

文章编号 1004-924X(2017)03-0779-07

一种分布式人脸识别方法及性能优化

闵卫东^{1,2*}, 石 杰¹, 韩 清², 王 玮¹

(1. 天津工业大学 计算机科学与软件学院, 天津 300387;

2. 南昌大学 信息工程学院, 江西 南昌 330031)

摘要:传统的集中式人脸识别方法在时间效率和可扩展性等方面存在不足,已经不能满足大规模实时人脸识别的需求。针对这个技术瓶颈,本文提出了一种分布式人脸识别方法,该模型由多个代理和一个服务器组成。代理能够同时对多个视频中的行人进行检测、跟踪以及特征提取,服务器则对视频中的行人执行识别操作。针对代理处理的任务分布不均而导致处理视频时间过长、任务量过大引起的 CPU 利用率激增问题,设计了代理的负载均衡来进行性能优化。利用代理统计处理的视频总数及每个视频中的行人数量,并将统计数据发送给服务器。服务器通过负载均衡将视频重新分配给每个代理进行人脸识别。实验结果证明,分布式人脸识别有效地提高了人脸识别方法的效率和可扩展性。对一些较为极端的实验例子,经过性能优化后,代理中最大的 CPU 利用率降低近 40%,有效地缓解了时间延迟问题。

关键词:人脸识别;分布式;代理;负载均衡;性能优化

中图分类号:TP394.1 **文献标识码:**A **doi:**10.3788/OPE.20172503.0779

A distributed face recognition method and performance optimization

MIN Wei-dong^{1,2*}, SHI Jie¹, HAN Qing², WANG Wei¹

(1. School of Computer Science and Software Engineering,
Tianjin Polytechnic University, Tianjin 300387, China;

2. School of Information Engineering, Nanchang University, Nanchang 330031, China)

* Corresponding author, E-mail: minweidong@ncu.edu.cn

Abstract: Since the scale of video to be monitored and processed by face recognition system has been increasing, the traditional centralized face recognition methods are insufficient in time efficiency and scalability and no longer able to meet the needs of large scale and real-time face recognition. Aimed at such technical bottleneck, a distributed face recognition method was proposed. The model consisted of several agents and one server, in which the agent was able to detect, trace and extract features of pedestrians in several videos at the same time, while the server was able to identify those pedestrians in videos. Subject to the problems caused by uneven distribution of agent processing tasks, including long processing time, too much tasks and CPU usage explosion, a load balance of agent was designed for performance optimization. Firstly, the agent was used to count the total videos to be processed and number of pedestrians in each video, then, statistical data was sent to the server through the agent and at last, the server will re-allocate the videos to each agent for face recognition by the load balance.

收稿日期:2016-12-21;**修订日期:**2017-01-15.

基金项目:国家自然科学基金资助项目(No. 61302127);江西省自然科学基金重大项目(No. 20161ACB20004)

The results show that the distributed face recognition method can effectively promote the efficiency and scalability of face recognition methods. For those extreme cases, after performance optimization, the maximum CPU using rate in agent has been declined by approximate 40%, which could effectively alleviate the time delay.

Key words: face recognition; distributed; agent; load balance; performance optimization

1 引言

人脸识别已从最初对背景单一的正面灰度图像的识别,发展到基于视频的人脸识别。基于视频的人脸识别包括 4 个部分:人脸检测、人脸跟踪、特征提取和匹配识别。在人脸检测方面, YanWen Wu 等人^[1]提出了一种结合肤色信息的 Adaboost 算法用来检测人脸, Rachid Belaroussi 等人^[2]则使用肤色信息进行人脸检测,但是这种算法不是很稳定。人脸跟踪算法方面,余九华等人^[3]提出使用改进的 Camshift 算法对人脸进行跟踪,可以同时跟踪多人,并且能够克服遮挡问题。另外, Veronica Vilaplana 等人^[4]提出了基于区域的 Meanshift 人脸跟踪算法,提高了跟踪的准确性。在多人脸识别方面, Binu Muraleedharan Nair 等人^[5]通过结合多姿态人脸数据库和模块加权主成分分析(Principal Component Analysis, PCA)算法,有效地提高了识别率。类似的,模块 PCA^[6]也被用来识别多张人脸。目前对于人脸识别系统的研究已经取得了很大的进步,可以快速精确地进行人脸识别。

