doi：10.11835/j.issn.1000－582X.2022.08.012

智慧园区环境下的多模态多核学习身份识别算法研究

刘安强[1]，张碧川[1]，郭　栋[1]，甘　梅[2]，刘　航[3]，李　幸[3]，陈　婕[4]
（1. 陕西陕煤曹家滩矿业有限公司，陕西 榆林 719000；2. 中煤科工集团重庆研究院有限公司，重庆 400039；
3. 重庆梅安森科技股份有限公司，重庆 400050；4. 重庆邮电大学，重庆 400065）

摘要：智慧园区的建设推动着企业与城市的发展，传统的园区管理方式已不再适用于产业融合创新的智慧园区。以曹家滩园区为例，设计智慧园区平台总体框架，针对园区中身份识别存在识别环境差、效率低、准确率低等问题，提出一种基于多模态多核学习的身份识别算法。所提算法将视频数据中的数据分为图像、音频，并采集个人信息的文本，并将三种模态的信息输入同一样本空间中，通过引入间隔约束的多核学习算法，保留不同模态的差异性和相似性，并进行特征融合与决策融合，最终采用分类器与评分机制输出身份识别结果。通过公开的视频数据集与曹家滩园区数据集进行实验，实验结果表明本文所提算法最高准确率达到 97.2%，与传统算法相比有较大优势。

关键词：智慧园区；身份识别；多模态；多核学习
中图分类号：TP391　　**文献标志码**：A　　**文章编号**：1000-582X(2022)08-130-11

Research on multi-modal and multi-kernel learning identity recognition algorithm in smart parks

LIU Anqiang[1]，*ZHANG Bichuan*[1]，*GUO Dong*[1]，*GAN Mei*[2]，*LIU Hang*[3]，*LI Xing*[3]，*CHEN Jie*[4]
(1. Shaanxi Shanmei Coal Caojiatan Mining Co., Ltd., Yulin, Shaanxi 719000, P. R. China; 2. CCTEG Chongqing Research Institute Co., Ltd., Chongqing 400039, P. R. China; 3. Chongqing MAS Science and Technology Co., Ltd., Chongqing 400050, P. R. China; 4. Chongqing University of Posts and Telecommunications, Chongqing 400065, P. R. China)

Abstract: The construction of smart parks promotes the development of enterprises and cities, and traditional park management methods are no longer suitable for smart parks with industrial integration and innovation. This paper takes Caojiatan Park as an example to design the overall framework of the smart park platform. Aiming at the problems of poor recognition environment, low efficiency and low accuracy in the park's identity recognition, this paper proposes an identity recognition algorithm based on multi-modal and multi-kernel learning. The proposed algorithm divides the data in the video data into images and audio, and collects the text of personal information, and inputs the information of the three modalities into the same sample space. By introducing a multi-kernel learning algorithm with interval constraints, the difference is retained to the greatest extent. The difference and similarity of modalities are combined with

收稿日期：2021-01-06
基金项目：重庆市技术创新与应用发展专项重点资助项目(cstc2019jscx-fxydX0039)；曹家滩矿井智能化项目建设平台资助项目(CKH/2-2017)。
Supported by Chongqing Technology Innovation and Application Development Special Key Project (cstc2019jscx-fxydX0039) and Shaanxi Caojiatan Mining Intelligent Developing Project (CKH/2-2017).
作者简介：刘安强(1987—)，中级工程师，主要从事智慧矿山建设研究，(E-mail)627789682@qq.com。

feature fusion and decision fusion, and finally the classifier and scoring mechanism are used to output the identification results. Through experiments on the public video dataset and Caojiatan Park dataset, the experimental results show that the algorithm proposed in this paper has a maximum accuracy of 97.2%, which has a great advantage over traditional algorithms.

Keywords: smart park; identification; multi-modal; multi-kernel learning

在科技的发展和国家相关政策的推动下，以产业聚焦为手段的各类园区发展迅速。目前，各大传统园区及企业逐渐向新领域、新技术、新局面蓬勃发展。产业园区作为多方向多领域集群发展的有效途径，是区域经济与多维产业联动的桥梁，各类园区作为对外开放、招商引资、管理创新的主要载体，为各个产业之间的联动、共享和协作提供了可靠的发展平台[1]。目前，以大数据、机器学习及物联网等技术为核心的新一代智慧园区已成为各类工业园区、商业园区和文化产业园区的建设和发展目标[2]。人脸识别技术对于园区的环境监控、日常监控、安防监控等领域提供便捷又智能的身份识别服务[3]。

