第 57 卷 第 3 期 厦门大学学报(自然科学版) Vol. 57 No. 3
2018 年 5 月 Journal of Xiamen University (Natural Science) May 2018
doi:10.6043/j.issn.0438-0479.201712010

·研究简报·

基于视觉跟踪的实时视频人脸识别

任梓涵,杨双远*
(厦门大学软件学院,福建 厦门 361005)

摘要:目前基于深度学习的人脸识别方法准确率高,但是模型复杂,识别速度慢.为了实现监控视频中人脸的实时识别,提出了一种基于视觉跟踪的实时视频人脸识别(RFRV-VT)方法.首先将监控视频的帧序列分组,每一组中分为人脸识别帧和人脸跟踪帧;然后在人脸识别帧中使用基于深度学习的人脸检测和人脸特征提取方法,在人脸跟踪帧中使用基于核相关滤波(KCF)的视觉跟踪方法以加快识别速度.将该方法应用于数据集 YouTube Faces(YTF)上进行测试,实验结果显示该算法在监控视频中具有实时性和较高的识别准确性(99.60%).

关键词:视觉跟踪;人脸识别;监控视频
中图分类号:TP 391 **文献标志码**:A **文章编号**:0438-0479(2018)03-0438-07

随着摄像机在监控和移动设备中的广泛使用,大量的视频不断被捕获,与静态图像相比,视频往往包含更多信息,例如时间和不同视角信息等.公共场所的监控视频为社会的安全和执法提供了有力的保障,因此,视频监控系统的智能化是计算机视觉领域具有重要意义的研究方向之一,尤其是人脸识别技术在监控视频中的使用.

人脸识别的步骤主要包含人脸检测、人脸分割、人脸特征提取和人脸匹配 4 个部分[1].人脸识别算法发展到现在,目前主要分为以下几种:基于特征的方法、基于模版匹配的方法、基于外观的方法和基于深度学习的方法[2].基于特征的方法主要是使用手工提取特征来进行人脸识别,例如方向梯度直方图(HOG)、局部二值模式(LBP)、尺度不变特征变换(SIFT)、Gabor 等方法[3-4].基于模版匹配的方法是建立一个标准的面部模型,将人脸定义为函数[2-5],输入一张图像,分别计算出人脸轮廓、眼睛、鼻子、嘴巴标准模式的相关值,根据这些相关值来判定人脸是否存在.基于外观的方法主要是用整幅图像的信息来识别人脸,较为典型的是主成分分析(PCA)算法[6]和线性鉴别分析(LDA)算法[7],它们使用降维统计的方法保留图像的关键信息,避免了整幅图像的大量计算[3].以上方法在姿态转动、光照变化、遮挡等复杂场景下识别具有很大的局限性.

近年来,基于深度学习的人脸识别方法变得流行.深度学习中,卷积神经网络(CNN)[8]的非线性特征学习能力特别强,所以在人脸识别等计算机视觉任务中具有很高的准确率.户外标记人脸数据集(LFW)[9]是著名的静态人脸识别数据集,许多基于深度学习的人脸识别方法在 LFW 上的准确率一次次突破极限,最高已达到 99.63%,但是在视频人脸数据集 YouTube faces(YTF)[10]上的准确率却不高,只有 92.80%.与静止拍摄的人脸图像相比,视频中的人脸成像更加复杂,而且容易受到设备的影响,识别起来比较困难.2014 年,Hu 等[11]提出了一种新的判别式深度度量学习(DDML)方法,该方法训练了一个深度神经网络,将人脸映射到一个特征空间,并使用马氏距离度量来使类间差异最大化和类内差异最小化,该方法在 YTF 上取得 82.3%的最好准确率.Tran 等[12]提出了基于 CNN 的三维可变人脸模型(CNN-3DMM)方法,该方法将人脸静态图像输入到训练好的一个 CNN 模型,得到三维的可变人脸模型,在 YTF 数据集上的准确率为 88.8%.Taigman 等[13]提出了深度人脸识别(DeepFace)方法,是一种将人脸图像对齐到一般的

