基于改进的 Fisher 准则的多示例学习视频人脸识别算法

王 玉 [1, 2, 3] 申铉京 [1, 3] 陈海鹏 [1, 3]

摘 要 视频环境下目标的姿态变化使得人脸关键帧难以准确定位, 导致基于关键帧标识的视频人脸识别方法的识别率偏低. 为解决上述问题, 本文提出一种基于 Fisher 加权准则的多示例学习视频人脸识别算法. 该算法将视频人脸识别视为一个多示例问题, 将视频中归一化后的人脸帧图像作为视频包中的示例, 采用分块 TPLBP 级联直方图作为示例纹理特征, 示例特征的权值通过改进的 Fisher 准则获得. 在训练集合的示例特征空间中, 采用多示例学习算法生成分类器, 进而实现对测试视频的分类及预测. 通过在 Honda/UCSD 视频库和 Youtube Face 数据库中的相关实验, 该算法达到了较高的识别精度, 从而验证了算法的有效性. 同时, 该方法对均匀光照变化、姿态变化等具有良好的鲁棒性.

关键词 视频人脸识别, 局部二值模式, 多示例学习, Fisher 准则

引用格式 王玉, 申铉京, 陈海鹏. 基于改进的 Fisher 准则的多示例学习视频人脸识别算法. 自动化学报, 2018, **44**(12): 2179−2187

DOI 10.16383/j.aas.2018.c170090

Video Face Recognition Based on Modified Fisher Criteria and Multi-instance Learning

WANG Yu[1, 2, 3] SHEN Xuan-Jing[1, 3] CHEN Hai-Peng[1, 3]

Abstract Due to the pose variation of target in video, it is difficult to accurately locate the face key frame and have a high recognition rate of the video face recognition based on key frame identification. To solve these problems, a video face recognition algorithm based on multi-instance learning is proposed in this paper. The algorithm takes each face video as a bag, and each normalized face frame as an instance in the bag. The feature of each instance is represented by cascading histograms of block TPLBP codes, and the weight of the instance feature is obtained by the improved Fisher criteria. The classifier is obtained in the feature space of training set by using a multiple instance learning algorithm, and then classification and prediction of test bag are realized. Experiments on the Honda/UCSD and YouTube Face databases show that the algorithm can achieve a higher recognition accuracy, and at the same time, the method is robust to illumination variation and expression variation.

Key words Video-based face recognition, local binary patterns (LBP), multi-instance learning, Fisher criteria

Citation Wang Yu, Shen Xuan-Jing, Chen Hai-Peng. Video face recognition based on modified Fisher criteria and multi-instance learning. *Acta Automatica Sinica*, 2018, **44**(12): 2179−2187

视频人脸识别是计算机视觉、模式识别、视频分析与理解等领域的重要研究课题. 视频人脸识别的研究不仅在理论上具有重大意义, 同时在生物特征鉴别、视频监控、信息安全等领域具有广泛的应用前景, 已经成为人脸识别领域的研究热点和难点问题[1].

视频环境下人脸识别中的关键问题是确定整个视频中对识别结果起决定性作用的最具代表性帧, 而对整个视频进行分析并选取关键帧的做法往往是不现实的. 目前, 绝大多数的视频人脸识别都是通过提取视频序列中包含人脸的关键帧, 采用基于静态图像的人脸识别算法达到视频分类的目的, 其中包括多视角融合、子空间或流形分析[2] 等. 该类方法中关键帧的选择歧义性较大, 需要对整个视频进行分析才能实现关键帧的准确定位, 降低了视频人脸识别系统的效率和实时性要求, 适合于人物目标配合、光照及视角良好并且视频质量较高的环境下的应用.

近年来, 基于图像集合和基于视频序列的视频人脸识别方法得到了广泛关注. 其中, 基于图像集合

收稿日期 2017-02-24 录用日期 2017-07-12
Manuscript received February 24, 2017; accepted July 12, 2017
国家青年科学基金 (61305046, 61602203), 吉林省优秀青年人才基金 (20180520020JH) 资助
Supported by National Science Foundation for Young Scientists of China (61305046, 61602203) and Outstanding Young Talent Foundation of Jilin Province (20180520020JH)
本文责任编委 杨健
Recommended by Associate Editor YANG Jian
1. 吉林大学计算机科学与技术学院 长春 130012 2. 吉林大学应用技术学院 长春 130012 3. 吉林大学符号计算与知识工程教育部重点实验室 长春 130012
1. College of Computer Science and Technology, Jilin University, Changchun 130012 2. Applied Technology College, Jilin University, Changchun 130012 3. Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012

的视频人脸识别方法是将视频作为一个无序的帧图像集合, 通过流形[3−5]、子空间[6]、Affine Hull[7]、协方差矩阵[8] 等对图像集合进行建模以实现视频人脸的识别. Cevikalp 和 Triggs[9] 在 IEEE 国际计算机视觉与模式识别会议 (IEEE Conference on Computer Vision and Pattern Recognition, CVPR) 上提出了基于图像集合的视频人脸识别方法, 指出该类方法包括两个主要方面: 1) 如何对人脸图像集合进行建模; 2) 如何度量模型之间的相似度, 通过凸包和仿射变换对图像集合建模并度量模型之间的几何距离以实现视频人脸识别. Hu 等[10] 在此基础上提出了基于 SANP 的图像集合视频人脸识别方法, 相比之前的算法, 能得到更好的性能. 于谦等[11] 提出了判别性联合多流形分析 (Discriminative joint multi-manifold analysis, DJMMA) 方法, 该方法将基于视频的人脸识别转换为图像集识别问题, 并提出用类间流形示每个图像集的平均脸信息, 用类内流形表示每个图像集的所有原始图像的信息. 采用分片技术学习两种流形的投影矩阵, 并提出了与分片技术相匹配的流形之间的距离度量方法. 实验结果表明, 该方法在几个公开研究的视频库中, 比现有的方法具有更高的识别正确率.

