

# 一种基于 QPSO 优化的流形学习的视频人脸识别算法

刘宇琦<sup>1,2</sup> 赵宏伟<sup>1,2</sup> 王玉<sup>1,2,3</sup>

**摘要** 视频场景复杂多变, 视频采集设备不一致等原因, 导致无约束视频中充斥着大量的遮挡和人脸旋转, 视频人脸识别方法的准确率不高且性能不稳定. 为解决上述问题, 本文提出了一种基于 QPSO 优化的流形学习的视频人脸识别算法. 该算法将视频人脸识别视为图像集相似度度量问题, 首先帧图像对齐后提取纹理特征并进行融合, 再利用带有 QPSO 优化的黎曼流形大幅度简约维度以获得视频人脸的内在表示, 相似度则由凸包距离表示, 最后利用 SVM 分类器获得分类结果. 通过在 Youtube Face 数据库和 Honda/UCSD 数据库上与当前主流算法进行的对比实验, 验证了本文算法的有效性, 所提算法识别精度较高, 误差较低, 并且对光照和表情变化具有较强的鲁棒性.

**关键词** 视频人脸识别, 量子微粒群优化, 黎曼流形学习, 视频相似度

**引用格式** 刘宇琦, 赵宏伟, 王玉. 一种基于 QPSO 优化的流形学习的视频人脸识别算法. 自动化学报, 2020, 46(2): 256–263

**DOI** 10.16383/j.aas.c180359



开放科学 (资源服务) 标识码 (OSID):

## Video Face Recognition Method Based on QPSO and Manifold Learning

LIU Yu-Qi<sup>1,2</sup> ZHAO Hong-Wei<sup>1,2</sup> WANG Yu<sup>1,2,3</sup>

**Abstract** The highly complex video scene and the inconsistent video acquisition equipment have made the unconstrained videos full of occlusion and face rotation, thereby, resulting in both low accuracy and unstable performance of video face recognition. To solve the problem, we propose a novel method by integrating the quantum behaved particle swarm optimization (QPSO) and the Riemannian manifold learning. It outperforms the existing state-of-art methods owing to the followed contributions: 1) the algorithm treats each face video as an image set, so that the texture features can be extracted from the aligned frame image; 2) the internal representation of video face is obtained by the QPSO Riemannian manifold, enabling the similarity measurement using the distance between convex hulls; 3) the classification is conducted using the common-practiced SVM classifier, to some extent, guaranteeing the good prediction performance. The experiments on both the YouTube Face database and the Honda/UCSD database have shown that the proposed algorithm is not only of higher accuracy, but also more robust to the illumination and expression changes, as compared to the other methods.

**Key words** Video-based face recognition, quantum-behaved particle swarm optimization, Riemannian manifold learning, video similarity

**Citation** Liu Yu-Qi, Zhao Hong-Wei, Wang Yu. Video face recognition method based on QPSO and manifold learning. *Acta Automatica Sinica*, 2020, 46(2): 256–263

收稿日期 2018-05-29 录用日期 2018-10-06  
Manuscript received May 29, 2018; accepted October 6, 2018  
国家青年科学基金 (61101155), 吉林省优秀青年人才基金项目 (20180520020JH), 吉林省科技计划重点科技研发项目 (20180201064SF), 国家重点科技研发计划项目 (2018YFC0830103) 资助

Supported by the National Science Foundation for Young Scientists of China (61101155), Outstanding Young Talent Foundation of Jilin Province (20180520020JH), Key Projects of Jilin Province Science and Technology Development Plan (20180201064SF), National Key and Development Plan (2018YFC0830103)

本文责任编辑 左旺孟

Recommended by Associate Editor ZUO Wang-Meng

1. 吉林大学计算机科学与技术学院 长春 130012 2. 吉林大学符号计算与知识工程教育部重点实验室 长春 130012 3. 吉林大学应用科技学院 长春 130012

1. College of Computer Science and Technology, Jilin University, Changchun 130012 2. Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education,

自图像人脸识别技术兴起以来, 尤其是近些年多媒体技术的发展使得图像采集成本的大规模下降, 人脸识别一直是图像识别领域的重要研究方向. 中央监控设备的大规模普及, 如雨后春笋般的社交媒体更让基于视频人脸识别的角色、主题等高层特征的分析与研究有更大的普适性. 所以, 在视频采集成本如此低廉的今天, 视频环境下的人脸识别相比于基于图像的人脸识别具有更大的研究价值.

目前, 人脸识别的方法可以分为全局方法和局部方法两大类<sup>[1]</sup>. 全局统计方法通常基于经典的统计分析技术, 忽略脸部的某些特征点, 人脸图像以矩阵的形式作为整体进行处理, 并且通常被转换成更容易处理的特征向量. 全局统计方法的主要优点是

Jilin University, Changchun 130012 3. Applied Technology College, Jilin University, Changchun 130012

实现简单、可移植性好、复杂度较低, 但该类方法对照明、位置和面部表情变化非常敏感. 局部识别方法首先检测兴趣点, 然后提取位于这些兴趣点上的特征. 这类方法多数基于特定几何结构的提取, 例如额头的宽度、鼻子的尺寸等, 最后, 将该数据用于训练得到分类器从而用于识别不同的类别.