收稿日期:2017-12-07 录用日期:2018-03-22
基金项目:福建省自然科学基金(2015J01288)
*通信作者:yangshuangyuan@xmu.edu.cn
引文格式:任梓涵,杨双远.基于视觉跟踪的实时视频人脸识别[J].厦门大学学报(自然科学版),2018,57(3):438-444.
Citation:REN Z H,YANG S Y.Real-time face recognition in videos based on visual tracking[J].J Xiamen Univ Nat Sci,2018,57(3):438-444.(in Chinese)

3D 形状模型的级联方法,在 YTF 上的准确率达到了 91.4%.2015 年,Wu 等[14]提出了一个轻量级的 CNN (a lightened CNN)方法,该方法用 4 个卷积层构成一个浅 CNN 模型,并提出一个新的最大特征映射(MFM)激活函数来获取紧凑的人脸特征信息,在 YTF 数据集上达到 91.6%的准确率.

近年来,监控视频中的目标跟踪是一个重要研究方向.在 2010 年之前,目标跟踪领域中最常用的方法是经典跟踪方法,如均值漂移(Meanshift)、粒子滤波、卡尔曼滤波和基于特征的光流算法等[15],由于这些方法不能够适应和处理视频中复杂的运动变化,在基于相关滤波和深度学习的跟踪方法出现后,经典跟踪方法已经较少使用.2012 年,Henriques 等[16]提出了循环结构的核跟踪(CSK)方法,这是一个基于循环矩阵和核函数的跟踪方法,使用循环矩阵完成密集采样,并利用傅里叶变换实现快速检测,该方法在处理速度上能够达到每秒 320 帧,为相关滤波系列方法在实时性应用中打下了基石.Henriques 等[17]又后续提出 CSK 的改进算法核相关滤波/判别相关滤波(KCF/DCF),在保证高速处理的同时提高了跟踪的准确性.2013 年 Wang 等[18]提出"深度学习跟踪"方法,这是第一个将深度网络运用于单目标跟踪的跟踪算法.Hyeonseob 等[19]的多域网络(MDNet)方法,利用多类跟踪序列预训练网络,并在在线跟踪时微调模型.基于深度学习的跟踪算法不断发展,目前性能上还是没有办法和相关滤波方法相比,但是其端到端输出的特点使其具有光明的前景.

综上所述,虽然基于 CNN 的人脸识别算法在图像人脸识别任务中取得很好的结果,但是在基于视频的人脸识别任务中的准确性不高.另外,由于这些算法的网络层次深,模型结构复杂,时间复杂度和空间复杂度都比较高,无法满足视频中实时识别的要求.因此本研究提出了一种基于视觉跟踪的实时视频人脸识别(RFRV-VT)方法,为了简化模型,采用视频分组的方法,对视频序列以组为单位进行识别;采用一种新的特征融合方法提高人脸识别准确率;并将视觉跟踪方法引入到人脸识别框架中来提高识别速度;最后在 YTF 数据集上对 RFRV-VT 进行验证.

1　RFRV-VT 方法介绍

本研究提出的 RFRV-VT 方法总体框架图如图 1 所示,首先提出将视频图像序列进行分组识别,每一组中包括人脸识别帧(图 1 中用 R 表示)和人脸跟踪帧(图 1 中用 T 表示).在人脸识别帧中实现人脸检测、人脸特征提取和人脸匹配过程,在人脸跟踪帧中对之前检测到的人脸进行跟踪.并将视频图像帧序列以 N 帧分为一组,每个组中第一帧为人脸识别帧,第 $2\sim N$ 帧为人脸跟踪帧,N 的设置需进行最优选择.相邻两个组之间,本研究提出一种双重匹配方法实现连接.

