基于视频的人脸识别研究进展

严 严[1),2)] 章毓晋[1),2)]
[1)] (清华大学信息科学与技术国家实验室 北京 100084)
[2)] (清华大学电子工程系 北京 100084)

摘 要 近年来基于视频的人脸识别已成为人脸识别领域最为活跃的研究方向之一.如何充分利用视频中人脸的时间和空间信息克服视频中人脸分辨率低,尺度变化范围大,光照、姿态变化比较剧烈以及时常发生遮挡等困难是研究的重点.文中对近期(主要近5年)基于视频的人脸识别研究进行了详细的介绍和讨论,在对相关方法分类的基础上,分析了各类方法中典型技术的优缺点,并概况介绍了常用的视频人脸数据库和实验结果,最后展望了基于视频人脸识别未来的发展方向和趋势.

关键词 模式识别;人脸识别;基于视频的人脸识别;进展
中图法分类号 TP391 **DOI号**: 10.3724/SP.J.1016.2009.00878

State-of-the-Art on Video-Based Face Recognition

YAN Yan[1),2)] ZHANG Yu-Jin[1),2)]
[1)] (*National Laboratory for Information Science and Technology, Tsinghua University, Beijing* 100084)
[2)] (*Department of Electronic Engineering, Tsinghua University, Beijing* 100084)

Abstract Recently, video-based face recognition has become one of the hottest topics in the domain of face recognition. How to fully utilize both spatial and temporal information in video to overcome the difficulties existing in the video-based face recognition, such as low resolution of face images in video, large variations of face scale, radical changes of illumination and pose as well as occasionally occlusion of different parts of faces, is the focus. The paper reviews most existing typical methods for video-based face recognition (especially for the last 5 years) and analyses their respective pros and cons. Two commonly used video face databases and some experimental results are given. The prospects for future development and suggestions for further research works are put forward in the end.

Keywords pattern recognition; face recognition; video-based face recognition; progress

1 引 言

人脸识别具有非常重大的理论意义和应用价值.人脸识别的研究对于图像处理、模式识别、计算机视觉、计算机图形学等领域的发展具有重大的推动作用,同时在生物特征认证、视频监控、安全等各个领域也有着广泛的应用.

经过多年研究,人脸识别技术已取得了长足的进步和发展.随着视频监控、信息安全、访问控制等应用领域的发展需求,基于视频的人脸识别已成为人脸识别领域最为活跃的研究方向之一[1-4].如何充分利用视频中人脸的时间和空间信息克服视频中人脸分辨率低,尺度变化范围大,光照、姿态变化剧烈以及时常

收稿日期:2006-09-03;最终修改稿收到日期:2009-04-06.本课题得到国家自然科学基金(60872084)和教育部高等学校博士学科点专项科研基金(SRFDP-20060003102)资助.严 严,男,1984年生,博士,主要研究领域为人脸识别等.章毓晋(通信作者),男,1954年生,博士,教授,博士生导师,主要研究领域为图像工程(图像处理、图像分析、图像理解及其技术应用).E-mail: zhangyj@ee.tsinghua.edu.cn.

发生遮挡等困难是研究的重点.国内外众多的大学和研究机构,如美国的 MIT[5]、CMU[6-7]、UIUC[8-9]、Maryland 大学[10-12]、英国的剑桥大学[13-15]、日本的 Toshiba 公司[16-18]和国内的中国科学院自动化所[19-21]都对基于视频的人脸识别进行了广泛而深入的研究.鉴于目前现有的人脸识别国内外综述文献主要针对基于静止图像的人脸识别研究[1-3],因此有必要对现阶段基于视频的人脸识别研究情况进行分析和总结,期望能够更好地指导未来的研究工作.

2　人脸识别概述

一个自动的基于视频的人脸识别系统包括了人脸检测模块、人脸跟踪模块、人脸特征提取模块和人脸识别模块[22].关于人脸检测、人脸跟踪和人脸特征提取的研究进展可以参考综述文献[1-3].本文重点介绍基于视频的人脸识别研究进展.

人脸识别问题可以定义成:输入(查询)场景中的静止图像或者视频,使用人脸数据库识别或验证场景中的一个人或者多个人[1-2].基于静止图像的人脸识别通常是指输入(查询)一幅静止的图像,使用人脸数据库进行识别或验证图像中的人脸.而基于视频的人脸识别是指输入(查询)一段视频,使用人脸数据库进行识别或验证视频中的人脸.如不考虑视频的时间连续信息,问题也可以变成采用多幅图像(时间上不一定连续)作为输入(查询)进行识别或验证.因此按照上面的分析,根据输入(查询)和人脸数据库的不同,人脸识别可以分成如表 1 所示的 4 种情况.