为了更好地处理无约束数据来源的背景、光照、视角的变化, Wolf 等<sup>[2]</sup> 提出利用背景样本集的思想, 利用两个视频及其在背景集中的近邻点分别训练两个 SVM (Support vector machine) 分类器, 然后再用这两个分类器与另一个视频作比较, 将得到的两个置信度取均值作为最终的相似度度量. 流形距离也是用来度量相似度的重要手段, 于谦等<sup>[3]</sup> 利用两个流形来描述一组图像集, 类间流形描述了每个图像集的平均信息, 称之为“平均脸”, 类内流形表示每个图像集的所有原始图像的信息, 并采取分片技术学习两种流形的投影矩阵, 以期学习到更具判别性的高层特征. Wang 等<sup>[4]</sup> 利用由多个带有先验概率的高斯分量组成的高斯混合模型来描述一个图像集, 希望从不同的类别中提取出不同的高斯分量用于相似度度量, 提出了一个用高斯分量的先验概率作为权值的核判别分析方法 (Kernel discriminant analysis). 於俊等<sup>[5]</sup> 提出了一种在粒子滤波框架下的结合在线外观模型和柱状人头模型人脸三维运动跟踪方案, 相关主观实验表明: 由跟踪得到的人脸运动参数合成的虚拟人脸具有较高的辨识度.

基于图像集合的方法是视频人脸识别领域中的常见方法, 而基于图像集合方法通常包括两个方面<sup>[16]</sup>, 一是如何对人脸图像集合进行建模; 另外就是如何度量模型之间的相似度. 现有工作通常是针对其中的某一方面进行研究与改进, 而本文在 MBGS (Matched background similarity) 等视频相似度方法的启发下, 提出一种基于量子微粒群优化 (Quantum-behaved particle swarm optimization, QPSO) 的黎曼流形学习 (Riemannian

manifold learning, RML) 的视频相似度度量方法并用以解决无约束环境下的视频人脸识别问题. 该算法首先对视频图像集进行特征提取并融合, 再利用黎曼流形学习减少因多特征融合带来的超长维度, 通过低维特征映射到仿射子空间来进行相似度度量, 最后使用 SVM 实现视频人脸的识别与分类.

## 1 本文工作

设训练集中共包含  $C$  个图像集合, 每个图像集合所属的类别不同. 图像集合中的所有人脸图像样本被表示为  $d$  维特征向量  $x_{ci} \in \mathbf{R}^d$ , 其中  $c = 1, 2, \dots, C$ . 另外,  $i = 1, 2, \dots, n_c$  用于表示第  $c$  个图像集合包含  $n_c$  个样本.

### 1.1 算法描述

本文提出的基于量子微粒群优化的黎曼流形学习视频人脸识别算法属于一种基于图像集合方式的人脸识别算法. 该算法将输入视频以图像集合的形式进行表示, 对图像集合中的每幅图像分别提取 LBP (Local binary pattern)、CSLBP (Center-symmetric LBP) 以及 FPLBP (Four-patch LBP) 特征, 对提取出三种特征进行级联融合以获得人脸图像高维表示, 所得到的融合特征能够有效表示人脸纹理, 并且对人脸表情及光照变化较为鲁棒. 然后, 本文对融合特征空间进行黎曼流形学习以减少因多特征融合带来的维度冗余, 从而获得样本数据内嵌的流形结构. 而在流形学习过程中, 需要确定最优的参数组合, 这里, 采用了量子微粒群优化算法以在样本集中寻优得到最优的参数组合, 从而获得低维特征的黎曼流形表示. 最后, 将低维特征表示映射到仿射子空间<sup>[16]</sup>, 并通过凸包来对图像集合进行建模, 并利用凸包之间的距离作为视频对之间的相似度, 将不同视频对的相似度信息送入 SVM 进行训练已得到最终的分分类器并实现对测试视频对的确认与分类, 本文提出的整体算法框架如图 1 所示.

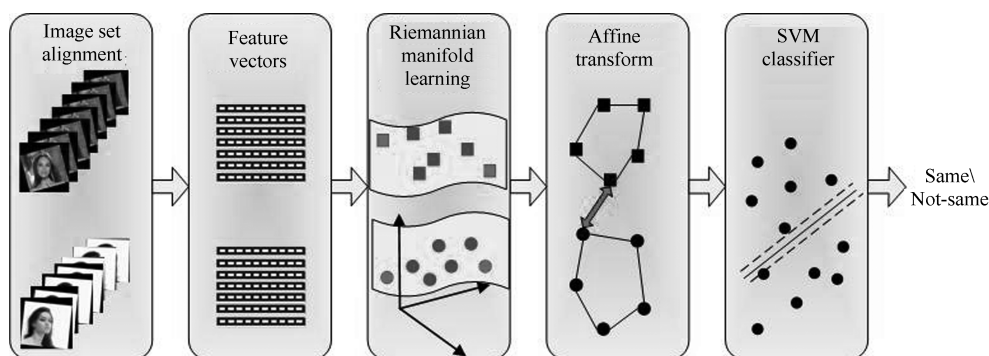


图 1 基于 QPSO 优化黎曼流形的视频人脸识别算法框架

Fig. 1 The framework of proposed video face recognition algorithm based on manifold learning

本文提出的基于 QPSO 优化黎曼流形的视频人脸识别算法主要包括以下步骤:

1) 图像对准和特征提取. 对给定的两个图像集进行图像对准, 然后提取图像集的 LBP、CSLBP、FPLBP 三种特征并进行级联操作以获得融合的高维特征;