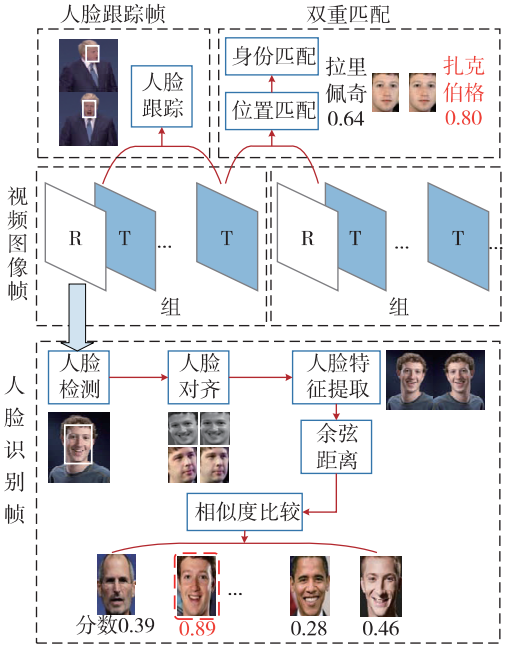

图 1　RFRV-VT 方法框架图

Fig. 1　Framework of RFRV-VT

1.1　人脸识别

在人脸识别帧中,首先使用多任务级联 CNN (MTCNN)[20]来实现视频图像帧中的人脸检测;其次,对 MTCNN 检测到的人脸按照人脸框分割,并使用 lightened CNN 提取分割出来的人脸图像的特征;最后,对提取出的特征采用余弦距离进行度量,从而实现人脸匹配.

1.1.1　人脸检测

人脸检测在图像中检测到人脸并返回人脸框坐标的过程,往往容易受到图像质量、光照、人脸转动的等因素的影响.深度学习中,CNN 可以在人脸检测中获取更高层次的语义信息,与传统的基于手工特征的检测方法相比,基于 CNN 的算法更具有鲁棒性.人脸图像经过 CNN 中的卷积操作,可以产生大量的人脸候选窗口,将人脸候选框输入到 softmax 分类器,能够得到其是否为人脸的分类结果,从而实现人脸检测.

本文中使用的 MTCNN 人脸检测方法由如下 4 步构成.首先,对视频帧中的图像进行不同尺度采样,

作为卷积神经网络结构的输入;其次,使用包含 3 个卷积层的全卷积网络粗略获取一部分人脸窗口候选集,并使用非最大抑制方法来提高检测结果的准确性,如图 2(a)所示;然后将其送入一个 4 层的卷积神经网络,该网络在第一个网络的基础上增加了一个全连接层,用来去掉更多非人脸的区域,如图 2(b)所示;最后将结果输入到一个 5 层的卷积神经网络做精细的处理,并输出人脸检测框坐标和人脸 5 个关键点的位置,如图 2(c)所示.MTCNN 使用 3 个 CNN 级联的方式,实现了由粗到细的算法结构,每个 CNN 网络是一个分类器,能够得到最有可能的人脸区域.该方法通过减少滤波器数量、设置小卷积核和增加网络结构的深度,实现了使用较少的运行时间获得更好的性能,在人脸转动、光照变化和部分遮挡等情况下能够得到很好的人脸检测结果.

1.1.2 人脸特征提取及其改进

人脸特征提取目的是为了提取出人脸的深层抽象特征,这个抽象特征能够具有区分两个不同人脸的特性.本研究基于 lightened CNN[14] 来提取人脸的深层特征.如图 3 所示,模型中包含 4 个卷积层、4 个最大采样层以及 2 个全连接层,全连接层输出 256 维的特征向量.模型中使用 MFM 激活函数,它在输入的卷积层中选择两层,取相同位置的最大值作为输出.假设有输入卷积层 $\boldsymbol{C}\in\mathbf{R}^{h\times w\times 2n}$,MFM 激活函数的数学表达式为

$$l_{ij}^{k}=\max_{1\leqslant k\leqslant n}(C_{ij}^{k},C_{ij}^{k+n}),\tag{1}$$

其中,输入卷积层的通道数为 $2n$,$1\leqslant i\leqslant h$,$1\leqslant j\leqslant w$,$\boldsymbol{l}\in\mathbf{R}^{h\times w\times n}$.