基于视频序列的视频人脸识别算法是通过设计视频纹理描述算子, 引入视频上下文信息等方式提高识别精度和效率. Zhao 等[12] 提出了基于 LBP 的具有旋转不变性的视频纹理描述算子, 该方法在 DynTex 数据库中的识别率达到了 98.57 %. 但是, 视频纹理算子及视频上下文信息的获取十分困难, 计算复杂度较高. 同时, 这类算法对人脸表情和目标姿态变化等影响识别性能的因素不够鲁棒[1]. 视频相对于图像提供的可用信息更多, 但也会带来更多的噪音等干扰因素.

自然视频大多数是非专业人员采集的, 视频采集设备有限, 视频环境光照条件较差, 目标姿态多变并且伴随运动模糊. 同时, 为了便于存储及传输, 通常以压缩格式存储, 这些噪音因素的存在都使得解决视频人脸识别问题具有极大的挑战性[13]. 为实现这种低分辨率、目标姿态多变条件下的视频人脸的鲁棒识别, 设计能够适应这种复杂环境下的视频人脸识别学习算法就变得尤为重要.

1 主要工作

视频采集会受到光照、表情, 尤其是姿态变化的影响, 视频人脸序列中有用信息的出现概率较低, 甚至会出现严重的数据缺失. 视频中有用信息出现的时间没有规律可循, 用常规的学习方法很难解决这种复杂条件下的视频人脸识别.

为解决视频环境下人脸识别问题中关键帧难以准确定位导致的识别率偏低等问题, 本文提出一种基于多示例学习的视频人脸识别算法, 多示例学习[14] 算法对于解决视频分类这种多示例问题性能优越, 已经应用于视频事件检测[15]、恐怖视频分类[16] 等领域. Yang 等[17] 将多示例学习模型用于新闻视频中的人脸标识, 并提出 Exclusive diversity 和 Iterative ED 两种可辨别概率模型用于解决这类多示例问题. 多示例学习模型被认为非常适合低信噪比或数据缺失严重的环境下的概念学习, 受到机器学习界广泛的重视并成为当前研究的热点之一.

对于解决像视频人脸识别这样的实际应用问题, 虽然单个视频对象中只包含同一类目标, 我们可以将视频中的每一个帧图像作为一个目标描述, 但是由于光照、姿态变化及表情的影响, 一个视频对象可能同时存在多个描述, 而这些图像描述中往往只有一个或几个描述具有决定性, 即所谓的关键帧, 而到底哪个图像描述可以决定视频对象的类别往往是不确定的. 而多示例学习算法正是解决这种 “对象: 描述: 类别” 之间的 “$1:N:1$” 关系的学习模型, 并被认为是与三种传统学习框架并列的第四种学习框架.

本文的主要工作是将多示例学习方法应用于视频人脸识别, 以解决传统方法需要在视频中准确定位代表性关键帧的难题. 本文算法将复杂环境下的视频人脸识别问题视为一个多示例问题, 将训练集合中的每个视频视为一个包 (Bag), 将视频包中归一化处理后的视频帧图像视为包中的示例 (Instance). 视频包带有标记而视频包中的示例没有标记, 如图 1 所示, 这里包含一个正包 (Positive bag) 和一个负包 (Negative bag), 利用有效的多示例学习算法在训练集合样本空间中学习并生成分类器, 以实现对测试包的预测及分类.

另外, 视频采集环境的光照变化、目标的姿态变化等, 都在一定程度上造成了视频人脸识别上的困难, 为此, 本文在算法实现过程中采用了基于改进的 Fisher 加权准则的 TPLBP (Three-patch local binary patterns) 进行示例的纹理特征表示, 该算子具有较强的可辨别能力, 并且对均匀光照变化是鲁棒的. 实验结果表明, 本文算法具有较高的识别精度和效率.

1.1 算法描述

本文算法主要步骤如下:

步骤 1. 视频包预处理. 训练视频集中的每个视频 (包) 带有标记, 而视频中的每个帧图像 (示例) 没有标记. 从视频包中的帧图像 (示例) 中检测出含人脸图像并预处理后, 以双眼坐标为基准进行人脸归一化处理.

步骤 2. 示例局部纹理特征提取. 对包中的每

个示例图像划分分块, 对每个分块提取 TPLBP 纹理特征, 获得每个分块级的局部直方图统计信息, 以获得人脸示例的局部纹理信息.

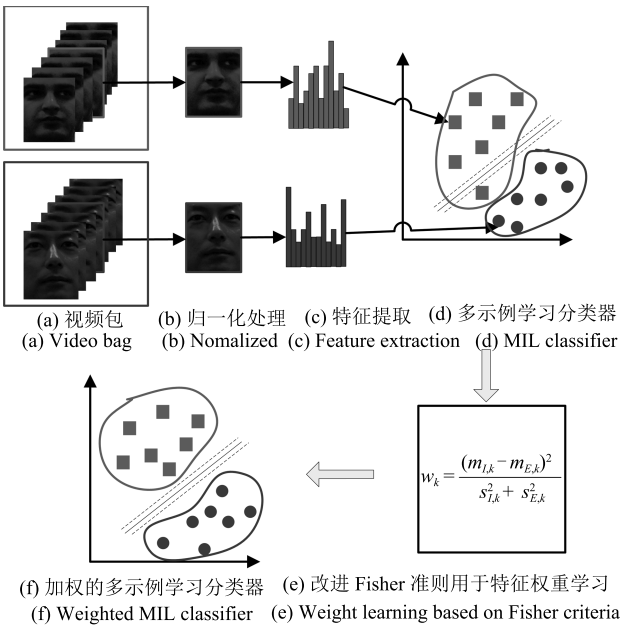


图 1　本文提出的基于多示例学习框架的视频人脸识别算法框架

Fig. 1　The framework of proposed video face recognition algorithm based on multi-instance learning

步骤 3. 示例全局纹理特征表示. 将每个示例的所有分块的 TPLBP 特征直方图进行级联操作, 构成每个示例的全局纹理特征直方图, 从而获得视频包中所有人脸示例的全局纹理特征表示.