人脸识别(face recognition)技术[4]是指通过获取的图像、视频或者是红外摄像获取的人像，通过面部信息的挖掘建模确定本人在先验数据库中的身份。人脸识别因其广泛应用性受到学者们的广泛研究，并在长期研究中产生了多样化的方法，具有较高的研究热度[5,6]。人脸识别技术更贴合智慧园区的应用，同时现有的研究已经证明多模态技术能够大大提升人脸识别的准确性[7-9]。人脸作为固有的生物特征之一，不同个体之间具有很强的辨识性，为身份识别的挖掘建模过程中提供了一个有效特征。然而传统的人脸识别对于图像采集大部分是在光线充裕的环境下进行，忽略了人脸因角度或是人为因素而无法有效采集的问题。因此传统的人脸识别算法在真实环境下难以达到一个稳定的表现。

由于不同模态信息存在多源异构性(图像、音频、文本等)，而且存在不同的空间中，导致不同模态之间的信息难以处理。目前的多模态融合算法主要是图像合成的方法，即将图像作为基础，将其他模态转化为图像的形式并与原图像建立关系，利用这种关系解决多模态的匹配问题。常用的算法包括马尔可夫随机场，本征变化[10,11]，耦合字典等。文献[12]针对相同图像不同分辨率之间存在着相同的稀疏系数，提出利用耦合字典作为中间工具、低分辨率的图像作为输入进行高分辨率图像的合成。虽然图像合成的方法能够通过多模态之间的联合学习保证特征的可靠性，但由于合成算法的特性融合，使其在多模态下的合成与识别缺乏普适性[13]。

在本文中，通过引入了间隔约束扩展 MKL 的方法并引入维度规范化核函数对多模态学习进行间隔维度约束与特征融合约束，并加入决策融合算法，提出了融合多模态的身份识别新框架，通过多核学习算法提升算法的适用性，并使融合后的特征发挥出最大的判别能力。

1　智慧园区平台总体架构

园区智慧管理建设作为曹家滩智慧矿山建设的重要组成部分，有力推动着全区的全方位发展。目前，曹家滩办公园区约占地 20.3 万 m^2(井口以上地面部分)，园区内现有应用领域包括：智慧办公、智慧服务、智慧管理，新建业务应用系统 16 套，集成现有系统 7 套(消防系统、安防系统、人员定位系统、培训系统、停车系统、智慧餐厅、人力资源系统)。

而人脸识别技术作为人工智能现实应用中的一部分，在近年来得到了飞速的发展。人脸识别技术所具有的安全、便捷、可靠等特性，促进其在各行业的应用与推广，它能够对特定身份进行生物特征识别[14]。通过对人脸面部数据的提取、特征数据转化处理和对比分析来准确识别个人的身份信息，相较于传统密码类非生物识别技术而言，人脸识别技术更加准确、便利和经济，既能迎合管理方和企业追求高效的目的，又能满足员工对低时延、便利的需求。将人脸识别技术应用于园区的智慧化管理当中，能够为智慧化写字楼、智慧化生活区等提供安全便捷的人员出入识别核查管理，提升园区内使用人员的舒适感。在园区的生活区域中，人脸识别技术与员工日常生活所采用的生物特征识别模式基本相同，具有良好的自然性和便捷性[15]。将人脸识别技术运用于曹家滩园区的智慧管理建设方案当中，通过建立人脸识别系统，完成对园区员工的身份识别等功能，其逻辑结构如下图 1 所示。