然而,为了提高人脸识别效率,学者们已经不再拘泥于传统的人脸识别系统。有的学者提出一种分布式并行人脸识别系统^[7],该系统通过分割人脸数据库并由多个从属服务器同时进行匹配识别,有效地提高了人脸识别效率,但该分布式人脸识别系统仅适用于超大规模人脸数据库的人脸识别。另外,文献^[8-10]还提出了基于数据融合的分布式人脸识别模型,该模型根据人脸的五官将人脸图像分割成 4 部分,分别提取特征,然后对得到的特征数据进行融合。该方法在保证面部识别准确性的前提下提高了人脸识别的效率。然而,这种针对五官分割面部图像的分布式人脸识别方法在分割图片时需要手工切割,不仅使人脸识别更为繁琐,而且降低了效率。另外,上述分布式人

脸识别方法都是针对人脸图片进行人脸识别操作的,不能针对视频中的人脸进行识别操作。基于此, Rajeshwari J 等人^[11]提出了针对视频流进行识别的分布式人脸识别系统,该系统的基本思想是将 n 帧视频分别发送给 N 个客户端,各个客户端分别对接收到的图像执行人脸识别操作,但这种分布式人脸识别系统实质上仍然是针对图像中的人脸进行识别,效率较低,还十分耗费计算机资源。

显然,上面提到的人脸识别系统都不能同时对多个视频执行人脸识别操作。而且随着科技的不断进步,需要进行人脸识别的视频也逐渐增加,为了增加人脸识别系统的可扩展性,本文提出了一种分布式人脸识别方法。这种方法是将传统的人脸识别分成两部分,其中人脸检测、人脸跟踪和特征提取等操作,这些操作均由代理来处理,而人脸匹配识别则由服务器来执行。这种分布式人脸识别可以有多个代理,如果代理间的任务量相差很大,那么处理任务量较大的代理将会产生时间延迟,导致整体的人脸识别时间过长,并且代理端的 CPU 利用率也会激增至最大。为了解决这些问题,需要使用一种能够均衡分配各代理任务量的优化算法。任务均衡问题类似于任务调度问题,而遗传算法(GA)一直以来都是解决这类问题的常用算法^[12-13]。另外,遗传算法也适用于负载均衡问题^[14-20]。然而传统遗传算法还存在容易过早收敛及不稳定等缺点。因此,本文对传统遗传算法的交叉概率和变异概率进行了限制,并且增加了阈值函数来减少迭代次数。这种优化方法可有效提高基于代理的分布式人脸识别模型的时间效率和可扩展性。

2 分布式人脸识别

2.1 分布式人脸识别模型

本文提出的分布式人脸识别模型如图 1 所

示。由图 1 可知,人脸识别模型被分割成两部分:服务器和代理。每个代理可以同时处理多个视频,每个视频中有一个或者多个行人。代理端和服务器端的信息通过 SOAP (Simple Object Access Protocol) 进行相互传递。

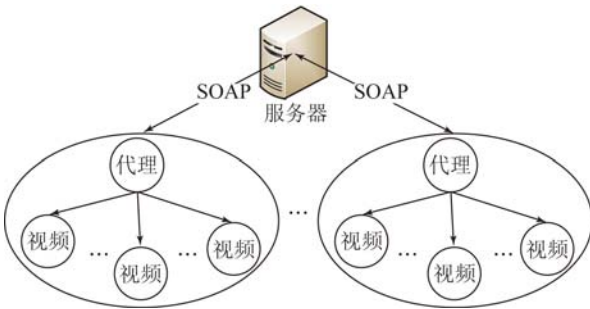


图 1 分布式人脸识别的模型图

Fig. 1 Distributed face recognition model

代理和服务器之间的关系如图 2 所示,代理主要对视频中的行人执行如下操作:人脸检测、人脸跟踪和特征提取。代理提取人脸特征以后,将特征信息通过 SOAP 发送给服务器,服务器接收特征信息,并将收到的特征信息与人脸数据库中的特征信息进行匹配识别,最后把识别结果返回给代理,代理会显示出识别结果。