2) 带有 QPSO 优化的 RML 降维. 将数据库训练集中的 RML 降维结果中未可达点数量作为适应度, 并利用量子微粒群优化策略在训练集中寻优以得到 RML 最优参数组. 利用寻优后的 RML 对输入图像集的高维特征进行降维处理得到各图像集的低维内在特征, 从而, 同一图像集内所含图像的低维特征向量组成了该图像集的特征表示;

3) 仿射包相似度度量. 这里, 将每个图像集的特征表示映射到仿射空间, 并利用凸包模型来对每个图像集进行建模, 通过两个凸包之间的距离来度量这两个图像集之间的相似度;

4) 分类与识别. 计算数据库训练集中的所有视频对的相似度信息送入线性 SVM 分类器进行训练, 得到最终分类器. 对于待检测视频对, 以同样方式可以计算得到一个相似度信息, 送入最终分类器都可以确定该视频对是否属于同一目标, 从而得到该视频匹配与否的识别结果.

## 1.2 本文所提方法的贡献

1) 采取多特征融合后再降维的方法最大限度地利用了纹理特征信息. 另外, 相对于其他线性降维算法, 流形学习能够更好地保持局部信息的同时保持图像流形的内在结构<sup>[6]</sup>.

2) 利用 QPSO 对黎曼流形的参数选择进行了进一步优化, 且考虑到数据库内部的巨大的个体差异 (即样本数), 利用各视频帧数为参数进行调整, 以达到为存在巨大个体差异的群体找到最佳参数组的目的.

## 2 基于黎曼流形学习的视频人脸识别方法

图像的特征可以以简单的高维向量进行表示, 这里共使用了三种描述符, 局部二值模式 (Local binary patterns, LBP)、中心对称局部二值模式 (CS-LBP) 及四分块局部二值模式 (Four-patch LBP).

### 2.1 特征提取

基于纹理的特征提取方法在图像识别尤其是人脸识别领域应用广泛, 其中 LBP 是纹理描述的经典算子<sup>[7]</sup>. LBP 是一种描述图像单一像素属性 (灰度或彩度图像的不同通道) 局部大小关系的二进制表述方法, 该算子具有计算简单、尺度不变、旋转不变性等优点. CSLBP (Center-symmetric local bi-

nary patterns)<sup>[8]</sup> 是利用对角位置的两个像素进行编码的 LBP 改进算子, 在行人检测问题中, CSLBP 编码的纹理特征信息显著, 被广泛应用. 对于第  $i$  个图像集的第  $j$  个帧的第  $k$  个像素处的 CSLBP 的编码定义如下所示:

$$CS_{i,j,k} = \sum_{i=0}^{(N-2)/2} g(p_{i,j,k}, p_{i,j,[k+(m_{i,j}/2)]}) \times 2^k$$

$$g(p_{i,j,k_1}, p_{i,j,k_2}) = \begin{cases} 1, & |p_{i,j,k_1} - p_{i,j,k_2}| > t_{CS} \\ 0, & \text{否则} \end{cases} \quad (1)$$

其中, 阈值  $t_{CS}$  通常取较小的值.

Four-patch LBP (FPLBP) 是由 Wolf 等<sup>[9]</sup> 提出的基于分块的 LBP 纹理描述算子. 该算子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. FPLBP 算子的编码定义为:

$$FPLBP_{r_1, r_2, S, \omega, \alpha}(p) = \sum_i^{S/2} f(d(P_{1,i}, P_{2,(i+\alpha) \bmod S}) - d(P_{1,i+S/2}, P_{2,(i+S/2+\alpha) \bmod S})) \times 2^i$$

$$f(x) = \begin{cases} 1, & \text{若 } x > t_{FP} \\ 0, & \text{否则} \end{cases} \quad (2)$$

其中,  $r_1, r_2, S, \alpha, \omega$  代表 FPLBP 的选块策略,  $P_{1,i}$  代表第 1 圈的第  $i$  个分块.  $d(\cdot)$  表示两个分块之间的某种距离.

### 2.2 带有 QPSO 的黎曼流形学习

黎曼流形学习是由 Lin 等<sup>[10]</sup> 于 2008 年提出的, 它基于如下假设: 高维输入样本集存在本征维度, 且样本集分布在这样维度的黎曼流形上. 于是, 在这个低维黎曼流形上的再表示就是黎曼流形降维的主要思想<sup>[11]</sup>.

黎曼流形的参数主要包括三类: 1) 基准点  $p$ , 决定着低维坐标系的位置, 选择适合的基准点能够大幅地降低计算量; 2) 目标维度  $d$ ; 3) 近邻区域的选择, 其中,  $k$  代表基准点附近的近邻个数,  $m$  代表其他样本点的近邻个数,  $r$  代表近邻最大半径.

对于基准点  $p$ , 本文选择采用流形中心. 计算样本点间的欧氏距离并构建无向图及距离矩阵  $D$ , 求得多源最短路径<sup>[12]</sup>. 每一个样本点到其他样本点的最短路径的最长距离定义为该点的几何半径, 将几何半径最小的点作为流形中心.