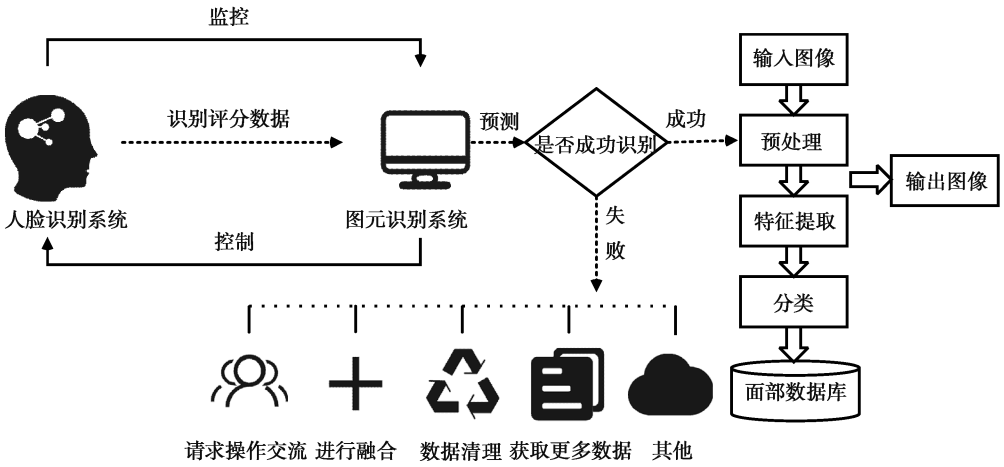


图1 园区人脸识别技术设计方案

Fig. 1 Scheme of face recognition technology in the park

首先,根据人脸识别系统数据库中已有数据来分析员工身份,若识别成功,则进行图像输入、面部表情数据特征预处理、特征提取、特征分类等步骤,并输出最终的人脸图像供平台调用;若识别失败,则用户可以再次请求系统交互操作,并对数据进行融合、清洗等操作,获取更多的数据特征,供系统再次进行判定。

此外,在园区内所设立的人脸识别设备无需携带卡片或摆出特定动作或指示,即可完成身份识别,其用户体验及操作难度优于虹膜、指纹等识别方式。在如今疫情常态化的情况下,人脸识别技术无需接触识别设备,即可完成对用户的识别,提高了园区内安全卫生管理,保障园区疫情防控措施落实。此外,人脸识别技术具有并发性,在人员基数大、分布相对集中的园区环境下,利用人脸识别技术可以同时进行多个人员的面部特征识别,提高用户工作效率,优化用户体验。

目前,园区的建设和发展主要以AI、物联网及大数据等新兴技术构建智能园区为重点,实现园区场景智能化、管理精细化、运营可视化等智慧管理,通过智能化场景提升用户体验,以精细化管理提升园区管理效率,采用数字化运行增加园区效能。曹家滩园区智慧管理总体架构设计如下图2所示,其中主要由IoT(the Internet of Things)应用、WEB应用、数据服务、设备模拟、数据分析、设备开发等模块组成。

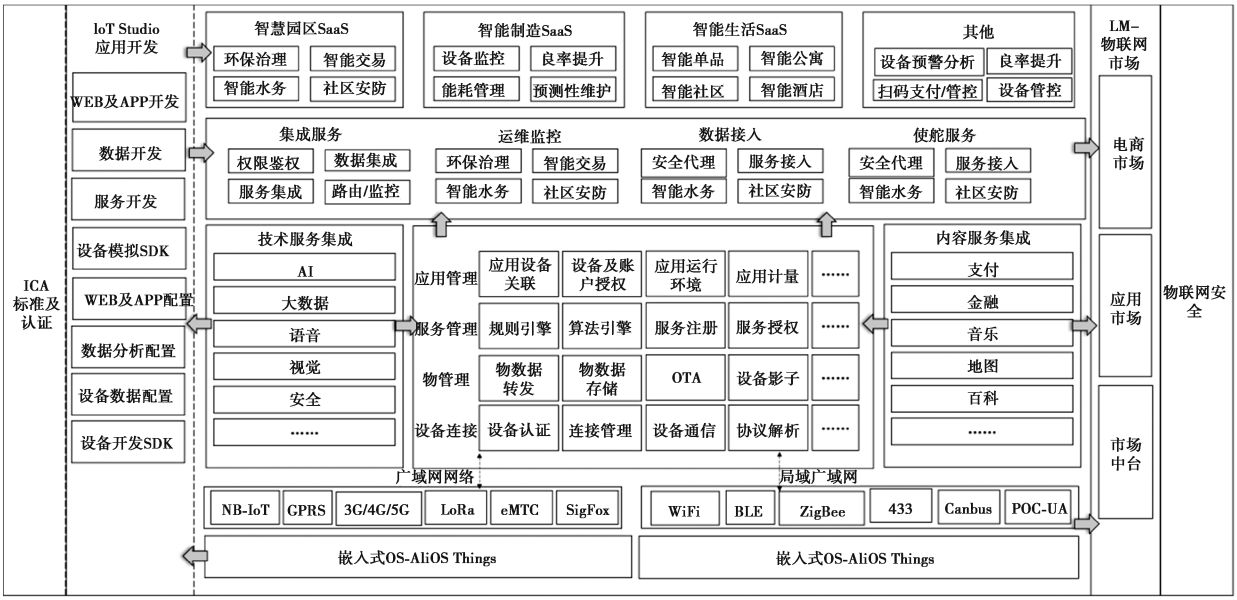


图2 曹家滩园区管理平台架构图

Fig. 2 Caojiatan Park management platform architecture diagram

在该架构模式下，曹家滩园区智慧管理以技术服务、内容服务集成以及 LP-IoT 基础 PaaS 平台为支撑，提供集成服务、运维监控、数据接入及使能服务等功能，实现智能城市、智能制造、智能生活等 IoT 功能应用。此外，应用 IoT 技术实现无感、便捷、高效的智慧应用以及用餐、购物、通行、体检等高效的用户体验，通过物联网、云计算以及人工智能打造“云工作台＋聚合共享应用”的智慧园区管理办公平台，聚焦智能化场景应用，构建园区智慧管理新模式。