表 1　输入(查询)和数据库不同情况下的人脸识别

	识别方式	
	数据库中图像(多幅图像)	数据库中视频
输入(查询)图像	图像-图像(多幅图像)	图像-视频
输入(查询)视频	视频-图像(多幅图像)	视频-视频

表中"图像-图像(多幅图像)"人脸识别就是传统的基于静止图像的人脸识别[1-3].而"图像-视频"人脸识别是指利用人脸图像作为输入采用视频人脸数据库进行识别或验证.通常的应用领域是基于人脸的视频信息检索.本文重点介绍的基于视频的人脸识别主要是指后面两种情况,即"视频-图像(多幅图像)"人脸识别和"视频-视频"人脸识别."视频-图像(多幅图像)"人脸识别是指输入(查询)一段人脸视频,利用静止图像人脸数据库进行识别或验证."视频-视频"人脸识别是指输入和数据库都利用视频进行人脸识别或验证.相对于前面 3 种情况,"视频-视频"人脸识别可以利用的信息最多.视频中可以利用的信息包括[4]:多幅同一个人的人脸图像,视频中人脸在时间和空间上的连续性,利用视频生成的三维(3D)人脸模型等.需要强调的是本文这样分类的目的是为了能够对整个人脸识别领域的研究现状有一个宏观上的认识,并区分不同情况下的人脸识别.事实上不同情况下人脸识别采用的技术可以是相同的,例如对所有人脸视频序列的处理是按照某种规则(如大小、姿态、清晰度等)提取一张人脸图像,则上面的情况都可以采用基于静止图像的人脸识别技术.

本文首先对现阶段基于视频的人脸识别研究现状进行了详细的分析和讨论,接着介绍了常用的视频人脸数据库和实验结果,最后展望了未来的发展方向.本文假设已经得到图像或者视频中需识别人脸的位置.对静止图像中人脸的定位可参见文献[23-24],对视频中人脸的定位和分割可参见文献[25].

3　基于视频的人脸识别

根据上一节的讨论,下面把基于视频的人脸识别分成"视频-图像(多幅图像)"人脸识别和"视频-视频"人脸识别两种情况分别给予综述.

3.1　"视频-图像(多幅图像)"人脸识别

"视频-图像(多幅图像)"人脸识别是指采用人脸视频作为输入(查询)利用静止图像人脸数据库进行识别或验证.由于现有的大部分人脸数据库都是静止图像人脸数据库,如何充分利用视频中的人脸信息更好地进行人脸识别是现阶段迫切需要解决的问题.

解决这类问题的传统做法[26-28]可以分成两大类:一类方法对输入视频中的人脸进行跟踪,寻找满足一定规则(如大小、姿态、清晰度等)的人脸图像,然后利用基于静止图像的人脸识别方法.这类方法的缺点是规则很难定义,并且没有最大限度地利用人脸视频中的时间和空间连续信息.另一类方法利用视频中的空间信息进行人脸识别.通过对输入视频中每一幅人脸或者若干幅人脸采用基于静止图像的人脸识别方法[1-3],利用各种联合规则[29](如多数投票或者概率/距离累加等方法)再进行最终的识别.这类方法的缺点是联合规则常有相当的随机性[4].

近年来,一些研究者开始利用视频中人脸的时间和空间连续信息进行识别.文献[10]讨论了在贝叶斯理论的框架下统一解决人脸识别和跟踪问题,采用时间序列模型刻画人脸的动态变化,把身份变

量和运动矢量作为状态变量从而引进时间和空间的信息;利用序贯重要度采样(Sequential Importance Sampling,SIS)的方法有效估计出身份变量和运动矢量的联合后验概率分布,通过边缘化提取出身份变量的概率分布.实验结果表明了该算法的有效性.不过当姿态变化时识别率只有 57 %.之所以出现姿态变化时识别率低的原因是对时间连续性的利用体现在人脸外观一致上,而随着光照或姿态的变化会导致外观的明显不同.因此文献[11]进一步提出了自适应外观变化模型并且采用自适应运动模型更准确地处理姿态的变化,对运动模型中噪声的方差和采样算法中的粒子数根据计算得到外观模型的误差进行更新,采用鲁棒统计学(robust statistics)处理脸部遮挡问题.利用基于贝叶斯人脸识别[30]方法的似然函数进行权重更新使得整个算法更加有效.

文献[31-32]通过对输入视频中人脸的脸部特征或外观的跟踪进行人脸验证.基本思想是,如果是正确的输入(对应数据库中要验证的人脸),则跟踪的轨迹基本一致;而如果是不正确的输入,则跟踪轨迹没有规律性.相应的数学模型就是考虑所得到的运动矢量分布,如果呈现尖峰(一致的运动参数)则认为是正确的人脸.如果没有呈现尖峰,而是类似均匀分布,则认为是错误的人脸.

上述方法中都采用贝叶斯理论引进了时间信息,极大地提高了识别率.并且采用序贯重要度采样克服非高斯分布和非线性系统带来的难以估计概率密度的问题.但是估计概率密度需要大量的粒子,导致其计算量比较大.

3.2 “视频-视频”人脸识别

“视频-视频”人脸识别是指输入和数据库中的人脸均是以视频的形式存在.大量的文献对如何同时利用输入和数据库中的人脸视频进行了深入的研究.现有文献中对视频中人脸信息的描述方式总结起来有下面几种:

(1) 利用一幅代表性的图像得到的特征(矢量表示),如主成分分析(PCA)降维后的矢量等;

(2) 利用所有图像得到的特征(矩阵表示),如特征空间、示例(exemplar)等;

(3) 利用概率密度函数刻画视频中的人脸分布,如高斯模型等;

(4) 利用动态模型刻画视频中人脸随时间的动态变化,如隐马尔可夫模型等;

(5) 利用流形(manifold)刻画视频中的人脸分布,如分段线性 PCA 子空间等.

