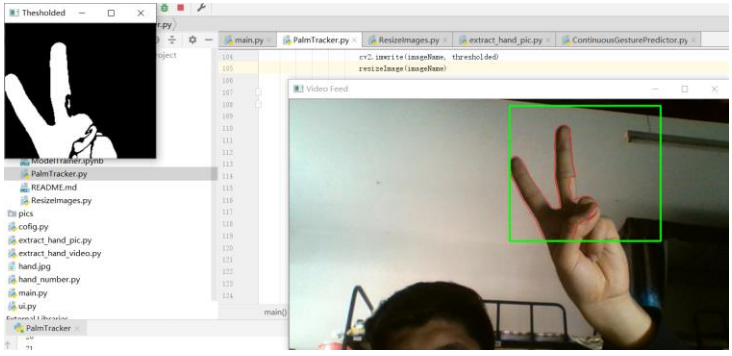
近年来，深度神经网络得到广泛应用，在计算机视觉领域更是如此，其特点是不需要人工设计算法进行特征提取，且对不同种类的手势有普适性。因此，我训练了一个卷积神经网络进行手势数字的识别。

1、网络结构

使用 9 层的卷积神经网络，卷积核大小为 2，激活层为 relu，每一个卷积层后面接一层 maxpooling，最后设置两层全连接层。

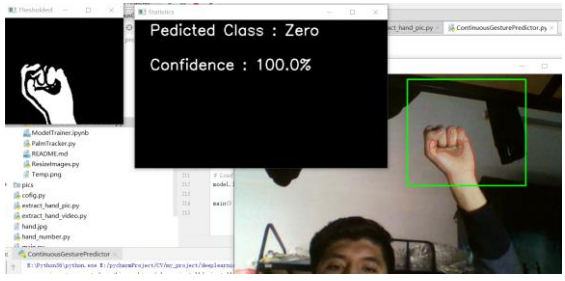
2、收集训练数据

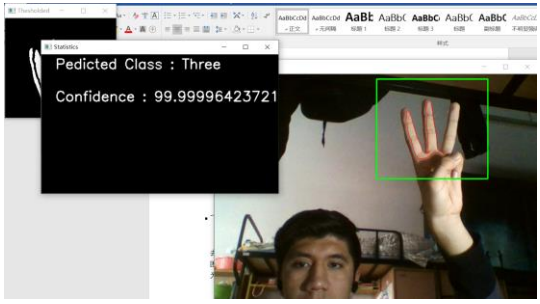
使用摄像头收集训练数据，用第三部分的方法分离前景与背景，提取手部形状，保存为图片。每个手势收集 1000 张训练数据，100 张测试数据。



3、训练与预测

使用收集到的数据进行训练，在预测时分离前背景的方法与第三部分相同。普遍来说预测结果比较可靠

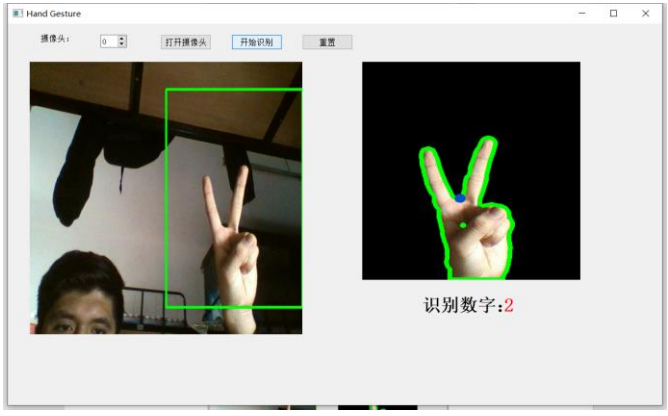




六、结果

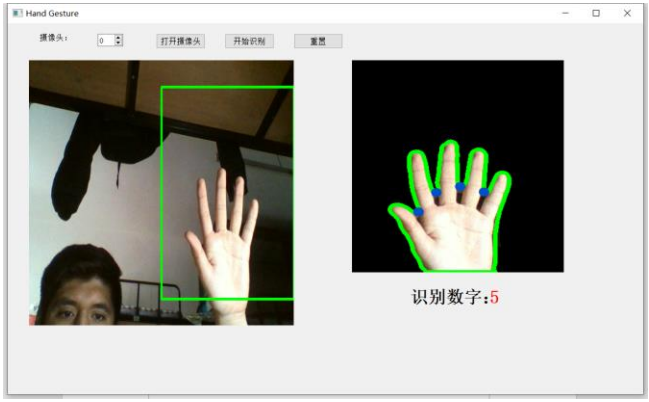
传统方法和基于神经网络的方法都能得到基本准确的识别结果，然而基于神经网络的方法鲁棒性较强。传统方法的结果如下，两种方法我都拍摄了演示视频。


Hand Gesture
摄像头：
0
打开摄像头
开始识别
重置
识别数字:1


Hand Gesture
摄像头：
0
打开摄像头
开始识别
重置
识别数字:2


Hand Gesture
摄像头：
0
打开摄像头
开始识别
重置
识别数字:3


Hand Gesture
摄像头：
0
打开摄像头
开始识别
重置
识别数字:4



七、总结

本次大作业完成了一个比较完整的项目，解决问题时充分利用到了问题的特性，并结合了课堂上很多学过的知识，比如 OTUS 阀值分割、图像形态学、区域紧性、最小凸包等等，体会到了知识的力量，提高了动手能力和灵活运用知识的能力，也算是对这门课的一个交代。

卷积神经网络的实现让我感受到深度神经网络的神奇之处，无需人工编写复杂的算法，也不用人工提取特征，神经网络就能以不错的性能完成任务。

总之，经过本次大作业的推进，我对手势识别领域有了一定的了解，完成大作业的过程中也提高了我发现问题、解决问题的能力，受益匪浅。

八、参考资料

Eddy, W. F. J. A. T. M. S. (1977). A new convex hull algorithm for planar sets. *3*(4), 398-403.

Liu, Z.-f., You, Z.-s., Jain, A. K., & Wang, Y.-q. (2003, 27-30 Sept. 2003). *Face detection and facial feature extraction in color image.* Paper presented at the Proceedings Fifth International Conference on Computational Intelligence and Multimedia Applications. ICCIMA 2003.

Otsu, N. J. I. t. o. s., man,, & cybernetics. (1979). A threshold selection method from gray-level histograms. *9*(1), 62-66.

Zivkovic, Z., & Van Der Heijden, F. J. P. r. l. (2006). Efficient adaptive density estimation per image pixel for the task of background subtraction. *27*(7), 773-780.