根据式(1),激活函数的梯度可以表示为

$$\frac{\partial l_{ij}^{k}}{\partial C_{ij}^{k'}}=\begin{cases}1, C_{ij}^{k}\geqslant C_{ij}^{k+n},\\ 0, \text{其他},\end{cases}\tag{2}$$

其中,k'为常数,$1\leqslant k'\leqslant 2n$,并且有

$$k=\begin{cases}k', 1\leqslant k'\leqslant n,\\ k'-n, n+1\leqslant k'\leqslant 2n.\end{cases}\tag{3}$$

由式(2)可以看出,激活层有 50% 的梯度为 0,MFM 激活函数能够得到稀疏的梯度.MFM 选择 2 个卷积特征图候选节点之间的最大值,采用聚合统计的方法得到最紧凑的特征表示,而且实现了变量选择,比常用 Relu 激活函数的高维稀疏梯度更具优点.Lightened CNN 采用轻量的结构,在取得比较好的人

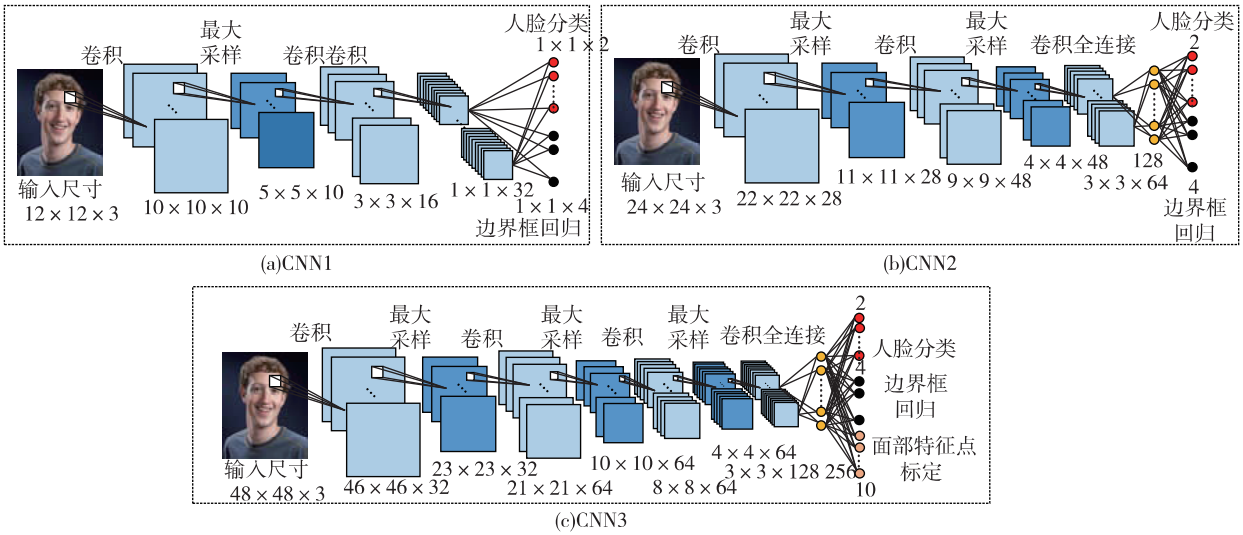

图 2 MTCNN 结构

Fig. 2 Structure of MTCNN

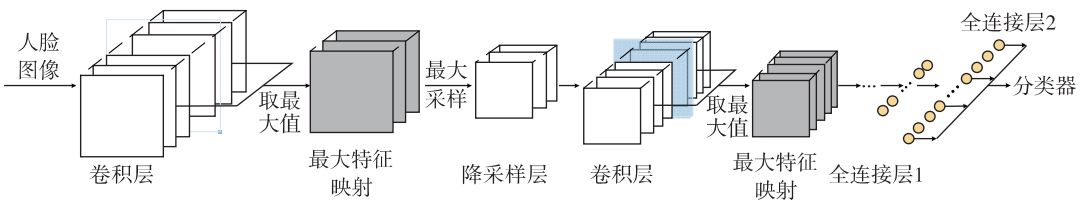

图 3 lightened CNN 结构

Fig. 3 Structure of lightened CNN

脸识别效果的同时,加快了识别速度,减小了存储空间的占用,对于监控视频的实时人脸识别具有良好的效果.