上述各种描述方式(矢量、矩阵、概率密度、动态模型、流形)之间可能的度量如表 2 所示.

表 2 描述方式之间的度量

输入	度量函数				
	矢量(x)(数据库)	矩阵(X)(数据库)	概率密度(f)(数据库)	动态模型(D)(数据库)	流形(M)(数据库)
矢量(x)	$d(x,x)$	$d(x,X)$	$f(x)$	$D(x)$	$M(x)$
矩阵(X)	$d(X,x)$	$d(X,X)$	$f(X)$	$D(X)$	$M(X)$
概率密度(f)	$f(x)$	$f(X)$	$d(f,f)$	\	\
动态模型(D)	$D(x)$	$D(X)$	\	$d(D,D)$	$d(D,M)$
流形(M)	$M(x)$	$M(X)$	\	$d(M,D)$	$d(M,M)$

表 2 中 d 代表两个模型之间的距离或相似度,$f(X)$,$M(X)$ 代表概率/距离累加或多数投票,$D(X)$ 代表各帧后验概率.

下面按照对输入描述方式的不同,分成矢量、矩阵、概率、动态模型、流形 5 个小节分别给予介绍.

3.2.1 矢 量

利用矢量作为输入描述方式的基本思想是利用视频得到一个反映输入人脸视频特性(如均值人脸图像、最好的正面图像等)的特征(矢量表示),和数据库中的人脸视频描述方式进行匹配.数据库中人脸视频的描述方式可以是矢量、矩阵、概率、动态模型、流形等.

文献[33]对数据库中的每类人脸建立一个 PCA 子空间,利用与各个人脸子空间的距离对输入视频中的所有人脸进行标注.文献[34]介绍了一种基于视频的人脸验证方法,采用形状和归一化纹理的联合外观模型(Active Appearance Model,AAM)来表示人脸,通过加入类别信息的改进线性鉴别分析(Linear Discriminant Analysis,LDA)分离出身份变量(identity)和其他变化因素(姿态、光照和表情).采用卡尔曼滤波器(Kalman filter)对身份变量进行跟踪得到的稳定值就是身份稳定估计量.人脸验证就可以通过对输入人脸视频进行跟踪得到的估计量和数据库得到的身份估计量进行比较看是否大于某个阈值来实现.与基于静止图像的人脸验证方法相比,采用基于视频的人脸验证能利用更多的信息,效果更好.算法采用 ASM(Active Shape Model)进行人脸定位可以避免误配准带来的影响.但是一旦定位不准,则对后继的参数跟踪会产生很大的误差,导致识别率下降.并且对于达到稳定估计值需要

的视频长度无法计算和估计.另一个借助视觉约束的人脸跟踪和识别方法可见文献[35].

3.2.2　矩　阵

采用矩阵作为输入描述方式进行人脸识别的算法可以分成两类,一类是利用得到的特征(矩阵表示)逐个与数据库中的人脸描述方式进行比较(相当于每次取出矩阵的一行或者一列),然后利用多数投票或者概率(距离)累加最大的方法进行识别.另一类是把得到的特征(矩阵表示)看成一个整体和数据库中的人脸描述方式进行比较.相对于前面一种方法,后者采用矩阵作为整体更能利用视频的空间连续信息.数据库中人脸视频的描述方式可以是矢量、矩阵、概率、动态模型、流形等.

文献[36]采用总体 PCA 方法进行降维,在低维空间中采用混合高斯模型(Gaussian Mixture Model, GMM)来表示数据库中每个人脸.通过计算输入视频中每一帧人脸的后验概率,采用多数投票和概率累加最大的方法得到最终结果.文献[37]对数据库的每类人脸建立多个匹配模板,并根据视频中的动态的信息(如人脸姿态、运动模糊等)对多个模板进行自适应的融合.文献[16]对输入的人脸序列和数据库中的人脸序列分别建立一个 PCA 特征子空间,两个特征子空间之间的距离由它们之间的夹角确定.为了进一步去除光照、姿态、表情等的影响,把子空间重新投影到限制子空间(constraint subspace)中,限制子空间只包含对识别有用的成分(身份)[17].为了解决限制子空间中需要大量样本的问题,进一步利用整体学习(ensemble learning)的方法训练出 M 个限制子空间,通过投影到这 M 个限制子空间的距离加权和作为人脸之间距离的度量[18].该类算法的主要缺点在于没有考虑每一类人脸的整体概率分布,没有利用每一类的均值和特征值,在投影到限制子空间时可能会产生一定的问题,并且参数的设定和空间维数都需要通过经验给出.

由于人脸在姿态、光照、表情变化时呈现非线性分布,文献[38]在线性空间中通过核的方法映射到高维的非线性空间(核 Hilbert 空间),在高维空间中的夹角(核主成分夹角)作为矩阵的相似性度量,并且利用正定的核函数就可以和 SVM(Support Vector Machine)结合起来提高分类的性能.文献[19-20]首先通过 LDA 进行线性降维,然后对每个人的人脸视频通过矢量量化技术或者 K 均值聚类形成 K 个类别,每个类别用聚类中心和聚类的权重来表示.最后采用 EMD(Earth Mover's Distance)距离作为相似性度量进行人脸识别.