本文对目标维度和近邻区域尺寸的选择采用了 QPSO 参数寻优. PSO 算法基于生物群体模

型<sup>[13]</sup>, 在标准 PSO 算法中, 粒子的速度总是有限的, 导致算法不能确保以概率 1 收敛到全局最优解. 而 QPSO 算法能保证全局收敛<sup>[14]</sup>, 并具有控制参数少、进化过程简单、收敛速度快和运算简单等优点<sup>[15]</sup>.

这里, QPSO 算法中的种群规模  $ps$  通常选择 20, 第  $i$  个微粒在第  $t$  代的位置  $P_i^t$  代表一组可能的参数,  $pbest_i^t$  代表第  $i$  个微粒在前  $t$  代的历史最佳位置,  $gbest^t$  代表前  $t$  代的全局历史最佳位置, 而每次参数迭代之后都会根据所有微粒的适应度分布进行如下运动更替位置, 又称为种群进化:

$$mbest^{t+1} = \left( \frac{1}{ps} \right) \sum_{i=1}^{ps} pbest_i^t \quad (3)$$

$$PP_i^{t+1} = \delta pbest_i^t + (1 - \delta) gbest^t \quad (4)$$

若设  $T = |mbest^{t+1} - P_i^t|$ , 则有:

$$p_i^{t+1} = \begin{cases} PP_i^{t+1} + T\alpha(t)\ln(\frac{1}{\mu}), & \mu \geq 0.5 \\ PP_i^{t+1} - T\alpha(t)\ln(\frac{1}{\mu}), & \mu < 0.5 \end{cases} \quad (5)$$

式中,  $ps$  表示种群规模,  $\alpha(t)$  表示收缩膨胀系数, 关系到收敛速度,  $mbest^{t+1}$  表示在前  $t$  代种群中历史最佳位置的平均值,  $\delta, \mu \in [0, 1]$  是服从均匀分布的随机数.  $PP_i^{t+1}$  用于表示前  $t$  代中第  $i$  个粒子的历史最佳位置  $pbest_i$  和前  $t$  代全局历史最佳位置  $gbest^t$  之间的随机点. 其中,  $\alpha(t)$  通常取值:

$$\alpha(t) = m - \frac{t(m-n)}{MaxIter} \quad (6)$$

其中,  $MaxIter$  为最大允许迭代次数,  $m, n$  为常数, 这里, 选用  $m = 1, n = 0.5$  的经验值. QPSO 优化黎曼流形参数过程伪代码如算法 1 所示.

**算法 1. 黎曼流形参数 QPSO 优化过程伪代码**

**Input:**  $X = X_1, X_2, \dots, X_c$

**output:**  $gbest = \{p, d, k, m, r\}$

- 1) Initialize  $ps = 20; MaxIter = 100; f_{gbest} = 1 \times 10^7$
- 2) **for**  $i = 1 : ps$
- 3) set  $p_i^0 = \{p_i, d_i, k_i, m_i, r_i\}$
- 4) RML  $X$  by  $P_i^0$  to get  $X'$  and  $F_i$
- 5)  $pbest_i = P_i^0$
- 6) **if**  $f_{pbest_i} < f_{gbest}$
- 7)  $f_{gbest} = f_{pbest_i}$
- 8)  $gbest = pbest_i$
- 9) **endif**
- 10) **endfor**

- 11) **for**  $t = 1 : MaxIter$
- 12) compute  $mbest^t$  by (3)
- 13) **for**  $i = 1 : ps$
- 14) update  $P_i^t$  by (4) ~ (6)
- 15) RML  $X$  by  $P_i^t$  to get  $X'$  and  $F_i$
- 16) **if**  $F_i < f_{pbest_i}$
- 17)  $f_{pbest_i} = F_i$
- 18)  $pbest_i = P_i^t$
- 19) **if**  $f_{pbest_i} < f_{gbest}$
- 20)  $f_{gbest} = f_{pbest_i}$
- 21)  $gbest = pbest_i$
- 22) **endif**
- 23) **endif**
- 24) **endifor**
- 25) **if**  $|mbest^t - mbest^{t-1}| < 1 \times 10^{-7}$  and  $|mbest^{t-1} - mbest^{t-2}| < 1 \times 10^{-7}$
- 26) **return**  $gbest$
- 27) **endif**
- 28) **endifor**

### 2.3 相似度量

对于包含  $n_c$  个图像的图像集合  $X_c$  来说, 经过上述的黎曼流形优化选择过程后, 会得到一个低维样本重新构成的图像集合  $X'_c$ , 其凸包形式  $CH(X'_c)$  为:

$$CH(X'_c) = \left\{ \sum_{i=1}^{n_c} X'_{ci} a_{ci} \mid \sum_{i=1}^{n_c} a_{ci} = 1, 0 \leq a_{ci} \leq 1 \right\} \quad (7)$$

若令  $a_c = [a_{c1}, a_{c1}, \dots, a_{cn_c}]^T$ ,  $X'_c = [X'_{c1}, X'_{c1}, \dots, X'_{cn_c}]^T$ , 上式可改写为:

$$CH(X'_c) = \left\{ X'_c a_c \mid \sum_{i=1}^{n_c} a_{ci} = 1, 0 \leq a_{ci} \leq 1 \right\} \quad (8)$$

对于给定的两个图像集合  $X_i$  和  $X_j$ , 他们之间的相似度距离可通过解决下面的约束凸优化问题获得:

$$\begin{aligned} (\hat{a}_i, \hat{a}_j) &= \arg \min_{a_i, a_j} \|X_i a_i - X_j a_j\|^2 \\ \text{s.t.} \quad &\sum_k a_{ik} = 1, \sum_{k'} a_{jk'} = 1 \\ &0 \leq a_{ik}, a_{jk'} \leq 1 \end{aligned} \quad (9)$$