为了进一步加快人脸识别速度和提高人脸识别准确度,本研究提出了一种特征融合的新方法.

首先,使用人脸识别图像中人脸区域图像和其水平旋转 180°的人脸镜像图像,分别输入到 lightened CNN 中,得到两个 256 维的向量.在一般图像任务的神经网络模型训练过程中,将训练数据集中的图像进行镜像、旋转等,能够产生更多的有效数据,提高模型的推广能力.受此启发,在本研究中,使用人脸原图像和镜像图像提取特征,以期得到更丰富的人脸信息,有提高识别准确率.

接着,使用特征融合函数将两个特征向量每一维的最大值融合形成一个新的特征向量,具体的特征融合函数为

$$h_x=\max(a_x,b_x),x=1,2,\cdots,m, \tag{4}$$

其中,x 表示第 x 维,m 是特征向量的维数.

最后,在人脸特征提取中,提取的特征维数太多会导致特征匹配时过于复杂,消耗系统资源,为了降低运算复杂度,本研究使用 PCA 算法将融合得到的 256 维特征向量进行压缩降维,最后得到 128 维的特征向量,在后续特征匹配中能够大大加快计算速度,对本研究中要求的算法实时性具有重要意义.

1.1.3　人脸匹配

本研究使用余弦距离来度量人脸比对结果.两个人脸分别提取出 128 维特征向量,计算两个特征向量的余弦距离,代表两个人脸的相似程度,如果余弦距离超过某个阈值则认为是同一个人的人脸.当一个人脸与多个人脸进行比对,如果超过阈值的有多张人脸,则取相似度最高的为最终结果.

1.2　人脸跟踪

人脸跟踪过程是已知在某一帧中检测到人脸,并标记脸部区域,然后在后续视频帧中继续跟踪标记的脸部位置.在人脸跟踪帧中,本研究使用基于核相关滤波的高速跟踪方法(KCF)[15],将跟踪问题当作一个二分类问题,从而找到目标与背景的边界.

如图 4 所示,KCF 实现目标跟踪的步骤为:假设在视频图像序列 i 帧中有目标对象,其位置坐标为 $L(i)$.首先,在 $L(i)$ 附近采集负样本,$L(i)$ 作为正样本,训练一个目标检测器,将图像样本输入检测器,能够得到该样本的一个检测响应值;接着,在视频图像序列 $i+1$ 帧中,在 $L(i)$ 附近进行采样,并将样本输入目标检测模型中,得到每个样本的检测响应值,即其是人脸区域的概率值;最后,取响应值最大的样本位置为 $i+1$ 帧的目标位置 $L(i+1)$.

在采集样本的步骤中,KCF 利用循环矩阵的性质,将图像候选区域像素矩阵乘以一个循环矩阵,用来表示候选框上下左右移动后新样本的窗口,这样可以快速制造大量新的样本,更多的样本数量能够训练更好的检测器.训练目标检测器时,KCF 使用脊回归算法进行训练,计算相邻两个视频帧之间的相关性,算法过程中利用循环矩阵在傅里叶域可对角化的特点,将时域上的卷积运算转换为频域上向量的点乘,这样能够大大减少计算量,如果使用方向梯度直方图(HOG)特征来跟踪,KCF 能够达到 172 帧/s 的跟踪速度,并且有很高的准确率.