文献[39]利用聚类的方法建立局部参数模型,对数据库中的每个人脸建立多个局部流形.首先对数据库中的每段人脸视频经过 LDA 进行线性降维,通过采取 ISOMAP(Isometric feature Mapping)[40]提取各点的测地距离(geodesic distance)作为人脸之间的距离,从而可以更准确地刻画各点在流形空间中的位置关系,然后采用 HAC(Hierarchical Agglomerative Clustering)聚类方法得到 K 个示例,对每一示例采用类似文献[41]的方法对每个局部模型建立双子空间(dual subspace)概率模型,使用概率测度作为相似性度量,采用多数投票进行识别.文献[6]对每段人脸视频建立一个特征空间并把视频中人脸的变化看成一个非平稳的随机过程(AR 模型),采用逐步更新特征空间的方法并且引进了权重的概念,对新的样本权重大,对以前的样本权重小.该文中针对每个人脸建立两个特征空间,包括训练集中的特征空间和识别后不断更新建立的新的特征空间来解决过慢学习的问题.文献[42-43]利用数据库中的人脸视频得到三维模型生成查询人脸视频条件下的光照和姿态变化,然后逐一进行比对,采用距离累计最大的方法得到识别结果.

3.2.3　概　率

采用概率作为输入描述形式的基本思想是把视频中人脸的动态变化看成是满足一定的概率分布的高维随机变量.一般对数据库中视频的描述方式也是概率方式,通过比较概率密度函数的相似性来度量人脸之间相似性.

文献[13]采用 GMM 模型学习不同姿态和光照条件下的人脸分布,对输入人脸视频和数据库中的人脸视频都利用 GMM 模型进行建模,采用 K-L 散度(Kullback-Leibler divergence)作为人脸之间相似性度量.文献[5]把人脸识别问题看成是一个假设检验问题,证明了如果人脸视频中每一帧之间是相互独立的,则得到的最优准则是 K-L 散度.假设每个人脸服从高斯分布,采用 K-L 散度作为相似性度量.但是由于假设是单高斯分布,因此无法刻画由于光照或者是姿态变化导致人脸呈现流形的情况,并且 K-L 散度本身是一种非对称的度量方式.文献[14]采用基于核函数方法把低维空间映射到高维空间,这样就可以在高维空间中利用低维空间中的线性方法(如 PCA)来解决一般的复杂的非线性问题,采用 RAD(Resistor-Average Distance)作为人脸相似性度量.为了解决配准误差所带来的识别率下降的问题,利用了多幅图像和 RANSAC(Random Sample Consensus)算法来解决.另外文献[44]利用了核的

方法,把原来的矢量空间映射到高维非线性空间 RKHS(Reproducing Kernel Hilbert Space)中计算概率分布之间的距离.

3.2.4 动态模型

无论是矢量、矩阵和概率都没有利用时间连续的信息,所以可以自然地推广到多幅人脸图像(时间上不必连续)作为输入时的人脸识别问题.而动态模型则利用了人脸的时间和空间连续变化的信息,能够更好地刻画人脸的动态变化特性.数据库通常的描述方式可以是矩阵、动态模型、流形.

文献[10]中采用 3.1 节中介绍的概率模型,通过自动选择人脸视频中的示例(在线 K 均值聚类),把人脸示例的索引也作为状态变量,采用 SIS 的方法估计出联合概率密度分布,最后通过边缘化求出身份变量的分布进行人脸识别.文献[7]中对数据库中的每段人脸视频采用 PCA 变换建立了特征子空间,在特征子空间中建立一个自适应隐马尔可夫模型(Hidden Markov Model,HMM),识别阶段就可以计算每个识别序列的后验概率作为相似性度量,并且当满足一定条件时对 HMM 模型进行更新.文献[12]把运动人脸建模成一个 ARMA(Auto-Regressive and Moving Average)模型(用姿态作为状态量,采用外观作为观测量),采用 ARMA 子空间之间的夹角作为相似性度量.

文献[8-9]和文献[10]的想法类似,认为应该把跟踪和识别结合起来,减少跟踪的误配准对识别的影响.对不同姿态下的人脸构造一个低维分段线性流形.为了引进时间信息,采用贝叶斯推理的方法,建立了不同姿态之间的转移矩阵,该文的算法能够很好地处理人脸的大规模旋转时的识别和跟踪问题.文献[21]首先对所有的人脸利用 LLE(Locally Linear Embedding)降维后建立整体分段线性模型,根据到各个分段子流形的距离采用贝叶斯推理的方法计算最大后验概率.在文献[45]中作者通过实验结果指出,利用时空结构的 HMM[7] 大于一定长度时要优于基于静止图像的多数投票方法,但是当视频的长度过短时则不一定.这说明时间长短对动态模型的识别率会有一定影响.

3.2.5 流 形

人脸在不同的光照、姿态变化下会构成一个的低维空间的流形[39,46].所以利用流形作为输入描述可以更好地描述人脸的分布.一般对数据库中的人脸采用同样的描述方法.比较输入和数据库中流形的相似性作为度量.