步骤 4. 特征权重学习. 对 Fisher 准则加以改进, 同类样本的不同样本个体之间的相关性分布构成了类内相关度空间, 不同类样本之间的相关性分布形成了类间相关度空间. 统计计算得到特征直方图中每个特征的类内相关度均值和方差, 进而获得每个特征的对应权重信息.

步骤 5. 多示例分类器学习. 在所有训练集合构成的多示例特征空间中训练生成多示例学习分类器. 在分类识别阶段, 将待检测的人脸视频包送入分类器, 通过直方图相交的匹配方式对测试包进行识别, 并通过相关对比实验验证本文算法的性能.

1.2　本文所提方法的贡献

1) 采用多示例学习方法解决视频人脸识别问题, 可以有效避免传统方法需要定位视频中的代表性帧的问题. 另外, 该方法对姿态变化等因素较为鲁棒.

2) 采用改进的 Fisher 加权准则对示例纹理特征的权重进行学习. 同类样本的不同样本个体之间的相关性分布构成了类内相关度空间, 不同类样本之间的相关性分布形成了类间相关度空间. 统计计算得到特征直方图中每个特征的类内相关度均值和方差, 进而获得每个特征的对应权重信息, 有效提高了算法的识别性能.

3) TPLBP 算子具有灰度平移不变性, 即 TPLBP 算子对均匀光照变化是鲁棒的, 同时该纹理描述算子的使用使得该算法对人脸表情变化较为鲁棒.

2　基于 Fisher 加权准则的视频人脸识别算法

如图 1 所示, 本文将每个人脸视频视为一个包, 假设视频集合有 $m+n$ 个视频包, 其中 n 个正包和 m 个负包. 每个视频包包含 p 个帧图像, 将其视为包中的中 p 个示例, 即表示第 i 个包中的第 j 个示例.

2.1　示例的纹理特征提取

2.1.1　TPLBP 纹理特征

LBP 算子[18] 是一种从纹理局部近邻定义中衍生出来、灰度范围内的纹理度量算子, 具有很强的分类能力、较高的计算效率、灰度平移不变性和旋转不变性等特点. 在 LBP 的各种模式中, 有一部分模式出现的概率相当高, 占据了绝大多数的纹理信息, 称为等价模式 LBP (Uniform LBP, ULBP). 等价模式的特点是在 LBP 的二进制编码中, 最多有两个 0 到 1 (或 1 到 0) 的变化. 表示一种等价模式的 LBP 算子, 采用了等价模式后, 二进制模式大大减少, 模式的数量从开始的 2^p 减少到 $p\ (p-1)+2$ 种 (降维). TPLBP 是 Wolf 等[19] 2008 年提出的基于分块的 LBP 纹理描述算子, 在人脸及图像相似度学习中具有较好的性能, 已经在很多目标多级识别系统获得成功. 在纹理分割系统中, 通过观察中心分块与边缘分块之间的交叉关系获得局部纹理特征对图像类型及局部变化具有较好的鲁棒性.

TPLBP 算子的计算公式如下:

$$TPLBP_{r,S,\omega,\alpha}(p) = \sum_{i}^{S} f(d(C_i, C_p) - d(C_{(i+\alpha)modS}, C_p))2^i$$

$$f(x) = \begin{cases} 1, & \text{若 } x > \xi \\ 0, & \text{若 } x \le \xi \end{cases} \tag{1}$$

其中, 参数 S 表示邻域像素块数, 参数 α 表示参与计算的邻域像素块间隔, r 代表邻域半径, ω 表示像素块窗口大小. 通过三个块之间的比较获得每个

像素编码的取值, 取值决定于其中两个块哪个块与中心块的相似度更大. C_i 和 $C_{(i+\alpha)modS}$ 代表邻域圆周上间隔为 α 的两个像素块, C_p 代表中心像素块. $d(\cdot,\cdot)$ 表示两个像素块之间的距离, 通过 ξ 值保证统一区域的稳定性, 这里取 $\xi = 0.01$. 人脸识别过程中, 光照变化对识别率的影响很大, 从 TPLBP 算子的计算公式可以看出, TPLBP 算子具有灰度平移不变性, 即 TPLBP 算子对均匀光照变化是鲁棒的.

2.1.2 纹理特征级联表示

对于一幅 $M \times N$ 大小的人脸示例图像来说, 当图像中的每一个像素点都统计得到 TPLBP 编码值后, 可以按如图 2 所示的方法建立一个统计直方图用于表示示例图像的纹理, 直方图特征的建立方式如下:

$$H(k) = \sum_{m-1}^{M}\sum_{n-1}^{N} f(TPLBP_{r,S,\omega,\alpha}(m,n),k)$$
$$k \in [0,d],\ \ f(x,y) = \begin{cases} 1, & 若\ x > y \\ 0, & 其他 \end{cases} \tag{2}$$

其中, d 代表最大的 TPLBP 编码值. 通过 TPLBP 对人脸描述的基本思想是先对人脸图像划分分块, 利用纹理描述算子对每个分块提取局部人脸描述特征, 然后把局部描述直方图级联组合以形成全局纹理描述, 已广泛应用于人脸识别与图像纹理表示.

2.2 多示例学习与权重分配

2.2.1 多示例学习算法

DD (Diverse-density) 算法[20] 和基于 EM (Expectation-maximization) 策略的变体 EM-DD 算法[21] 是目前应用最为普遍的多示例学习算法. DD 算法是一种基于概率统计的多示例学习算法, 特征空间中某个点的多样性密度的定义为有多少个不同的正包有距离该点足够近的示例, 同时来自反包的示例远离该点的程度的度量. 特征空间中某点附近来自正包的示例越多, 来自负包的示例越远, 则该点的多样性密度越大. DD 算法的目的就是找到特征空间中的多样性密度最大点 $Objective$.