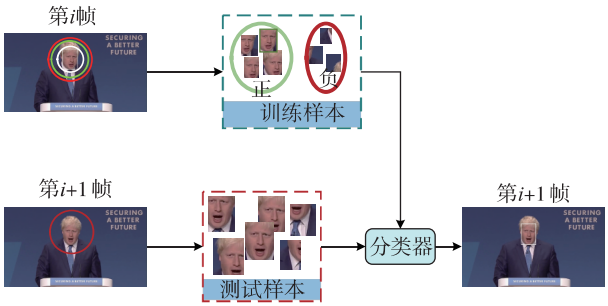

图 4　跟踪算法步骤

Fig. 4　Tracking algorithm steps

1.3　视频组间匹配

本研究中人脸识别以视频组为单位独立进行,为了实现相邻组间的信息连接,本研究提出一种双重匹配方法.

相邻的两个视频帧序列组 g_i 和 g_{i+1} 之间,标记 g_i 中的最后一帧为 f_n^g,g_{i+1} 组中的第一帧为 f_1^{g+1},本研究使用双重匹配方法实现相邻组的连接.

1) 位置匹配.算法过程中保存 f_n^g 中所有人脸的框的位置,当在 f_1^{g+1} 中完成人脸检测算法时,计算 f_1^{g+1} 中每个人脸和 f_n^g 中每个人脸的欧氏距离,如果距离小于某个阈值,则认为是同一个人的人脸.若 f_1^{g+1} 中有某一人脸与 f_n^g 中所有人脸的欧氏距离都大于设定的阈值,则判定为新出现的人脸.

2) 身份匹配.f_1^{g+1} 帧完成人脸识别过程后,比较 f_1^{g+1} 和 f_n^g 相对应的人脸身份信息.如果不相同,则需要比较两个身份的相似度大小,取相似度比较大的结果为最终识别结果.如果相同,则不需要更新.

通过上述的双重判别方法,两个视频帧序列组 g_i 和 g_{i+1} 之间实现了连接,同时,身份匹配能够对前一

次的人脸识别结果进行矫正,提高人脸识别的准确率.

2 实验和结果分析

2.1 实验设置及数据

本研究实验平台为 6 核,Intel(R) Core(TM) i7-5820K @3.30 GHz,内存 64 GB 的 CPU 和 GeForce GTX TITAN X,显存 12 GB 的 GPU.

首先在公开视频 YTF[10] 上进行测试,然后再使用监控摄像头在真实场景下进行测试.YTF 数据集共包含从 YouTube 网站上收集的3 425个视频,其中有 1 595个不同的人.每个视频的姿态、光照和表情有很大的变化,每个视频剪辑的平均长度为 181.3 帧.对于监控摄像头真实场景下的测试,本实验使用海康威视 IP 摄像头在实验室进行监控视频采集,数据包含 20 个不同人的 50 个视频,每个人都有不同角度、表情,每个视频剪辑的平均长度为 203 帧.

本实验参数设置如表 1 所示.人脸特征提取的结果为 128 维的特征向量.视频组间位置匹配时使用欧氏距离度量,如果相邻两帧两个人脸的欧式距离小于 20,则认为是同一个人.人脸识别帧中,如果人脸比对的余弦距离大于 0.75,则认为是同一个人.

表 1 实验参数

Tab.1 Experimental parameters

实验名称	参数	参数值
人脸特征提取	维度	128
位置匹配	欧氏距离阈值	20
人脸相似性	余弦距离阈值	0.75

2.2 实验数据处理

YTF 数据集:根据 YTF 数据集标准的评估协议,实验中使用5 000个视频对,分为 10 个组,每个组包含 250 个正样本对和 250 个负样本对.从每个视频中随机选取 100 个样本并计算其平均相似度和处理速度.

摄像头数据集:对 50 个视频中的 22 个人采集不同角度人脸的照片并保存在数据库中,实验中对监控摄像头拍摄的视频样本与数据库中人脸进行比对,每个视频取 100 个样本并计算其平均相似度和处理速度.