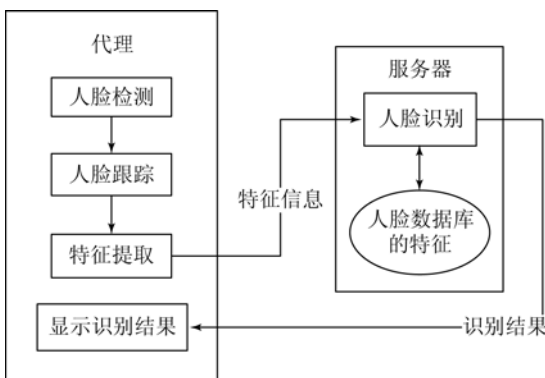


图 2 代理和服务器之间的关系图

Fig. 2 Relationship between agent and server

2.2 分布式模型中的多人脸识别

分布式人脸识别模型通过结合 PCA 和 FisherFace 对视频中的多个行人执行人脸识别。首先,使 PCA 将 K-L 变换后的特征值从大到小排列,挑选出相对大的特征值所对应的特征向量,构成一个 K-L 变换特征空间的子空间,来进行特

征提取,然后对特征数据进行降维。其次,将降维后的图像投影到 $N-c$ 空间中,再结合线性鉴别分析方法求出最优投影矩阵。其中 N 表示训练样本的总数目, c 表示的是人脸类别数。求解 FisherFace 最优投影矩阵的公式如下:

$$W_{pca} = \arg \max_w |W^T S W|, \quad (1)$$

$$W_{pca} = [W_{pca1} | W_{pca2} \cdots W_{pca k}], k \leq N - c, \quad (2)$$

$$\tilde{S}_w = W_{pca}^T S_w W_{pca}, \quad (3)$$

$$\tilde{S}_b = W_{pca}^T S_b W_{pca}, \quad (4)$$

$$W_{LDA} = \arg \left\{ \max \left[\frac{|W^T \tilde{S}_b W|}{|W^T \tilde{S}_w W|} \right] \right\}, \quad (5)$$

$$W_{opt} = W_{pca} W_{LDA}, \quad (6)$$

其中: w 对应了 W 中的列向量, S_b 和 S_w 分别代表高维空间中的类间离散度矩阵和类内离散度矩阵。与之对应的, \tilde{S}_b 和 \tilde{S}_w 分别代表低维空间中样本的类间离散度矩阵和类内离散度矩阵。LDA 的准则函数则为类内离散度矩阵和类间离散度矩阵的行列式的比值。

本文使用的人脸数据库是 ORL 数据库,并增加了 4 个测试视频中行人的人脸图像。由于使用的网络摄像头分辨率较低,所以增加的人脸图像也是低分辨率的。本文主要解决的是行人的正脸识别。图 3 是 ORL 数据库中增加的人脸图片。为了能够更精确地正面识别到行人,每个人取 10 张角度变化不大的正脸图片作为 ORL 数据库中的样本。



图 3 ORL 数据库中增加的面部图像

Fig. 3 Facial images added to ORL database

3 分布式人脸识别的性能优化

本文提出的分布式人脸识别模型中有多个代理,并且每个代理可以同时识别多个视频中的行人,为了避免代理处理任务量不均导致的时间延迟,对每个代理处理的任务量采用遗传算法(GA)进行负载均衡。

假设有 k 个代理要处理 m 个视频,每个视频

中有多个行人,第 i 个代理处理的 ID 号为 AID_i ,处理的视频集合为 $V_i = \{V_{i1}, V_{i2}, \dots, V_{in_i}\}$,其中 i 的范围是 $(0, k]$, n_i 是第 i 个代理中的视频个数, $(0 < n_i \leq m)$, V_i 中第 j 个视频的行人个数为 P_{ij} , $(0 < j \leq n_i)$ 。