文献[47-49]使用流形来解决基于视频的人脸识别问题,首先建立了一个多视角动态人脸模型,包含了一个 3D 模型,一个和形状姿态无关的纹理模型,一个仿射变化模型.其基本思想是基于分析的合成,通过最小化损失函数,求解出模型的参数.在视频序列中该问题可以进一步简化,利用 Kalman 滤波求解出形状和纹理.人脸纹理通过 KDA(Kernel Discriminant Analysis)降维后对单个人脸序列建立一个分段的线性流形(特征矢量随着姿态的变化).接着就可以通过比较轨迹的匹配程度进行人脸识别.但是要进行 3D 模型的估计需要大量的多视角图像,计算复杂度较大.

现有文献中的典型算法总结如表 3 所示.

表 3 典型的"视频-视频"人脸识别的方法

输入描述	数据库中人脸的描述	度量方法	典型文献
矢量	PCA 特征子空间	重构误差 $d(x,X)$	文献[33]
矢量	LDA 降维后跟踪得到身份稳定估计量	欧式距离 $d(x,x)$	文献[34]
矩阵	混合高斯模型	多数投票/概率累加 $f(X)$	文献[35]
矩阵	PCA 特征子空间	子空间夹角 $d(X,X)$	文献[16-18]
矩阵	核 Hilbert 空间	核主成分角 $d(X,X)$	文献[37]
矩阵	矢量量化/ K-均值聚类得到示例	EMD $d(X,X)$	文献[19-20]
矩阵	每个人脸由多个局部模型组成 每个局部模型建立双子空间概率模型	多数投票 $f(X)$	文献[38]
矩阵	两个 PCA 特征子空间	多数投票 $d(X,X)$	文献[6]
矩阵	3D 模型得到的合成人脸图像	距离累加 $d(X,X)$	文献[42]
概率	混合高斯模型	K-L 测度 $d(f,f)$	文献[13]
概率	单高斯模型	K-L 测度 $d(f,f)$	文献[5]
概率	核 PCA 建立的单高斯模型	RAD 测度 $d(f,f)$	文献[14]
动态模型	在线 K 均值聚类得到示例	最大后验概率 $D(X)$	文献[10]
动态模型	隐马尔可夫(HMM)模型	最大后验概率 $d(D,D)$	文献[7]
动态模型	自回归滑动平均(ARMA)模型	ARMA 子空间夹角 $d(D,D)$	文献[12]
动态模型	PCA 子空间内分段线性流形	最大后验概率 $d(D,M)$	文献[8-9]
动态模型	LLE 降维后分段线性流形	最大后验概率 $d(D,M)$	文献[21]
流形	KDA 降维后分段线性流形	轨迹匹配 $d(M,M)$	文献[47-49]

3.3 小　结

综上可以看出“视频-图像(多幅图像)”人脸识别和“视频-视频”人脸识别研究的主要问题包括:

(1) 如何对高维的人脸图像降维;

(2) 如何对降维后的人脸序列进行描述;

(3) 如何刻画描述方式之间的度量;

人脸数据降维的目的是得到表达性特征(如主成分分析等)或鉴别性特征(如线性鉴别分析等)以降低高维人脸数据的计算复杂度和减弱噪声、表情、光照等因素的影响[50]. 对各种常见线性和非线性的降维方法研究的介绍可参考文献[51].

现阶段对降维后的人脸序列描述方式包括矢量、矩阵、概率、动态模型、流形等. 其中采用概率和流形的方法需要大量反映人脸分布的样本才能更准确地刻画人脸的分布,达到较好的性能. 利用动态模型能够很好地利用时间和空间的信息,但是方法相对比较复杂,计算量一般都比较大. 而利用矢量作为输入描述方式的主要缺点是样本选取的随机性. 矩阵方式最为简单,并且可以应用到时间上不连续的多幅图像情况,但如何更好地刻画矩阵之间的度量是一个值得研究的内容.

4 常用的视频人脸数据库及一些实验结果

目前基于视频的人脸识别常用的视频人脸数据库包括 Mobo(Motion of body)数据库[52]和 Honda/UCSD 数据库[8-9]. Mobo 数据库最初是 CMU 为了 Human ID 计划进行步态识别而采集的数据库. 整个数据库包含 25 个人在跑步机上以四种不同的方式行走的视频序列. 行走的方式包括慢速行走、快速行走、斜面行走和拿球行走. 正面角度拍摄的视频序列共 99 段(一段丢失). UCSD/Honda 数据库包含 20 个人的共 52 段视频. 数据库中的人脸视频包含了大规模的 2D(平面内)和 3D(平面外)的头部旋转. 另外还有 DXM2VTS 数据库[53].

这些视频人脸数据库普遍的缺点是没有考虑到各种条件的变化. 大部分都是姿态的变化,其他的如光照、表情的变化等考虑较少,并且数据库的人偏少(<50 个人),无法进行大规模有效的实验来评价各种算法的优劣.

目前大部分的文献中采用的数据库以及训练,测试方法都不尽相同. 但为了对目前典型方法的实验结果有一个直观的认识,表 4 汇集了在视频人脸数据库上一些典型方法的实验结果.