若令  $X = (X_i - X_j)$ ,  $a = [a_i \ a_j]^T$ , 上式可改写为:

$$(\hat{a}_i, \hat{a}_j) = \arg \min_{a_i, a_j} \|Xa\|^2 \quad (10)$$

凸包作为一种约束型的仿射包,式(10)可通过对仿射包之间的欧氏距离类似方式进行求解,即上述问题可以转化为一个标准的最小二乘问题<sup>[16]</sup>.最后,  $X_i$  与  $X_j$  之间的距离可以表示为:

$$D(CH(X_i), CH(X_j)) = \|X'_i \hat{a}_i - X'_j \hat{a}_j\| \quad (11)$$

### 3 实验与分析

#### 3.1 YouTube Face 数据库中的相关实验

本文实验中用到的数据库是 YouTube Face 数据库,该数据库<sup>[2]</sup>由 Wolf 等于 2011 年创建,该数据库共包含 3 425 个视频段,累计包含 1 595 个不同类别,人均拥有 2.15 个视频.该数据库中的视频来源广泛,场景多变、姿态各异,同时包括各种遮挡的情形,这些因素的存在都对视频人脸的有效识别产生极大影响,也能更好地模拟无约束条件,从而测试算法在真实环境中的稳定性和准确度.该测试用数据库共包含 10 000 个视频段,其中,视频段最长包含 2 157 帧,最短包含 48 帧,平均长度为 184.2 帧.

在数据库中选取 5 000 对视频,并确保其中 2 500 对视频属于同一目标,而另外 2 500 对视频属于不同目标,并将这 5 000 对视频分为 10 个子集以采用交叉验证的实验方式.为了更好量化和对比算法的性能,与参考文献 [2] 相同,本文选用 4 个评价指标,分别是准确率 Acc、标准误差 SE (Standard error)、ROC (Receiver operating characteristic) 线下面积 AUC (Area under curve) 和等概率误差 EER (Equal error rate),其中,ROC 曲线是二类分类器常用的评价指标.

##### 3.1.1 算法性能与参数分析

文献 [2] 中已经给出在 YouTube Faces 数据库上进行算法分析的具体过程,为了便于比较,这里将本文算法与最简相似度度量方法进行比较,即将向量组的欧氏距离视为相似度.将每个视频段的纹理特征矩阵视为特征向量组,为视频对中的两个特征向量组求解一一对应的欧氏距离,并分别使用最大距离 (max dist)、最小距离 (min dist)、距离均值 (mean dist) 以及距离中值 (median dist) 来度量相

似度.另外,不以图像集顺序一一对应,而是在另一向量组中与本向量距离最近的距离值作为最小平均距离 (mean min),从而得到五种度量方法及结果.

为了获取更全面的纹理信息,本文算法采用了多特征融合来描述视频中的所有帧图像.文献 [2] 中给出了三种特征选取方法 LBP、CSLBP、FPLBP 的测试结果,为了验证多特征融合的优势,融合对比试验结果如表 1 所示.

从表 1 中可以明确看出,融合特征在 min dist 方法下获得最佳的实验结果,有效地验证了多特征融合较单一纹理特征的优势,但该方法使得特征维度增大从而带来时间复杂度较高的问题,另外,多种纹理特征的融合必定会带来高冗余,可以通过带有 QPSO 参数优化的黎曼流形学习为其进行“维度简约”.

QPSO 的寻优过程中,近邻尺寸和基准点的选择又明显依赖于各视频图像集的样本数,这里将参数值以参数系数来表示,而参数系数又选择以帧数倍数来表示.另外,近邻半径  $r$  用的是最大欧氏距离的倍数,以此达到为存在巨大个体差异的群体找到最佳参数组的目的.

收敛过程中各参数系数的变化如图 2 所示.迭代过程中各参数组使用在部分测试集内的黎曼流形降维带来的均不可达点数量作为寻优适应度,

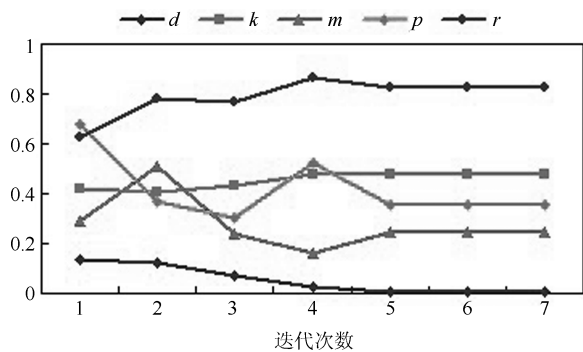


图 2 QPSO 优化过程各参数系数变化

Fig. 2 Optimization of parameters in QPSO optimization process

表 1 不同纹理描述算子在 YouTube Face 数据库上的识别率 (%)

Table 1 Recognition rate of different texture description operators on YouTube Face database (%)