首先对 GA 中的染色体进行初始化,以第 i 个代理为例,将 AID_i, V_{ij}, P_{ij} 这 3 个参数作为一个节点,由于第 i 个代理中有 n_i 个视频,所以就有 n_i 个这样的节点,将这 n_i 个节点作为第 i 个染色体的基因。第 i 个染色体由 $Chrom_i$ 来表示, $Chrom_i$ 初始化结果如下所示:

$$Chrom_i: (AID_i, V_{i1}, P_{i1}), (AID_i, V_{i2}, P_{i2}), \dots, (AID_i, V_{ij}, P_{ij}), \dots, (AID_i, V_{in_i}, P_{in_i}).$$

本文采用的适应值函数如下:

$$F_i = c / \sum_{j=1}^{n_i} P_{ij}, \quad (7)$$

其中 c 是小于 1 的随机数。

其次,使用轮盘赌方式进行选择操作,根据计算公式(8)获得选择结果。考虑个体优劣差距较大的情况,根据染色体的数量,选择不同质量的染色体进行遗传。

$$P_{chrom}(i) = F_i / \sum_{i=1}^k F_i. \quad (8)$$

在遗传操作中,选择两点交叉方法进行交叉操作,根据染色体长度的不同,设定这两个交叉点的选择范围,令第一个交叉点的位置为第一个基因节点,假设这两条染色体中最短的染色体长度为 L_{min} ,则第二个交叉点的选择范围即为 $[L_{min}/2, L_{min}]$ 。

为了避免过早收敛使算法进入局部最优,及保证种群的多样性,根据染色体的适应值对交叉概率 P_c 和变异概率 P_m 增加了一些随机性操作, P_c 和 P_m 的计算公式如下所示:

$$P_c = k_1 \frac{\bar{F}}{\text{Max}F_1} + k_2 \frac{\bar{F}}{\text{Max}F_2}, \quad (9)$$

$$P_m = k_3 \frac{\bar{F}}{\text{Max}F_1} + k_4 \frac{\text{Min}F}{\bar{F}}, \quad (10)$$

其中: $\text{Max}F_1$ 是染色体中最大的适应值, $\text{Max}F_2$ 是染色体中第二个大的适应值, $\text{Min}F$ 是染色体中最小的适应值, k_1, k_2, k_3 和 k_4 是 $[0, 1]$ 的随机数。

为了尽快缩小染色体之间的优劣差距,在传统的 GA 算法中加入了迁移操作,即通过将最差染色体中的最大基因迁移至最优的染色体中来完

成该操作。

最常用的停止准则是选择一个比较大的数作为迭代次数。但在本文中,为了不错过最优解,设置了一个阈值函数,公式如下:

$$\bar{F} / \text{Max}F > \epsilon, \quad (11)$$

其中:参数 ϵ 是一个阈值,这个阈值小于 1。

本文负载均衡操作的步骤如下:

步骤 1:服务器接收代理发送的信息,并初始化为染色体。

步骤 2:根据公式(7)计算每条染色体的适应值。

步骤 3:根据公式(8)执行选择运算。

步骤 4:计算公式(11),如果没有满足预定的阈值,则执行步骤 5~步骤 7;如果满足预定阈值,则停止。

步骤 5:根据公式(9)计算交叉概率 P_c ,判断是否执行交叉操作,并且重复步骤 2 和步骤 3。

步骤 6:根据公式(10)计算变异概率 P_m ,如果不能执行变异操作,则执行迁移操作,并且重复步骤 2 和步骤 3。

步骤 7:转到步骤 4 继续。

4 测量实验与结果

基于视频的集中式人脸识别系统不适合大规模人脸识别,在可扩展性方面具有局限性,而已有的分布式人脸识别系统则无法对视频进行处理。为了解决这些问题,本文提出了基于代理的分布式人脸识别模型,其不仅可以同时处理多个视频,并且增加了可扩展性。

本节将针对本文提出的基于代理的分布式人脸识别系统的性能进行评估,分多种情况进行仿真实验。所有的视频都是通过分辨率为 704×516 的 Axis 215 PTZ 网络摄像头采集的,视频的帧率为 5 frame/s。服务器和代理采用的处理器都为 Intel(R) Core(TM) i7-3770M CPU@3.40 GHz、操作系统为 Windows7 的计算机。改进的遗传算法中,停止准则的阈值 ϵ 设置为 0.87。