表 4　视频人脸数据库上典型方法的实验结果

方法名称	识别率	典型文献
PCA(多数投票)	87.1%(MoBo) 89.6%(Honda/UCSD)	文献[44]
LDA(多数投票)	90.8%(MoBo) 86.5%(Honda/UCSD)	文献[44]
隐马尔可夫模型	92.3%(MoBo) 91.2%(Honda/UCSD)	文献[44]
基于贝叶斯框架的 SIS 方法	92%以上(MoBo)	文献[10]
概率外观流形法	98.8%(无遮挡)(Honda/UCSD) 97.8%(有遮挡)(Honda/UCSD)	文献[9]
ARMA 模型	90%左右(Honda/UCSD)	文献[12]
混合高斯模型	94%	文献[13]
局部线性模型	95.62%	文献[21]

5 总结和展望

本文介绍了现阶段基于视频的人脸识别研究进展. 在对人脸识别不同情况分类的基础上,重点介绍了现阶段基于视频的人脸识别的主要方法,分析和讨论了各种方法的优缺点,还介绍了常用的视频人脸数据库及一些典型方法的实验结果.

现阶段基于视频的人脸识别一般都是把人脸视频看成一个整体来克服分辨率低的问题[35]. 对于光照或者姿态的单独变化可以通过矩阵、概率或者流形的方式部分解决[9,18],但是需要不同条件下的大量的训练样本. 对于遮挡问题可以采用鲁棒统计学[11]或者对脸部的分块处理[54]来解决.

随着研究的深入,基于视频的人脸识别需要进一步研究的工作包括:

(1) 人脸特征的准确定位

本文假设已经得到了图像或者视频中人脸的位置,并且人脸的特征已经准确定位. 但是在实际应用中,人脸视频的分辨率过低常会使得人脸的检测和准确的特征定位存在一定的困难. 人脸的误配准也会严重影响人脸识别的结果. 作为人脸识别的基础,准确和快速的人脸检测和特征定位方法是必不可少的.

(2) 人脸的超分辨率重建和模糊复原

视频序列中的人脸由于采集条件和运动的影响,人脸图像分辨率低且人脸模糊. 需要研究人脸图像超分辨率技术[55]和图像复原技术[56]以得到清晰的人脸图像也是未来需要重点解决的问题.

(3) 人脸的 3D 建模

现阶段基于二维的人脸识别方法可以在一定程

度上解决姿态或光照的变化问题.但是人脸是一个三维的物体,利用人脸的三维信息是解决姿态,光照变化问题的最本质方法.现阶段利用视频数据生成3D模型的计算复杂度很大[42,57-59],无法达到使用要求.更好地降低三维人脸建模的复杂度和提高建模的精度是未来发展的一个重要方向.

(4) 视频人脸数据库和测试方法的标准化

与基于静止图像的人脸识别相比,基于视频的人脸识别的最大问题是还没有一个包含各种条件变化的、统一的、大规模的视频人脸数据库和测试标准.许多文章采用的视频人脸数据库和测试方法都不尽相同,无法进行算法之间的比较.建立一个公共的、大规模的视频人脸数据库和标准的测试方法是该领域的一个首要任务.

(5) 多模生物特征认证

现阶段基于视频的人脸识别算法主要是基于室内的环境条件.室外条件下的人脸图像光照、姿态等的剧烈变化使人脸识别仍然面临着许多困难,融合多种生物特征提高识别的性能也将是未来研究的一个重点[60-62].

参 考 文 献

[1] Chellappa R, Wilson C, Sirohey S. Human and machine recognition of faces: A survey. Proceedings of the IEEE, 1995, 83(5): 705-740

[2] Zhao W, Chellappa R, Rosenfeld A, Phillips P J. Face recognition: A literature survey. ACM Computation Survey, 2003, 35(4): 399-458

[3] Li S Z, Jain A K. Handbook of Face Recognition. New York: Springer, 2005

[4] Zhou S, Chellappa R. Beyond a single still image: Face recognition from multiple still images and videos// Zhao W et al eds. Face Processing: Advanced Modeling and Methods. New York: Academic Press, 2005

[5] Shakhnarovich G, Fisher J W, Darrell T. Face recognition from long-term observations// Proceedings of the European Conference on Computer Vision. Bari, 2002: 851-868

[6] Liu X M, Chen T, Thornton S M. Eigenspace updating for non-stationary process and its application to face recognition. Pattern Recognition, 2003, 36(9): 1945-1959

[7] Liu X M, Chen T. Video-based face recognition using adaptive hidden Markov models// Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. Madison, 2003: 340-345

[8] Lee K C, Ho J, Yang M H, Kriegman D. Video-based face recognition using probabilistic appearance manifolds// Proceedings of the International IEEE Conference on Computer Vision and Pattern Recognition. Madison, 2003: 313-320

[9] Lee K C, Ho J, Yang M H, Kriegman D. Visual tracking and recognition using probabilistic appearance manifolds. Computer Vision and Image Understanding, 2005, 99(3): 303-331

[10] Zhou S, Krueger V, Chellappa R. Probabilistic recognition of human faces from video. Computer Vision and Image Understanding, 2003, 91(1): 214-245

[11] Zhou S, Chellappa R, Moghaddam B. Visual tracking and recognition using appearance-adaptive models in particle filters. IEEE Transactions on Image Processing, 2004, 13(11): 1434-1456