Method	LBP			CSLBP			LBP			CSLBP		
	Acc ± SE	AUC	EER	Acc ± SE	AUC	EER	Acc ± SE	AUC	EER	Acc ± SE	AUC	EER
min dist	65.7 ± 1.6	70.66	35.20	63.08 ± 1.0	67.29	37.36	65.60 ± 1.7	70.01	35.64	66.04 ± 2.24	71.21	34.88
max dist	57.90 ± 1.7	61.06	42.64	56.46 ± 2.2	58.80	43.76	55.70 ± 2.4	58.10	45.32	57.44 ± 2.21	59.91	43.20
mean dist	63.72 ± 2.2	68.34	36.84	61.10 ± 2.1	64.86	39.52	62.86 ± 1.4	66.98	38.20	63.88 ± 2.18	67.88	37.20
median dist	63.46 ± 2.0	68.16	36.80	60.84 ± 2.1	64.81	39.44	62.70 ± 1.5	66.81	38.36	63.50 ± 2.33	67.70	37.52
mean min	65.12 ± 1.7	69.99	35.84	62.62 ± 1.5	66.48	38.28	65.48 ± 1.8	69.22	36.56	65.48 ± 2.15	70.04	35.96

寻优适应度变化趋势如图 3 所示. 从图中可以看出, 第 5 次迭代可达到最优参数, 其适应度均不可达数量低至 0.5585, 并且在第 7 次完成收敛. 此时, 全局最优的参数系数为  $\{d = 0.006, k = 0.4805, m = 0.2452, p = 0.3560, r = 0.8257\}$ .

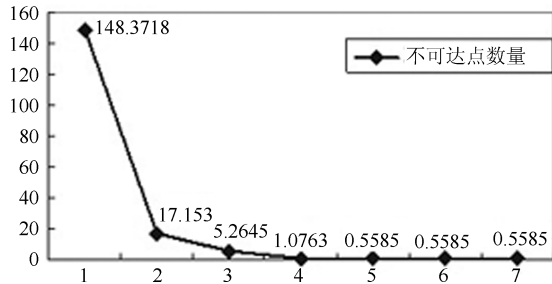


图 3 QPSO 优化过程中寻优适应度变化

Fig. 3 Optimization of fitness in QPSO optimization process

### 3.1.2 算法对比实验与分析

为了验证 RML 与线性降维方法的差异, 并验证他们在视频人脸识别问题上的优势, 本文将 RML 与 MDS (Multi dimensional scaling)<sup>[17]</sup> 在同一数据库上进行了对比, 结果如表 2 所示.

表 2 不同算法在 YouTube Face 数据库上的识别率 (%)

Table 2 Recognition rate of different algorithm on YouTube Face database (%)

Method	With Logmap			With MDS		
	Acc ± SE	AUC	EER	Acc ± SE	AUC	EER
min dist	49.60 ± 0.9	51.24	48.56	51.22 ± 1.4	49.39	50.40
max dist	50.00 ± 0.2	50.71	49.96	50.20 ± 2.0	50.74	49.56
mean dist	50.16 ± 0.6	50.64	49.48	49.64 ± 1.2	50.09	50.68
median dist	50.06 ± 0.6	50.60	49.64	49.18 ± 1.2	50.03	50.60
mean min	50.18 ± 0.7	50.43	49.48	50.16 ± 0.7	49.55	50.68

从表 2 中可以看出, MDS 远不及 RML 在该问题上的表现, 在四种评价指标下均呈现出了较大的差异. 另外, 帧数量对本文算法的影响较大, 在采取 QPSO 进行参数优化的过程中, 本文将参数组的选择由参数的数值改变为对各视频内帧数的倍数, 其中近邻半径  $r$  用的是最大欧氏距离的倍数, 以此达到为存在较大个体差异的群体找到最佳参数组的目的. 本文将选择出的一组结果较好的参数与未降维的多特征融合进行对比. 为了简化实验过程, 相似度度量选择了欧氏距离, 以此来验证降维对实验结果的影响, 表 3 所示的是带有固定参数的实验结果. 从实验结果的 AUC 和 ACC 值上来看, 加入黎曼流形学习后优势并不明显, 但标准误差 SE 值显著下降, 即算法稳定性得到了很大提升, 另外, 加入黎曼流形

学习之后特征维度大幅减少, 可以显著提升算法执行效率, 缩短算法运行时间.

为了衡量本文提出算法的有效性, 这里将本文中提出的算法与不同类型的基准算法进行了识别率上的对比, 对比实验结果如表 4 所示. 不同类型的基准算法说明可参考文献 [2] 和文献 [18].

表 3 黎曼流形在 YouTube Face 数据库上的识别率 (%)

Table 3 Recognition rate of manifold learning on YouTube Face database (%)

Method	With Logmap			Without Logmap		
	Acc ± SE	AUC	EER	Acc ± SE	AUC	EER
min dist	49.60 ± 0.9	51.24	48.56	66.04 ± 2.2	71.21	34.88
max dist	50.00 ± 0.2	50.71	49.96	57.44 ± 2.2	59.91	43.20
mean dist	50.16 ± 0.6	50.64	49.48	63.88 ± 2.1	67.88	37.20
median dist	50.06 ± 0.6	50.60	49.64	63.50 ± 2.3	67.70	37.52
mean min	50.18 ± 0.7	50.43	49.48	65.48 ± 2.1	70.04	35.96

表 4 不同算法在 YouTube Face 视频人脸数据库上的实验结果 (%)