由于篇幅有限,本文仅给出其中一个视频的部分识别结果。图 4(彩图见期刊电子版)表示代理显示的两个人脸的识别结果,分别给出了视频中第 75 帧、第 87 帧、第 92 帧、第 105 帧、第 112 帧和第 127 帧的识别结果。



图 4 代理显示的多人脸识别结果

Fig. 4 Multiple faces recognition results shown in agent

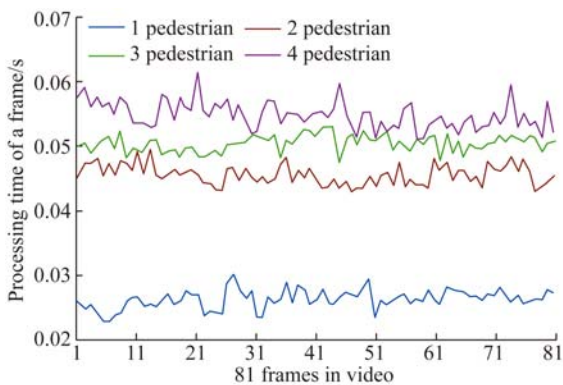


图 5 处理不同人数视频的处理时间

Fig. 5 Processing times of video with different number of pedestrians

为了分析处理时间与视频中人数的关系,先令代理单独处理一个视频,由于测试视频中行人总数最多是 4 人,所以做了 4 次实验。中间的 81 帧视频的处理时间如图 5(彩图见期刊电子版)所示,每一帧视频的处理时间包括代理的检测时间、

跟踪时间和特征提取时间,以及服务器的识别时间。

对于不同人数的视频,平均处理时间见表 1。从图 5 和表 1 中可以看出,处理时间随着视频中人数的增加而增加。

表 1 处理不同人数视频的平均时间

Tab. 1 The average processing times of video with different number of pedestrians (s)

视频中的人数	1	2	3	4
平均处理时间	0.025 8	0.044 3	0.048 2	0.518 5

接下来,模拟 2 个代理的情况,分别把 4、6、8、10、12、14 个视频随机分配给 2 个代理,未进行任务均衡前,2 个代理中所有视频中的行人总数分别是 7、13、17、22、28、34。将本文使用的遗传算法(GA)与随机划分方法(Random)和文献[18]中提出的粒子群优化算法(PSO)分别进行任务均衡,以对比方法性能。

图 6 是采用不同方法时,服务器处理一帧视频的平均时间。由图 6 可知,服务器仅执行人脸识别操作时,处理时间的数量级很小,使用 GA 均衡后,服务器端的人脸识别时间明显比使用 Random 方法和 PSO 算法要低很多。

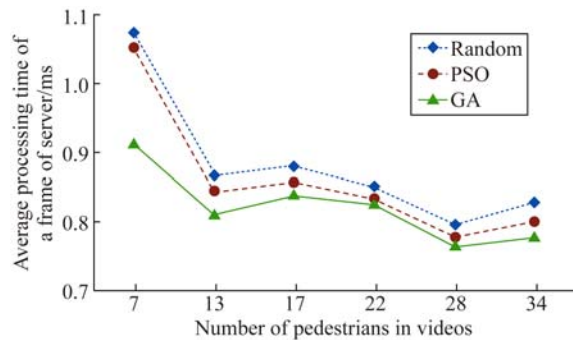


图 6 服务器使用不同方法处理一帧视频的平均时间

Fig. 6 Comparison of average processing times for one frame video in server

然后比较使用 3 种算法均衡任务后处理一帧视频的平均时间。比较结果如图 7 所示,随着任务量的增加,处理时间也随之增加,所以任务分配越均匀,处理一帧视频的平均时间就越小。平均来看,使用 GA 均衡后,一帧视频的平均处理时间