[12] Aggarwal G, Chowdhury A K R, Chellappa R. A system identification approach for video-based face recognition// Proceedings of the IEEE International Conference on Pattern Recognition. Cambridge, 2004: 23-26

[13] Arandjelovič O, Cipolla R. Face recognition from face motion manifolds using robust kernel resistor-average distance// Proceedings of the IEEE Conference on Compute Vision and Patter Recognition workshop. Washington D.C, 2004: 88-93

[14] Arandjelovič O, Shakhnarovich G, Fisher G, Cipolla R, Darrell T. Face recognition with image sets using manifold density divergence// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. San Diego, 2005: 581-588

[15] Arandjelovič O, Cipolla R. A pose-wise linear illumination manifold model for face recognition using video. Computer Vision and Image Understanding, 2009, 113(1): 113-125

[16] Yamaguchi O, Fukui K, Maeda K. Face recognition using temporal image sequence// Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition. Nara, 1998: 318-323

[17] Fukui K, Yamaguchi O. Face recognition using multi-viewpoint patterns for robot vision// Proceedings of the International Symposium of Robotics Research. Siena, Italy, 2003: 192-201

[18] Nishiyama M, Yamaguchi O, Fukui K. Face Recognition with the multiple constrained mutual subspace method// Proceedings of the 5th International Conference on Audio- and Video-Based Biometric Person Authentication. New York, 2005: 71-80

[19] Li J W, Wang Y H, Tan T N. Video-based face recognition using a metric of average Euclidean distance// Proceedings of the 5th Chinese Conference on Biometric Recognition. Guangzhou, China, 2004: 224-232

[20] Li J W, Wang Y H, Tan T N. Video-based face recognition using earth mover's distance// Proceedings of the International Conference on Audio- and Video-based person Authentication. New York, 2005: 229-239

[21] Fan W, Wang Y H, Tan T N. Video-based face recognition using Bayesian inference model// Proceedings of the International Conference on Audio- and Video-based Person Authentication. New York, 2005: 122-130

[22] Yan Y, Zhang Y J. State-of-the-art on video-based face recognition. Encyclopedia of Artificial Intelligence, 2008, 1455-1461

[23] Jia H X, Zhang Y J. Human detection in static images// Verma B, Blumenstein M. Pattern Recognition Technologies and Applications: Recent Advances. 2008: 227-243

[24] Liu X M, Zhang Y J, Tan H C. A new Hausdorff distance based approach for face localization. Sciencepaper Online, 2005, 200512-662(1-9)

[25] Srikantaswamy R, Samuel R D S. A novel face segmentation algorithm from a video sequence for real-time face recognition. EURASIP Journal on Advances in Signal Processing, 2007, 2007: 1-6

[26] Wechsler H, Kakkad V, Huang J, Gutta S, Chen V. Automatic video based person authentication using the RBF network// Proceedings of the International Conference on Audio- and Video-Based Person Authentication. Crans-Montana, 1997: 85-92

[27] Steffens J, Elagin E, Neven H. PersonSpotter: Fast and robust system for human detection, tracking and recognition// Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition. Nara, 1998: 516-521

[28] Mckenna S J, Gong S. Non-intrusive person authentication for access control by visual tracking and face recognition// Proceedings of the International Conference on Audio- and Video-Based Person Authentication. Crans-Montana, 1997: 177-183

[29] Kittler J, Hatef M, Duin R P W, Matas J. On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(3): 226-239

[30] Moghaddam B, Pentland A. Probabilistic visual learning for object representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 696-710

[31] Li B, Chellappa R. Face verification through tracking facial features. Journal of the Optical Society of America A, 2001, 18(12): 2969-2981

[32] Li B, Chellappa R. A generic approach to simultaneous tracking and verification in video. IEEE Transactions on Image Processing, 2002, 11(5): 530-554

[33] Torres L, Vila J. Automatic face recognition for video indexing applications. Pattern Recognition, 2002, 35(3): 615-625

[34] Edwards G J, Taylor C J, Taylor T F. Improving identification per formation by integrating evidence from sequences// Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. Fort Collins, 1999: 486-491

[35] Kim M Y, Kumar S, Pavlovic V, Rowley H. Face tracking and recognition with visual constraints in real-world videos// Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, 2008: 1-8

[36] McKenna S, Gong S, Raja Y. Face recognition in dynamic scenes// Proceedings of the British Machine Vision Conference. Colchester, 1997: 140-151

[37] Park U, Jain A K, Ross A. Face recognition in video: Adaptive fusion of multiple matchers// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, 2007: 1-8

[38] Wolf L, Shashua A. Kernel principal angles for classification machines with applications to image sequence interpretation// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Madison, 2003: 635-642

[39] Fan W, Yeung D Y. Locally linear models on face appearance manifolds with application to dual-subspace based classification// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York, 2006: 1384-1390

[40] Tenenbaum J B, Silva V D, Langford J C. A global geometric framework for nonlinear dimensionality reduction. Science, 2000, 290(5500): 2319-2323

[41] Moghaddam B, Jebara T, Pentland A. Bayesian face recognition. Pattern Recognition, 2000, 33(11): 1771-1782

[42] Xu Y, Roy-Chowdhury A, Patel K. Pose and illumination invariant face recognition in video// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, 2007: 1-7