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

Method	CSLBP		FPLBP		LBP		Fusion	
	AUC	EER	AUC	EER	AUC	EER	AUC	EER
min dist	67.29	37.36	70.01	35.64	70.66	35.2	71.21	34.88
max dist	58.8	43.76	58.1	45.32	61.06	42.64	59.91	43.2
mean dist	64.86	39.52	66.98	38.2	68.34	36.84	67.88	37.2
median dist	64.81	39.44	66.81	38.36	68.16	36.8	67.70	37.52
most frontal	63.61	40.36	64.24	40.04	66.5	38.72	66.23	38.4
nearest pose	63.24	40.32	64.35	40.2	66.87	37.88	66.29	38
MSM	64.64	40.04	63.85	40.24	66.19	38.28	66.33	38.28
CMSM	65.17	39.76	68.35	37.16	67.26	38.36	69.81	36.04
$\ U_1^T U_2\ $	67.68	37.4	69.37	35.8	69.78	35.96	70.64	35.32
Linear AHISD	60.06	42.32	60.14	42.28	64.55	39.24	64.71	39.28
Kernel CHISD	66.65	38.6	67.01	38.56	68.89	37.2	68.35	37.4
Proposed	67.52	30.55	74.21	29.55	79.43	28.34	77.35	32.02

通过实验结果可以看出, 本文提出的基于 QPSO 优化黎曼流形的视频人脸识别算法可以有效提高视频人脸识别的精度. 同时, 由于该算法使用融合的局部二值模式来表示图像样本的纹理, 使得该方法对光照、表情变化等具有较好的鲁棒性. 对于不同的视频人脸数据库, 优化后得到的参数不尽相同, 但获得最优参数后, 可以得到较高的识别率, 实现了复杂视频条件下的快速视频人脸识别.

### 3.2 Honda/UCSD 数据库实验与分析

Honda/UCSD 数据库提供了一个用于评估人脸跟踪与识别算法的标准视频数据库. 目前研究发现, 姿态变化是目前人脸识别问题所面临的最大挑

战, 所以该数据库中的所有视频序列均包含明显的 2-D (平面内) 和 3-D (平面外) 头部旋转. 在每一个视频中, 目标可以自由旋转头部位置, 而且通常旋转会持续 15 秒左右, 为该目标提供了大范围的姿势变化. 此外, 这些序列中的一些包含真实世界里人脸识别可能遇到的一些障碍, 如部分遮挡、视野受限、以及大规模变化等. 识别率 (Accuracy) 是衡量视频人脸识别算法性能的重要指标, 本文重点考察了算法在 Honda/UCSD 数据库上的首选识别率.

为便于在该数据库上对本文所提出的算法进行测试, 这里将每个视频序列提取出 7 个关键帧提取并归一化为  $183 \text{ px} \times 229 \text{ px}$ . 算法 1 步骤 1)~3) 的操作与前述一致, 在识别阶段采用最近邻算法, 即依据测试视频凸包与训练集合所有凸包的距离, 将测试视频归为距离最小的凸包类别. 本文算法与 VLBP (Volume LBP)、LBP-TOP (Local binary patterns from three orthogonal planes) 等算子在 Honda/UCSD 数据库上的首选识别率如表 5 所示, 所提算法得到了更高的识别精度, 验证了所提方法的有效性. 同时, 该方法采用 LBP 及其改进算子对视频进行纹理特征提取, 使得所提视频描述算子对光照变化、表情变化等具有良好的鲁棒性.

表 5 不同算法在 Honda/UCSD 视频人脸数据库上的首选识别率

Table 5 Recognition rate of different algorithms on Honda/UCSD database

Algorithm	Accuracy (%)
$GVLBP_{1,4,1} + 1NN$	53.8
$LBP-TOP_{4,4,4,1,1,1} + 1NN$	53.8
$LBP-TOP_{8,8,8,1,1,1} + 1NN$	56.4
$GLBP-TOP_{8,8,8,1,1,1} + 1NN$	66.7
$WGLBP-TOP_{8,8,8,1,1,1} + 1NN$	69.2
The proposed	74.4

为了横向对比本文所提算法与其他基准算法在视频人脸识别问题上的性能, 这里将本文算法与 LBP-TOP<sup>[19]</sup>、Gabor-VLBP、Gabor-LBP-TOP 以及采用了 Fisher 加权准则的 GLBP-TOP (WGLBP-TOP)<sup>[20]</sup> 四种不同算法进行了比较. 当视频帧集合大小为 7 帧时, 五种不同算法的 CMC 曲线如图 4 所示. VLBP 及 LBP-TOP 算子在视频纹理表示方面性能优越, 而本文提出的算法得到了更好的识别结果.

## 4 结论

无约束来源的视频环境不可控, 设备不一致导致视频中充斥着大量的遮挡和人脸旋转、数据量大等问题一直是视频人脸识别领域的研究热点问题.

本文提出的基于 QPSO 优化的流形学习的视频人脸识别算法将视频人脸识别视为图像集相似判别问题, 采取多种纹理特征融合描述符提取图像信息, 随后采用带有 QPSO 优化的黎曼流形学习大幅度简约维度, 再利用凸包距离进行相似度度量, 最后训练得到 SVM 分类器并对测试视频进行分类与识别. 在 Youtube Face 数据库及 Honda/UCSD 数据库将本算法与当前主流算法进行了对比实验, 从而验证了算法的有效性, 本文算法识别精度较高, 误差较低.