上述训练视频集合 D 中共有 n 个正包和 m 个负包, 并且, 这 $m+n$ 个包是 D 中的所有示例, 则特征空间中任一目标点 t 的多样性密度定义如下:

$$DD(t) = \Pr(t|B_1^+,\cdots,B_n^+,B_1^-,\cdots,B_m^-) \tag{3}$$

对式 (3) 应用 Bayes 规则, 得

$$DD(t) = \frac{\Pr(B_1^+,\cdots,B_n^+,B_1^-,\cdots,B_m^-|t)\Pr(t)}{\Pr(B_1^+,\cdots,B_n^+,B_1^-,\cdots,B_m^-)} \tag{4}$$

在对多示例学习问题的分析中, $\Pr(t)$ 代表与目标相关的先验知识, 可以视为常量. $\Pr(B_1^+,\cdots,B_n^+,B_1^-,\cdots,B_m^-)$ 也是一个与目标相关的常量. 多示例学习的目的是得到最大化的 t, 所以上述两个概率值可以被认为是一个归一化项, 不需要明确计算. 则式 (3) 可表示为

$$\Pr(B_1^+,\cdots,B_n^+,B_1^-,\cdots,B_m^-|t) \tag{5}$$

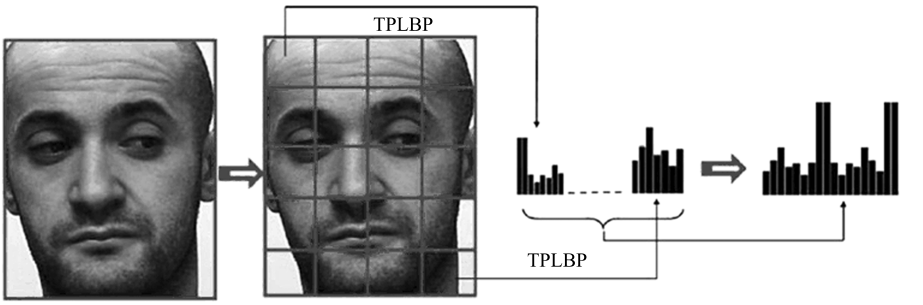
在各包条件独立的情况下, 对式 (5) 进一步简化, 得到多样性密度的最终表示为

$$\prod_{i=1}^{n}\Pr(B_i^+|t)\prod_{i=1}^{m}\Pr(B_i^-|t) \tag{6}$$

通过 DD 算法获得的多样性密度最大点 $Objective$ 为

$$Objective = \arg\max_{t\in\mathbf{R}^d} DD(t) \tag{7}$$

为便于计算 $DD(t)$, 假设各包独立, 由 Bayes 理论可将式 (6) 变换为



(a) 归一化后的人脸图像 (a) Normalized face image
(b) 分块的人脸图像 (b) Image divided into blocks
(c) 分块的 TPLBP 特征直方图 (c) Local TPLBP feature histogram
(d) 全局特征直方图 (d) Global feature histogram

图 2 通过局部分块特征直方图级联表示人脸示例纹理

Fig. 2 Face instance texture is represented by cascading local block feature histogram

$$\prod_{i=1}^{n}\frac{\Pr(t|B_i^+)\Pr(B_i^+)}{\Pr(t)}\prod_{i=1}^{m}\frac{\Pr(t|B_i^-)\Pr(B_1^-)}{\Pr(t)}=\prod_{i=1}^{n}\Pr(t|B_i^+)\prod_{i=1}^{m}\Pr(t|B_i^-) \tag{8}$$

假设每个包中有 p 个示例, 则式 (5) 中的概率值可通过 noisy-or 模型计算得到, 正包的概率为

$$\Pr(t|B_i^+)=\Pr(t|B_{i1}^+,\cdots,B_{ip}^+)=1-\prod_{j=1}^{p}(1-\Pr(B_{ij}^+\in c_t)) \tag{9}$$

负包的概率为

$$\Pr(t|B_i^-)=\prod_{j=1}^{p}(1-\Pr(B_{ij}^-\in c_t)) \tag{10}$$

2.2.2 纹理特征的权重学习

本文算法中的视频集合的每个示例特征是通过 TPLBP 特征直方图表示的, 特征直方图相似度匹配的一般方式包括直方图相交、Chi 平方概率统计及 Log 概率统计等. 本文选择直方图相交的匹配方式, 匹配方式为

$$D(S,M)=\sum_{k=1}^{d}\min(\boldsymbol{S}_k,\boldsymbol{M}_k) \tag{11}$$

其中, $\boldsymbol{S}_k$ 和 $\boldsymbol{M}_k$ 分别表示任意两个示例 S 和 M 的统计直方图的第 k 维特征向量.

Fisher 准则[22] 是一种传统的线性判别方法, 在模式识别领域得到广泛应用, 基本原理是寻找特征空间的某个投影子空间, 使得所有特征点在该子空间得到最好的分类. 针对本文提到的视频人脸识别问题, 对 Fisher 准则进行改进以应用于纹理特征的权值学习.

结合上述分析, 本文对 TPLBP 特征直方图的加权规则如下: 1) 相关特征设置的权值较大, 使得正包与 *Objective* 的距离变短, 正包中的示例变得更加紧密, 增大了多样性密度值; 2) 不相关特征设置的权值较小, 负包与 *Objective* 的距离增大, 同样增大了多样性密度值. 同类样本的不同样本个体之间的相关性分布构成了类内相关度空间, 不同类样本之间的相关性分布形成了类间相关度空间. 每个特征的类内相关度均值和方差可通过计算得到.