比 Random 算法降低了 18.92% 左右,比 PSO 算法降低了大约 8.7%。

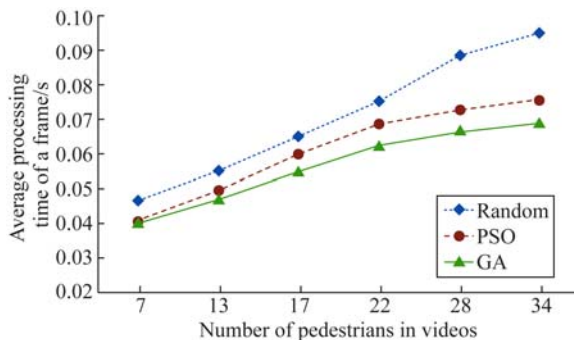


图 7 几种方法处理一帧视频的平均时间的结果对比
Fig. 7 Comparison of the average processing time of a frame processed by different methods

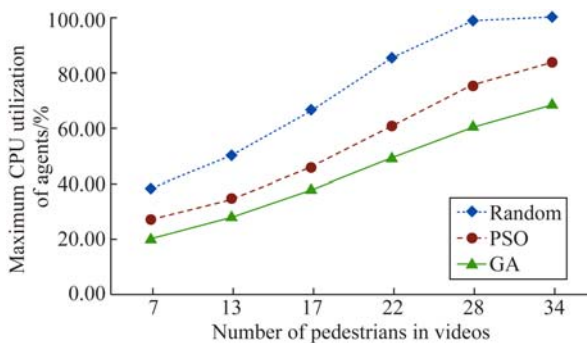


图 8 代理中最大的 CPU 利用率

Fig. 8 Compare the maximum CPU utilization of agents

最后,比较了使用 3 种算法进行负载均衡后 2 个代理中最大的 CPU 利用率,结果如图 8 所示,使用 Random 算法后,当视频中总人数达到 28 人时,代理中最大的 CPU 利用率已经几乎达

到 100%;使用 PSO 算法进行负载均衡后,代理中最大的 CPU 利用率达到了 74.73%,而使用 GA 优化后,代理中最大的 CPU 利用率为 60.56%,明显比前两种方法要低很多。

5 结 论

目前分布式人脸识别系统大部分是针对人脸图片设计的,即便有针对视频的分布式人脸识别模型也不具有可扩展性,因此提出了一种新型的可扩展的分布式人脸识别方法。该方法将人脸识别系统分为服务器端和代理端两部分,服务器执行人脸匹配识别操作,代理则执行人脸检测、跟踪和特征提取等操作。该方法可以同时识别多个视频中的行人,有效地提高了人脸识别效率,并增加了可扩展性。

在此基础上,针对多个代理由于处理任务量不均匀所导致的一系列问题,设计了代理的负载均衡方法对分布式人脸识别模型进行性能优化,并使用遗传算法对代理端的任务量进行了负载均衡。通过大量实验证明,在相对极端的情况下,相比随机划分方法和粒子群优化算法而言,使用遗传算法进行性能优化后,处理一帧视频的平均时间分别降低了约 19.37% 和 9.43%,并且代理中最大的 CPU 利用率平均分别降低了 28.52% 和 18.27%,有效缓解了处理视频时间过长的问题,并且降低了 CPU 利用率。这种优化方法有效地提高了分布式人脸识别方法的性能和可扩展性。

参考文献:

- [1] WU Y W, AI X Y. An improvement of face detection using AdaBoost with color information [C]. *International Colloquium on Computing, Communication, Control, and Management*. IEEE Computer Society, Washington, DC, USA: ISECS, 2008: 317-321.
- [2] BELAROUSSI R, MILGRAM M. A comparative study on face detection and tracking algorithms [J]. *Expert Systems with Applications*, 2012, 39(8):7158-7164.
- [3] 余九华,王敬东,李鹏. 基于 Camshift 的人脸跟踪算法[J]. *计算机技术与发展*, 2008, 18(9):12-15.
- [4] SHE J H, WANG J D, LI P. Face tracking algorithm based on Camshift [J]. *Computer Technology and Development*, 2008, 18(9):12-15. (in Chinese)
- [5] VILAPLANA V, MARQUES F. Region-based mean shift tracking: application to face tracking [C]. *IEEE International Conference on Image Processing, Barcelona, Spain: ICIP*, 2008: 2712-2715.
- [6] NAIR B M, FOYTIK J, TOMPKINS R, et al. Multi-pose face recognition and tracking system [J]. *Complex Adaptive Systems*, 2011, 6(1):381-386.