2.3 实验结果

实验 1:在本研究中,监控视频图像帧序列中每组设为 N 帧,对于 N 的设置,则要通过实验来选择.如表 2 所示,使用摄像头数据来进行测试,实验测试了当 N 分别取 3~8 时,算法的准确性和处理速度.当 N 取 3 的时候,识别准确率最高,为 99.73%,但是其处理速度为 19 帧/s,达不到实时处理的效果(通过测试,处理速度达到 28 帧/s 以上,视频才能够流畅显示).由于 N 的取值越大,越容易产生人脸漏检的情况,所以人脸识别的准确率会随 N 的增大下降.但 N 的增加使人脸跟踪帧的帧数增加,进而使算法的处理速度大大提高.由表 1 可知,当 N 取 5,其处理速度为 38 帧/s,满足视频实时性的要求;此时人脸识别准确率为 99.60%,有较高的准确率.综合考虑,本研究中图像帧序列以每组 5 帧处理,后续实验在这个最优值下进行.

表 2 监控视频不同分组方式下的算法准确率和处理速度

Tab.2 The accuracy and processing speed in different ways of grouping videos

N	准确率/%	处理速度/(帧·s^{-1})
3	99.73	19
4	99.65	24
5	99.60	38
6	98.40	46
7	97.90	55
8	97.90	67

实验 2:为了验证本研究提出的特征融合方法的有效性,比较了基于 CNN 的 DeepFace、视觉几何组(VGG)、lightened CNN、lightened CNN+特征融合方法以在两个数据集上的识别准确率.结果如表 3 所示,可以看出,本研究提出的特征融合方法能够提高人脸识别的准确率,在 YTF 数据集上可以达到 92.50%,有良好的识别效果.同时,在实际监控视频中可以达到 99.6%的识别准确率.

表 3 人脸识别准确率对比

Tab.3 Comparison of face recognition accuracy

方法	准确率/%	
	YTF	摄像头数据
DeepFace	91.40	95.97
WebFace	88.00	96.80
VGG	92.80	99.30
Lightened CNN	91.60	98.50
Lightened CNN+特征融合方法	92.50	99.60

实验 3:为了验证本研究中视觉跟踪算法对提高人脸识别速度的有效性,在使用特征融合方法的基础上,比较没有视觉跟踪算法和加入视觉跟踪算法之后的平均处理速度,并与 YTF 数据集上准确率较高的 VGG 方法相比.结果如表 4 所示,可以看出加入视觉跟踪技术能够极大地加快人脸识别处理速度,在实际的监控摄像头中可以达到平均 38 帧/s 的速度,完全满足监控摄像头中实时处理要求.

表 4　视觉跟踪方法处理速度对比

Tab. 4　The processing speed comparison of the visual tracking method

方法	处理速度/(帧·s^{-1})	
	YTF	摄像头数据
VGG	1.7	2
特征融合方法	15	16
特征融合+视觉跟踪	29	38

3　结　论

本研究提出了一个基于视觉跟踪的视频人脸识别方法,创新之处有:1) 提出了视频帧分组的方法,以组为单位进行识别.2) 将视觉跟踪加入到人脸识别方法中,以提高识别速度.3) 提出了一种特征融合方法,以提高人脸识别的准确度.使用该方法在 YTF 数据集和实际监控视频数据集上进行了实验,结果显示本研究提出的人脸识别框架具有较好的识别准确率,并且极大地提高了识别速度,满足了监控视频中实时处理的要求.

本研究提出的算法框架是组件式框架,可以随着行业的研究进行算法替换,实验中对于视频分组采取的是平均分组,后续考虑采取其他分组方式,用剪枝的方法,能够进一步提高识别速度,也将考虑结合数据库中的大规模人脸检索方法,在实际场景中达到更快的处理速度.

参考文献:

[1] 李武军,王崇骏,张炜,等.人脸识别研究综述[J].模式识别与人工智能,2006,19(1):58-66.

[2] JAFRI R,ARABNIA H R.A survey of face recognition techniques[J].Journal of Information Processing Systems,2009,5(2):41-68.

[3] ZHAO W,CHELLAPPA R,PHILLIPS P J.Face recognition:a literature survey[J].ACM Computing Surveys,2003,35(4):399-458.