$$m_{I,k}=\frac{1}{2}\left(\frac{2}{N^+(N^+-1)}\sum_{i=2}^{N^+}\sum_{j=1}^{i-1}D(S_k^i,M_j^k)\right)+\frac{1}{2}\left(\frac{2}{N^-(N^--1)}\sum_{i=2}^{N^-}\sum_{j=1}^{i-1}D(S_k^i,M_j^k)\right) \tag{12}$$

$$S_{I,k}^2=\sum_{i=2}^{N^+}\sum_{j=1}^{i-1}(D(S_k^i,M_j^k)-m_{I,k})^2+\sum_{i=2}^{N^-}\sum_{j=1}^{i-1}(D(S_k^i,M_j^k)-m_{I,k})^2 \tag{13}$$

其中, S_k^i 和 M_k^j 分别表示同一分类中的第 i 个和第 j 个样本第 k 个特征值, N^+ 表示正示例的数目, N^- 表示负示例的数目. 同理, 可以得到每个特征的类间相关度均值和方差.

$$m_{E,k}=\frac{1}{N^+N^-}\sum_{i=1}^{N^+}\sum_{j=1}^{N^-}D(S_k^i,M_k^j) \tag{14}$$

$$S_{E,k}^2=\sum_{i=1}^{N^+}\sum_{j=1}^{N^-}(D(S_k^i,M_k^j)-M_{E,k})^2 \tag{15}$$

其中, S_k^i 表示第 i 个正示例的第 k 个特征值, M_k^j 表示第 j 个负示例的第 k 个特征值, 最终第 k 特征值的权重为

$$w_k=\frac{(m_{I,k}-m_{E,k})^2}{S_{I,k}^2+S_{E,k}^2} \tag{16}$$

通过上述分析, 多示例学习完成后, 在获得 *Objective* 的同时得到一个权值向量 $\boldsymbol{w}$, 该权值向量反映了示例不同特征值对算法性能的贡献, 基于加权的直方图相交度量方式如式 (17) 所示.

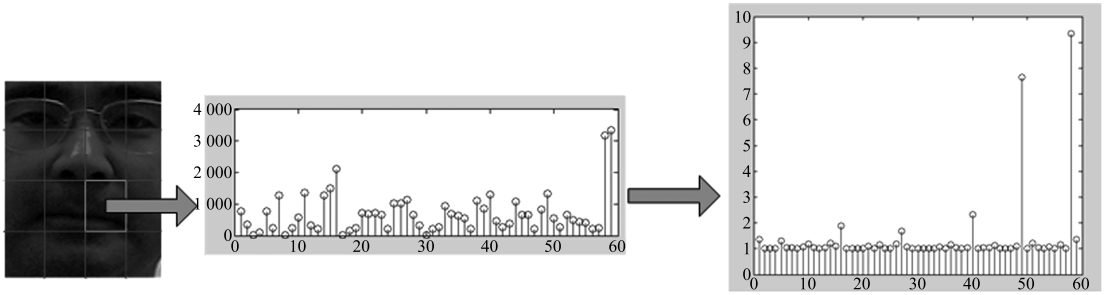
$$\Pr(B_{ij}\in c_t)=\exp\left[1-\sum_{k=1}^{d}w_k^2\left(\min(B_{ijk},c_{tk})\right)^2\right] \tag{17}$$

其中, $\boldsymbol{w}$ 代表获得的权值向量, 是一个非负向量. 从式 (17) 的定义可以看出, 如果向量值较小, 则说明这个属性很不相关; 若越大, 说明这个属性越重要. 图 3 是人脸示例一个分块的 TPLBP 特征直方图及对应的权值. 从图 3 可以看出, 特征 49 和 58 对该示例的分类具有决定性的作用, 可以进一步增强人脸的纹理表示性能.

2.2.3 预测及分类

DD 算法及 EM-DD 算法学习到的概念是空间中多样性密度最大点 *Objective*. 为了实现对测试包的预测及分类, 需要计算 *Objective* 与测试包之间的距离. 如果测试包与 *Objective* 之间的距离小于分类阈值, 那么将其作为正包, 否则, 将其归为反包. 任一视频包 $B_i=\{B_{i1},\cdots,B_{ij},\cdots,B_{ip}\}$ 与 *Objective* 间的距离计算如下:

$$Dis(B_i,Objective)=\min_j D(B_{ij},Objective) \tag{18}$$



(a) 分块的人脸图像 (b) 单个分块的 TPLBP 直方图 (前 60 维), X 轴表示直方图维度, Y 轴表示对应的像素数目 (c) 特征直方图对应的权值 (前 60 维), X 轴表示直方图维度, Y 轴表示对应的特征权值

(a) Blocked face image (b) TPLBP feature histogram for one block (first 60D) (c) Weight for feature histogram (first 60D)

图 3 分块 TPLBP 特征直方图 (前 60 维) 及各特征值对应的权值

Fig. 3 Feature value of the TPLBP histogram (the first 60 dimensions) and the corresponding weights

分类阈值可以在候选阈值中选择, 将 $Objective$ 与 $D = \{B_1^+; \cdots; B_n^+; B_1^-; \cdots; B_m^-\}$ 中的每个训练包的距离求出来, 然后将这些距离排序. 排序后相邻两个距离间的平均值就是一个候选阈值. 当某一候选阈值可以最大程度地将训练集合中的包进行正确分类时, 将该候选阈值作为最终的分类阈值.

3 实验与分析

目前, 基于视频的人脸识别常用的数据库包括 YouTube Face 数据库和 Honda/UCSD 数据库[23] 等. 本文分别在这两个视频人脸数据库中进行实验. 对于视频人脸识别问题, 准确率 (Accuracy) 是最直接有效的算法性能评价标准, 本文也重点从这一性能指标对算法性能进行测试.

3.1 Honda/UCSD 数据中的相关实验

Honda/UCSD 数据库包含 20 个人的视频信息, 其中训练集视频 20 段, 测试集视频 39 段, 都是在不同的时间段采集的. 该数据库中的视频中包含了大规模的 2D (平面内) 和 3D (平面外) 的头部旋转, 同时伴随着光照及表情变化. 训练集和测试集中的目标始终处于运动状态, 姿态是变化的, 这些因素在很大程度上影响了视频人脸识别算法的精度和效率. 在该数据库的测试集合中, 每个测试视频在训练集都有对应的分类, 而且每个测试视频都会最终归于这 20 个分类中的一类.