- [6] FOYTIK J, SANKARAN P, ASARI V. Tracking and recognizing multiple faces using kalman filter and modular PCA [J]. *Procedia Computer Science*, 2011, 6(1):256-261.
- [7] JIANG C H, SU G D, LIU X D. A distributed parallel system for face recognition [C]. *International Conference on Parallel and Distributed Computing, Applications and Technologies, Chengdu, China: PDCAT*, 2003:797-800.
- [8] ZHANG Z, GUO Y, SONG G. A distributed face recognition framework based on data fusion [J]. *International Journal of Database Theory & Application*, 2014, 7(4): 87-98.
- [9] YAN Y J, OSADCIW L A. Distributed wireless face recognition system [C]. *Electronic Imaging, International Society for Optics and Photonics, San Jose, America: SPIE*, 2008:68200A-68200A-12.
- [10] RAZZAK M I, KHAN M K, ALGATHBAR K, et al. . Energy efficient distributed face recognition in wireless sensor network [J]. *Wireless Personal Communications*, 2011, 60(3):571-582.
- [11] RAJESHWARI J, KARIBASAOOA K. Face recognition in video systems on homogeneous distributed systems [J]. *International Journal of Advanced Computer and Mathematical Sciences*, 2013, 4(1): 143-147.
- [12] WANG L, CAI J C, LI M. An adaptive multi-population genetic algorithm for job-shop scheduling problem [J]. *Advances in Manufacturing*, 2016,4(2):142-149.
- [13] LU Q, LI S, ZHANG W, et al. . A genetic algorithm-based job scheduling model for big data analytics [J]. *Eurasip Journal on Wireless Communications & Networking*, 2016, 2016(1):1-9.
- [14] NOURZADEH R, EFFATPARVAR M. A genetic-fuzzy algorithm for load balancing in multiprocessor systems [J]. *International Journal of Computer Applications*, 2014, 101(10):39-42.
- [15] TED S, KENNETH N B. Wireless LAN load-balancing with genetic algorithms [J]. *Knowledge-Based Systems*, 2009, 22(7):529-534.
- [16] JIAN A, CHAUHARI N S. Genetic algorithm based concept design to optimize network load balance [J]. *Ictact Journal on Soft Computing*, 2012, 2(4):357-360.
- [17] DASGUPTA K, MANDAL B, DUTTA P, et al. . A Genetic Algorithm (GA) based load balancing strategy for cloud computing [J]. *Procedia Technology*, 2013, 10(2):340-347.
- [18] MIN W D, LIU Y H, KE Y Z, et al. . Using particle swarm optimization algorithm to improve multi-agents network management [J]. *Journal of Computational Information Systems*, 2014, 10(2):739-746.
- [19] 劳达宝,周维虎,李万红,等. 基于遗传算法的柱面光栅测角技术研究[J]. *红外与激光工程*, 2015, 44(7):2182-2188.
- LAO D B, ZHOU W H, LI W H. Cylindrical grating angle measurement technology based on genetic algorithm [J]. *Infrared and Laser Engineering*, 2015, 44(7):2182-2188. (in Chinese)
- [20] 郭天太,洪博,潘增荣,等. 基于改进的SVM在矿井气体定量分析中的应用[J]. *红外与激光工程*, 2016, 45(6):203-210.
- GUO T T, HONG B, PAN Z R, et al. . Application of improved SVM in quantitative analysis of mine gas concentration [J]. *Infrared and Laser Engineering*, 2016, 45(6):203-210. (in Chinese)

作者简介:



闵卫东(1966—),男,江西会昌人,博士,教授,博士生导师,1989年、1991年、1995年于清华大学分别获得学士、硕士和博士学位,1995年至2014年在加拿大 Alberta 大学和 Corel 等跨国公司从事研究工作,主要从事图形图像处理、图形学、分布式系统、智慧城市等的研究。E-mail:minweidong@ncu.edu.cn



石杰(1990—),女,内蒙古赤峰人,硕士研究生,2013年于天津工业大学获得学士学位,主要研究分布式人脸识别及其性能优化,分布式网络管理。E-mail:714491893@qq.com