[43] Xu Y L, Roy-Chowdhury A, Patel K. Integrating illumination, motion, and shape models for robust face recognition in video. Eurasip Journal on Advances in Signal Processing, 2008, 2008: 1-13

[44] Zhou S, Chellappa R. From sample similarity to ensemble similarity: Probabilistic distance measures in reproducing kernel Hilbert space. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(6): 917-929

[45] Hadid A, Pietikäinen M. From still image to video-based face recognition: An experimental analysis// Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition. Seoul, 2004: 813-818

[46] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding. Science, 2000, 290(5500): 2323-2326

[47] Li Y, Gong S, Lidell H. Modeling faces dynamically across views and over time// Proceedings of the IEEE International Conference on Computer Vision. Vancouver, 2001: 554-559

[48] Li Y, Gong S, Lidell H. Video-based online face recognition using identity surfaces// Proceedings of the IEEE International Conference on Computer Vision. Vancouver, 2001: 40-46

[49] Li Y, Gong S, Lidell H. Constructing facial identity surfaces in a nonlinear discriminating space// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Kauai, 2001: 258-263

[50] Yan Y, Zhang Y J. Discriminant projection embedding for face and palmprint recognition. Neurcomputing, 2008, (16-18): 3534-3543

[51] Yan S C, Xu D, Zhang B, Zhang H J. Graph embedding and extensions: A general framework for dimensionality reduction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(1): 40-51

[52] Gross R, Shi J. The CMU Motion of Body (MoBo) database. Robotics Institute, Carnegie Mellon University: Technical Report CMU-RI-TR-01-18, 2001

[53] Teferi D, Bigun J. Damascening video databases for evaluation of face tracking and recognition —The DXM2VTS database. Pattern Recognition Letters, 2007, 28(15): 2143-2156

[54] Zhang Y, Martinez A M. A weighted probabilistic approach to face recognition from multiple images and video sequences. Image and Vision Computing, 2006, 24(6): 626-638

[55] Al-Azzeh M, Eleyan A, Demirel H. PCA-based face recognition from video using super-resolution// Proceedings of the 23rd International Symposium on Computer and Information Sciences. Istanbul, 2008: 1-4

[56] Goksel D. Exploiting space-time statistics of videos for face hallucination[Ph.D. dissertation]. Carnegie Mellon University, Pittsburgh, USA, 2007

[57] Chowdhury A, Chellappa R. Face reconstruction from monocular video using uncertainty analysis and a generic model. Computer Vision and Image Understanding, 2003, 91(1): 188-213

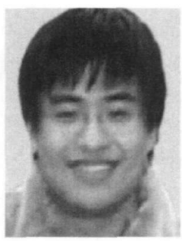
[58] Choudhury A, Clarkson B, Jebara T, Penland A. Multimodal person recognition using unconstrained audio and video// Proceedings of the Conference on Audio- and Video-based Biometric Person Authentication. Washington D. C, 1999: 176-180

[59] Zhang Z Y, Liu Z C, Adler D, Cohen M F, Hanson E, Shan Y. Robust and rapid generation of animated faces from video images: A model-based modeling approach. International Journal of Computer Vision, 2004, 58(2): 93-119

[60] Zhou X, Bhanu B. Integrating face and gait for human recognition at a distance in video. IEEE Transactions on Systems, Man and Cybernetics, Part B, 2007, 37(5): 1119-1137

[61] Jing X Y, Yao Y F, Zhang D, Yang J Y, Li M. Face and palmprint pixel level fusion and kernel DCV-RBF classifier for small sample biometric recognition. Pattern Recognition, 2007, 40(11): 3209-3324

[62] Yan Y, Zhang Y J. Multimodal biometrics fusion using correlation filter bank// Proceedings of the 19th IAPR International Conference on Pattern Recognition. Tampa, 2008, MoBT7.3(1-4)



YAN Yan, born in 1984, Ph.D.. His main research interests focus on pattern recognition.

ZHANG Yu-Jin, born in 1954, Ph.D., professor, Ph.D. supervisor. His main research interests include image engineering (image processing, image analysis, image understanding and technique application). http://www.ee.tsinghua.edu.cn/~zhangyujin/

Background

This work is supported by the National Natural Science Foundation of China under grant No. 60872084 and the Specialized Research Fund for the Doctoral Program of Higher Education under grant No. 20060003102.

Traditional still image-based face recognition has achieved great success in constrained environments. However, once the conditions, including illumination, pose, expression, age, etc., change too much, the performance declines dramatically. The recent FRVT2002 shows that the recognition performance of face images captured in an outdoor environment and different days is still not satisfying. Current still image-based face recognition algorithms are even far away from the capability of human perception system. On the other hand, psychology and physiology studies have shown that motion can help people for better face recognition.

During the past several years, many research efforts have been concentrated on video-based face recognition. Compared with still image-based face recognition, true video-based face recognition algorithms that use both spatial and temporal information started only a few years ago. No comprehensive survey in this field has been made, and a lot of issues in video-based face recognition still have not been addressed well. So the content of this paper gives an overview of the most existing methods in the field of video-based face recognition. A suitable classification for different methods has been made, the respective pros and cons of typical techniques in each method group are analyzed. The important issues which need to be solved, the prospects for future development and some suggestions for further research works are put forward to meet the goal of this paper.