3.1.1 算法参数分析

本文设定训练集合中每个分类含 6 个包, 每个包含 3 个示例, 每个测试包含 9 个示例, 采用各种纹理描述算子提取各示例的特征直方图作为示例特征. 本文提取的视频人脸图像经归一化后为 183 像素 × 229 像素.

对于 LBP 算子及其等价模式, P 和 R 的取值决定了纹理表示的细致程度, 同时决定了特征直方图的维度. 不同参数在视频人脸数据库中的首选识别率如表 1 所示. 从表 1 可以看出, 随着 P 值的增大, 纹理描述更加细致, 算法性能得到一定程度的提高, 但是会带来特征维度的增大, 增加了算法复杂度. 通过文献 [18] 中的相关实验分析, 为了有效控制算法的时间复杂度并保证算法性能, 本文选择 $P = 8$ 和 $R = 1$ 的经验值, 同时采用了 LBP 算子的等价模式, 以降低特征维度.

表 1 LBP 算子不同参数在 Honda/UCSD 视频人脸数据库中的首选识别率

Table 1 Recognition rate of different parameters of LBP operator on Honda/UCSD database

Algorithm	P	R	Dim	Accuracy (%)
ULBP + DD	8	1	59	71.8
ULBP + DD	4	1	15	66.7
ULBP + DD	8	2	59	71.8
ULBP + DD	4	2	15	64.1
LBP + DD	8	1	256	76.9
LBP + DD	4	1	16	66.7
LBP + DD	8	2	256	74.4
LBP + DD	4	2	16	66.7

对于 TPLBP 算子, 邻域块数量和邻域圆周半径对算子性能影响较大. 文献 [19] 中对该算子相关参数的影响进行了详细分析. 表 2 给出了不同视频人脸数据库中的首选识别率, 参数 S 和 γ 对算法性能的影响如图 4 所示. 从表 2 和图 4 的实验数据可以看出, 随着 S 和的值的变化, 纹理表示变化明显, 算法性能受到的影响较大. 为了保证算法效率, 本文选择 $S = 8$, 即考虑 8 个窗口为 3×3 的邻域像素块.

3.1.2 对比实验与分析

为了说明本文算法的优越性, 进行了与相关基准算法的对比实验. 考虑了不同类型的视频人脸识

别算法, 具体说明如下.

表 2　TPLBP 算子不同参数在 Honda/UCSD 视频人脸数据库中的首选识别率

Table 2　Recognition rate of different parameters of TPLBP operator on Honda/UCSD database

Algorithm	S	γ	ω	α	Dim	Accuracy
TPLBP + EMDD	8	2	3	5	256	76.9
TPLBP + EMDD	4	2	3	5	16	66.7
TPLBP + EMDD	8	4	3	5	256	74.4
TPLBP + EMDD	4	4	3	5	16	66.7

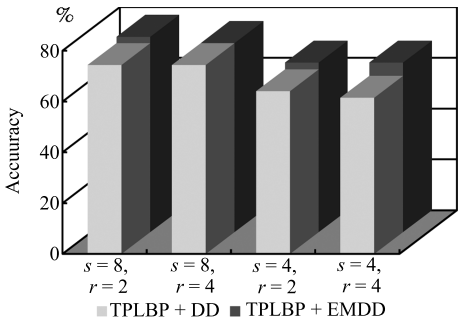


图 4　参数 S 和 γ 对算法性能的影响

Fig. 4　Parameter S and γ's effect on the performance of the algorithm

1) 多示例学习策略的方法 (本文算法). 训练集合中每个分类含 6 个包, 每个包中含 3 个示例, 每个测试包含 9 个示例, 多示例学习方法选择 DD 算法及 EMDD 算法.

2) 基于关键帧的方法. 为更好地评价算法性能, 进行了与经典的监督学习方法的对比实验. 支持向量机 (Support vector machine, SVM) 分类器采用了 Chang 等[24] 开发设计的 LibSVM 开发包, 并使用线性核函数作对比分析. 本文在实现过程中采用 One-against-one 的多分类模型, 即每两个不同分类建立一个二分类的 SVM 分类器, 对于 Honda/UCSD 数据库中的 20 个类别共需要 190 个子分类器.

3) 从帧集合角度进行识别的方法. GLBP-TOP (Gabor local binary patterns from three orthogonal planes) 算法[25] 和 VLBP (Volume local binary patterns) 算法[26], 该类方法将这 9 个示例作为一个帧集合提取时空连续性特征, 采用 Chi 平方概率统计来度量两个特征直方图序列之间的相似度, 并采用 1NN (Nearest neighbor) 方法进行分类.

本文分析了上述三类算法的识别性能, 各方法的分类结果如表 3 所示, 不同算法的 CMC 曲线如图 5 所示. 从实验结果可以看出, 通过采用多示例学习算法可以有效提高视频人脸识别问题的识别率, 通过在不同时间段采集的测试视频集合中与其他人脸识别方法进行对比, 在识别率上提高了近 10 %, 充分验证了该方法的有效性. 同时, 由于采用了有效的人脸纹理表示算子, 使得该方法对均匀光照和表情变化等具有良好的鲁棒性, 实现了在视频这种低信噪比环境条件下, 快速有效的视频人脸纹理表示与识别.