目前, 随着深度学习与人工智能方法的发展, 出现了一些用于视频人体行为识别、手势识别以及目标视觉检测的深度学习框架 (如 3D-CNN 等), 为深度学习在视频分析与理解的应用提供了借鉴, 成为视频人脸识别领域的研究方向之一. 另外, 真实视频场景中存在各种遮挡、姿态变化仍然是目前视频人脸识别面临的巨大挑战. 现有数据库虽然对类似影响因素都加以考虑, 但仍然无法达到无约束的视频环境, 如何进一步提高算法在真实视频场景中的通用性和提升算法的鲁棒性都会成为笔者今后的研究重点.

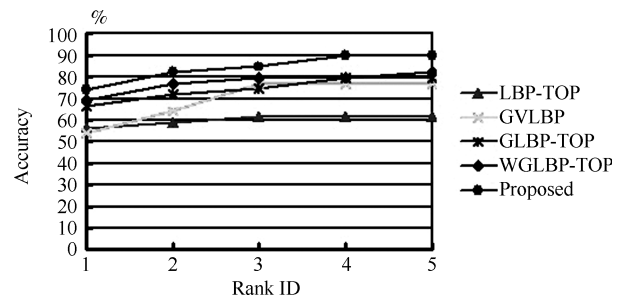


图 4 不同算法在 Honda/UCSD 上的 CMC 曲线

Fig. 4 The CMC curves of different algorithms on Honda/UCSD database

## References

- Jeremiah R B, Kevin W B, Patrick J F, Soma B. Face recognition from video: a review. *International Journal of Pattern Recognition and Artificial Intelligence*, 2012, **26**(5): 1266002-1-1266002-53
- 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: IEEE, 2011. 529-534
- 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, 庄蕴恺. 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)
- 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

- 5 Yu Jun, Wang Zeng-Fu. A robust and high accurate 3D facial motion tracking algorithm. *Journal of Computer Research and Development*, 2014, **51**(4): 802–812  
(於俊, 汪增福. 一种鲁棒高精度的人脸三维运动跟踪算法. *计算机研究与发展*, 2014, **51**(4): 802–812)
- 6 Wong K M, Zhang J K, Liang J P, Jiang H Y. Mean and median of PSD matrices on a riemannian manifold: application to detection of narrow-band sonar signals. *IEEE Transactions on Signal Processing*, 2017, **65**(24): 6536–6550
- 7 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
- 8 Heikkila M, Pietikainen M, Schmid C. Description of interest regions with local binary patterns. *Pattern Recognition*, 2009, **42**(3): 425–436
- 9 Wolf L, Hassner T, Taigman Y. Descriptor Based Methods in the Wild. In: Proceedings of the 2008 European Conference on Computer Vision. Marseille, 2008. 121–128
- 10 Lin T, Zha H B. Riemannian manifold learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(5): 796–809
- 11 Sanin A, Sanderson C, Harandi M T, Lovell B C. Spatio-temporal covariance descriptors for action and gesture recognition. In: Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision. Clearwater Beach, FL, USA: IEEE, 2013. 103–110
- 12 Cabello S, Chambers E W, Erickson J. Multiple source shortest paths in embedded graphs. *SIAM Journal on Computing*, 2012, **42**(4): 1542–1571
- 13 Ueno G, Yasuda K, Iwasaki N. Robust adaptive particle swarm optimization. In: Proceedings of the 2005 IEEE International Conference on Systems, Man and Cybernetics. Waikoloa, HI, USA: IEEE, 2005. 3915–3920
- 14 Zhang C M, Xie Y C, Liu D, Wang L. Fast threshold image segmentation based on 2D fuzzy Fisher and random local optimized QPSO. *IEEE Transactions on Image Processing*, 2017, **26**(3): 1355–1362
- 15 Peng C, Yan J, Duan S K, Zhang S L. Enhancing electronic nose performance based on a novel QPSO-KELM model. *Sensors*, 2016, **16**(4): 520
- 16 Cevikalp H, Triggs B. Large margin classifiers based on convex class models. In: Proceedings of the 2009 International Conference on Computer Vision. Kyoto, Japan, 2009. 101–108
- 17 Tang Z j, Huang Z Q, Zhang X Q, Lao H. Robust image hashing with multidimensional scaling. *Signal Processing*, 2017, **137**(C): 240–250
- 18 Cevikalp H, Triggs B. Face recognition based on imagesets. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 2010. 2567–2573
- 19 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
- 20 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



刘宇琦 吉林大学计算机科学与技术学院博士生. 主要研究方向为图像处理与模式识别. E-mail: liuyuq@jlu.edu.cn  
(LIU Yu-Qi Ph.D. candidate at the College of Computer Science and Technology, Jilin University. His research interest covers image processing and pattern recognition.)



赵宏伟 吉林大学计算机科学与技术学院教授. 2001 年获得吉林大学博士学位. 主要研究方向为多媒体技术, 计算机图像处理, 智能控制与嵌入式系统. E-mail: zhaohw@jlu.edu.cn  
(ZHAO Hong-Wei Professor at the College of Computer Science and Technology, Jilin University. He received his Ph.D. degree from Jilin University in 2001. His research interest covers multimedia technology, computer image processing, intelligent control and embedded system.)



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. 本文通信作者. 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 College of Computer Science and Technology, Jilin University in 2017. His research interest covers image processing and machine learning. Corresponding author of this paper.)