[4] 严严,章毓晋.基于视频的人脸识别研究进展[J].计算机学报,2009,32(5):878-886.

[5] SOLANKI K,PITTALIA P.Review of face recognition techniques[J].International Journal of Computer Applications,2016,133(12):20-24.

[6] DASHORE G,RAJ V C.An efficient method for face recognition using principal component analysis(PCA)[J].International Journal of Advanced Technology and Engineering Research,2012,19(2):23-29.

[7] CHAN L H,SALLEH S H,TING C M.Face biometrics based on principal component analysis and linear discriminant analysis[J].Journal of Computer Science,2010,6(7):693-699.

[8] LÉCUN Y,BOTTOU L,BENGIO Y,et al.Gradient-based learning applied to document recognition[J].Proceedings of the IEEE,1998,86(11):2278-2324.

[9] HUANG G B,MATTAR M,BERG T,et al.Labeled faces in the wild:a database for studying face recognition in unconstrained environments[R].Amherst:University of Massachusetts,2007.

[10] WOLF L,HASSNER T,MAOZ I.Face recognition in unconstrained videos with matched background similarity[C]// IEEE Conference on Computer Vision and Pattern Recognition.Colorado Springs:IEEE,2011:529-534.

[11] HU J L,LU J W,TAN Y P.Discriminative deep metric learning for face verification in the wild[C]// IEEE Conference on Computer Vision and Pattern Recognition.Columbus:IEEE,2014:1875-1882.

[12] TRAN A T,HASSNER T,MASI I,et al.Regressing robust and discriminative 3D morphable models with a very deep neural network[C]// IEEE Conference on Computer Vision and Pattern Recognition.Honolulu:IEEE,2017:1493-1502.

[13] TAIGMAN Y,YANG M,RANZATO M,et al.DeepFace:closing the gap to human-level performance in face verification[C]// IEEE Conference on Computer Vision and Pattern Recognition.Columbus:IEEE,2014:1701-1708.

[14] WU X,HE R,SUN Z N.A lightened CNN for deep face representation[EB/OL].(2015-11-09)[2017-12-01].https://arxiv.org/abs/1511.02683v1.

[15] WU Y,LIM J,YANG M H.Online object tracking:a benchmark[C]// IEEE Conference on Computer Vision and Pattern Recognition.Portland:IEEE,2013:2411-2418.

[16] HENRIQUES J F, RUI C, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels[J]. European Conference on Computer Vision, 2012, 7575(1): 702-715.

[17] HENRIQUES J F, RUI C, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(3): 583.

[18] WANG N, YEUNG D Y. Learning a deep compact image representation for visual tracking[C]// International Conference on Neural Information Processing Systems. Lake Tahoe: NIPS, 2013: 809-817.

[19] NAM H, HAN B. Learning multi-domain convolutional neural networks for visual tracking[C]// IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 4293-4302.

[20] ZHANG K, ZHANG Z, LI Z, et al. Joint face detection and alignment using multi-task cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.

Real-time Face Recognition in Videos Based on Visual Tracking

REN Zihan, YANG Shuangyuan*

(Software School, Xiamen University, Xiamen 361005, China)

Abstract: At present, face recognition methods based on deep learning yield high accuracies, but their complex models recognize face slowly. To achieve the real-time face recognition in surveillance videos, we propose a real-time face recognition method in videos based on visual tracking (RFRV-VT). Firstly, this algorithm divides the surveillance video frame sequence into several groups, and each group contains face recognition frames and face tracking frames. Then, face detection method and face feature extraction method based on deep learning are used in the face recognition frame, and visual tracking method based on kernelized correlation filters (KCF) is used to speed up the recognition in the face tracking frame. This method is applied to the YouTube Faces (YTF) dataset for testing. Experimental results show that the proposed algorithm exhibits real-time performances and high recognition accuracies in videos (99.60%).

Key words: visual tracking; face recognition; surveillance video