表 3　不同算法在 Honda/UCSD 视频人脸数据库中的首选识别率

Table 3　Recognition rate of different algorithms on Honda/UCSD database

Type	Algorithm	Accuracy
1	$\mathrm{TPLBP}_{2,8,3,5}$ + EMDD	76.9
1	$\mathrm{LBP}^{u2}_{8,1}$ + EMDD	74.4
1	$\mathrm{LBP}^{u2}_{8,1}$ + DD	71.8
2	$\mathrm{LBP}^{u2}_{8,1}$ + SVM[24]	61.5
3	GLBP-$\mathrm{TOP}^{u2}_{8,8,8,1,1,1}$ + 1NN[25]	64.1
3	$\mathrm{VLBP}_{1,4,1}$ + 1NN[26]	38.5

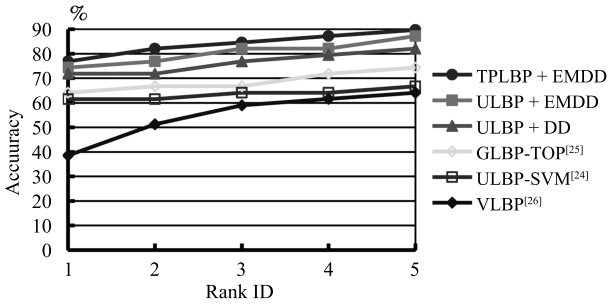


图 5　不同算法在 Honda/UCSD 上的 CMC 曲线

Fig. 5　The CMC curves of different algorithms on Honda/UCSD database

本文提出的算法框架中, 视频包中的示例数目, 即样本数目对识别率的影响较大, 本文进行了示例数目对算法性能影响的实验分析. 选择算法性能较好的 TPLBP + EMDD、GLBP-TOP 等四种算法进行对比分析, 分析结果如图 6 所示, 从实验结果可以看出, 随着示例数, 即样本数目的增大, 样本空间中的特征分布更加细致、准确, 使得算法性能得到一定程度的提高.

3.2　YouTube Face 数据库中的相关实验

YouTube Face 视频人脸数据库[27] 包含 1 595 个不同类别的 3 425 段视频. 平均每个类别可拥有 2.15 段视频. 这些视频中最短的视频持续时间是 48 帧, 最长的是 6 070 帧, 一个视频片段的平均长度是 181.3 帧. 该数据库中的视频背景复杂、姿态多变,

同时包括各种障碍的遮挡, 这些因素对视频人脸的有效识别产生极大影响.

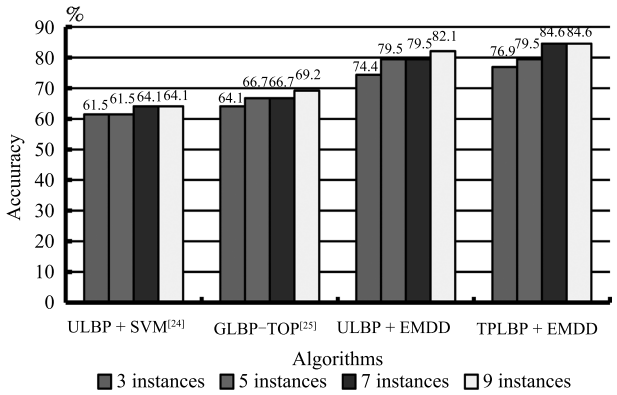


图 6 示例数目对算法性能的影响

Fig. 6 Effect of instance number on algorithm performance

文献 [27] 指出在该数据库中一般采用十字交叉验证的实验方式和一些基准算法, 为了便于比较, 将提出的算法直接与这些基准算法进行识别率上的对比. 不同基准算法的具体说明如下:

1) 视频序列的直接比较. 该类方法用向量集合表示一个视频序列, 其中每个向量是将每个视频帧特征提取后获得的特征向量, 通过计算所有向量集合之间的最大距离 (Max dist), 最小距离 (Min dist), 距离均值 (Mean dist) 以及距离中值 (Median dist) 以实现分类识别.

2) 基于关键帧的方法. 该类方法通过获得视频序列中的最正面人脸 (Most frontal) 或姿态变化最小的视频帧 (Nearest pose) 作为该视频的代表性帧进行识别.

3) 子空间方法. 该类方法是将每个视频序列看做特征空间中的向量分布, 通过计算子空间的相互关系 ($\|U_1^{\mathrm{T}}U_2\|$) 来达到识别的目的, 其中互子空间方法 (Mutual subspace method, MSM) 是较为经典的一种基于图像序列的人脸识别方法.

4) 本文提出的基于多示例学习的视频人脸识别算法, 具体的实验过程与上述在 Honda/UCSD 中的相似, 对比实验结果如表 4 所示. 从实验结果可以看出, 与其他算法比较, 本文提出的算法在一定程度上提高了视频人脸识别精度, 进一步验证了本文算法的有效性.

4 结论

本文提出了一种基于加权 Fisher 准则的多示例学习视频人脸识别算法, 该算法将每个人脸视频视为一个包, 将视频中的人脸图像作为包中的示例, 对包中的示例图像提取加权的 TPLBP 编码值并建立特征空间直方图来描述人脸得到示例特征, 通过多示例学习算法训练得到分类器以实现对测试人脸视频的分类预测. 本文算法在得到较高的识别精度的同时, 有效解决了目标姿态多变视频环境中的人脸视频关键帧难以定位的问题, 并且具有较强的抗干扰能力, 对均匀光照和姿态变化等也具有较好的鲁棒性. 但该算法时间复杂度较高, 学习算法的泛化能力还有待进一步加强.

表 4 不同算法在 YouTube Face 视频人脸数据库中的首选识别率 (%)

Table 4 Recognition rate of different algorithms on YouTube Face database (%)

类型	算法	TPLBP	LBP
1	min dist	71.53	70.66
1	max dist	62.1	61.06
1	mean dist	69.68	68.34
1	median dist	69.86	68.16
2	most frontal	68.54	66.5
2	nearest pose	67.53	66.87
3	MSM	68.34	66.19
3	$\|U_1^TU_2\|$	71.31	69.78
4	Proposed	75.28	73.43

References

1 Barr J R, Bowyer K W, Flynn P J, Biswas S. Face recognition from video: a review. *International Journal of Pattern Recognition and Artificial Intelligence*, 2012, **26**(5): Article No. 1266002

2 Belhumer P N, Hespanha J P, Kriegman D J. Eigenfaces vs fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 711−720

3 Wang R P, Shan S G, Chen X L, Gao W. Manifold-manifold distance with application to face recognition based on image set. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, AK, USA: IEEE, 2008. 1−8

4 Huang Z W, Wang R P, Shan S G, Chen X L. Projection metric learning on Grassmann manifold with application to video based face recognition. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015. 140−149

5 Harandi M T, Sanderson C, Shirazi S, Lovell B C. Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, CO, USA: IEEE, 2011. 2705−2712

6 Yang M, Zhu P F, van Gool L, Zhang L. Face recognition based on regularized nearest points between image sets. In: Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). Shanghai, China: IEEE, 2013. 1−7

7 Hu Y Q, Mian A S, Owens R. Sparse approximated nearest points for image set classification. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, CO, USA, USA: IEEE, 2011. 121−128

8 Wang R P, Guo H M, Davis L S, Dai Q H. Covariance discriminative learning: a natural and efficient approach to image set classification. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI, USA: IEEE, 2012. 2496−2503

9 Cevikalp H, Triggs B. Face recognition based on image sets. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 2010. 2567−2573

10 Hu Y Q, Mian A S, Owens R. Face recogniton using sparse approximated nearest points between image sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(10): 1992−2004

11 Yu Qian, Gao Yang, Huo Jing, Zhuang Yun-Kai. Discriminative joint multi-manifold analysis for video-based face recognition. *Journal of Software*, 2015, **26**(11): 2897−2911 (于谦, 高阳, 霍静, 庄韫恺. 视频人脸识别中判别性联合多流形分析. 软件学报, 2015, **26**(11): 2897−2911)

12 Zhao G Y, Ahonen T, Matas J, Pietikainen M. Rotation-invariant image and video description with local binary pattern features. *IEEE Transactions on Image Processing*, 2012, **21**(4): 1465−1477

13 Wang W, Wang R P, Huang Z W, Chen X L. Discriminant analysis on Riemannian manifold of Gaussian distributions for face recognition with image sets. *IEEE Transactions on Image Processing*, 2018, **21**(1): 151−163

14 Dietterich T G, Lathrop R H, Lozano P T. Solving the multiple instance problem with axis parallel rectangles. *Artificial Intelligence*, 1997, **89**(1−2): 31−71

15 Lai K T, Yu F X, Chen M S, Chang S F. Video event detection by inferring temporal instance labels. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014. 2251−2258

16 Ding Xin-Miao, Li Bing, Hu Wei-Ming, Guo Wen, Wang Zhen-Chong. Horror video scene recognition based on multi-view joint sparse coding. *Acta Electronica Sinica*, 2014, **42**(2): 301−305 (丁昕苗, 李兵, 胡卫明, 郭文, 王振翀. 基于多视角融合稀疏表示的恐怖视频识别. 电子学报, 2014, **42**(2): 301−305)

17 Yang J, Yan R, Hauptmann A G. Multiple instance learning for labeling faces in broadcasting news video. In: Proceedings of the 13th ACM International Conference on Multimedia. Hilton, Singapore: ACM, 2005. 31−40

18 Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(7): 971−987

19 Wolf L, Hassner T, Taigman Y. Descriptor based methods in the wild. In: Proceedings of the 2008 Workshop on Faces in Real Life Images Detection Alignment and Recognition. Marseille, France, 2008.

20 Maron O, Lozano-Pérez T. A framework for multiple-instance learning. In: Proceedings of the 10th International Conference on Neural Information Processing Systems. Cambridge, MA, USA: MIT Press, 1998. 570−576

21 Zhang Q, Goldman S A. EM-DD: an improved multiple-instance learning technique. In: Proceedings of the 14th International Conference on Neural Information Processing Systems. Cambridge, MA, USA: MIT Press, 2002. 1073−1080

22 Moghaddam B, Jebara T, Pentland A. Bayesian face recognition. *Pattern Recognition*, 2000, **33**(11): 1771−1782

23 Lee K C, Ho J, Yang M H, Kriegman D. Video-based face recognition using probabilistic appearance manifolds. In: Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Madison, WI, USA: IEEE, 2003. 313−320

24 Chang C C, Lin C J. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2011, **2**(3): Article No. 27

25 Wang Y, Shen X J, Chen H P, Zhai Y J. Dynamic biometric identification from multiple views using the GLBP-TOP method. *Bio-Medical Materials and Engineering*, 2014, **24**(6): 2715−2724

26 Zhao G Y, Pietikainen M. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(6): 915−928

27 Wolf L, Hassner T, Maoz I. Face recognition in unconstrained videos with matched background similarity. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, CO, USA, USA: IEEE, 2011. 529−534

王 玉 吉林大学应用技术学院副教授. 2017 年获得吉林大学计算机科学与技术学院博士学位. 主要研究方向为图像处理与机器学习.
E-mail: wangyu001@jlu.edu.cn
(**WANG Yu** Associate professor at the College of Applied Technology, Jilin University. He received his Ph. D. degree from the College of Computer Science and Technology, Jilin University in 2017. His research interest covers image processing and machine learning.)

申铉京 吉林大学计算机科学与技术学院教授. 1990 年获得哈尔滨工业大学博士学位. 主要研究方向为多媒体技术, 计算机图像处理, 智能测量系统, 光电混合系统. E-mail: xjshen@jlu.edu.cn
(**SHEN Xuan-Jing** Professor at the College of Computer Science and Technology, Jilin University. He received his Ph. D. degree from Harbin Institute of Technology in 1990. His research interest covers multimedia technology, computer image processing, intelligent measurement system, and optical-electronic hybrid system.)

陈海鹏 吉林大学计算机科学与技术学院教授. 主要研究方向为图像处理与模式识别. 本文通信作者.
E-mail: chenhp@jlu.edu.cn
(**CHEN Hai-Peng** Professor at the College of Computer Science and Technology, Jilin University. His research interest covers image processing and pattern recognition. Corresponding author of this paper.)