2　多模态身份识别

2.1　基于间隔维度约束的 MKL 模型——MDMKL

内核机器学习领域的一个最近发展方向就是采用多核学习（MKL），通过多个内核在同一个框架中进行优化，能使在其监督学习或半监督学习中发挥更好的作用。它不需要关心特征空间的数据异构性、数据无规律性、数据分布不均匀、数据量大等问题。MKL 具有自动调节内核参数、描述数据表示的各种特征，并能够并行处理各种多源异构数据的特点。同时，还能提升分类器的泛化能力增加模型的可解释性。

根据最新研究表明，MKL 方法能够在有效对具有鉴别性的基本特征进行有效融合的同时，忽略掉不具有鉴别性的特征。MKL 中包含了高斯 RBF 核，它具有将基本特征通过核函数转换到高维空间的作用。一般来说，为找到一个能够对不同特征都可以使用的内核参数是十分困难的，主要是由于不同参数对于不同特征的影响效果大不相同。因此，MKL 难以在多个模态中获取所有基本特征的鉴别能力。

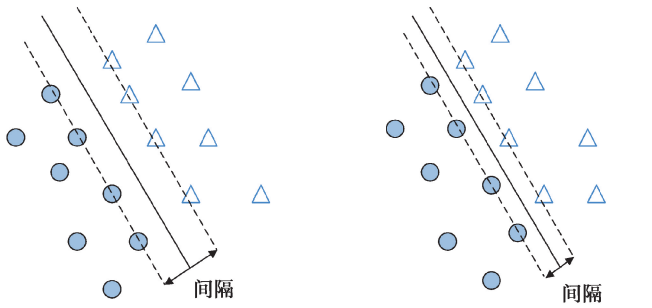
为此通过引入间隔约束，提出了基于间隔维度约束的多核学习（MDMKL）方法，将数据维度通过高斯 RBF 核归一化到同一空间中，并在该空间中利用多模态特征融合算法，结合使用间隔约束扩展 MKL 保证特征融合的有效性。MDMKL 方法会通过给予不同模态特征以不同的关注度来辨别不同模态特征的识别能力。相较于传统的 MKL，MDMKL 在构造最优组合核参数时，会将不重要的特征分配较低的权值而将具有鉴别能力的权值分配更大的权值，以保证不同模态之间能够充分利用相互之间的关系保证特征融合的准确性。

2.1.1　间隔约束

MKL 构建了一个良好的框架，能够通过给最具有鉴别性的基本特征赋予一个较大的权值，来保证特征融合的稳定性。不同于直接串联方法，MKL 方法可以有效地避免特征维度很大的鉴别性差的基本特征带来的污染。

MKL 方法存在选择的基本特征特别少的缺点，MKL 在进行样本区分中仅会选择一个或两个在高维空间中有区别的基本特征。由于不同模态的特征在最优高维空间中的核参数必然会显著不同，就会导致传统 MKL 无法充分利用基本特征的最大鉴别能力。

为了解决传统 MKL 无法很好地区分不同特征的缺点，借助于 SVM 算法的间隔约束理论将间隔维度约束引入到多核学习之中，提出了间隔维度约束多核学习（MDMKL）。图 3 即表示了不同特征的间隔图，在图 3(a)中，分离间隔距离较大时就能有效地区分基本特征，而图 3(b)中，用于区分特征的超平面间隔距离相对较小，代表了在该特征空间下两个类别的类间相似度小，不易区分。



(a)分离间隔距离较大　　(b)分离间隔距离较小

图 3　间隔约束图

Fig. 3　Margin constraint

利用分离的间隔作为评价多个类别之间的基本特征的鉴别能力的指标。而这种判别指标在MKL算法中能够有效地寻找最优的特征组合形式。在模型中损失函数定义为式(1),判别指标可以用式(2)表示,即对损失函数倒数的开平方根的形式。

$$f=\frac{1}{2}\|w\|^2+C\sum_i\xi_i+\sum_k\sigma_k d_k, \tag{1}$$

$$m_k=\frac{2}{\|w_k\|}\approx\frac{\sqrt{2}}{\sqrt{f_k}}=\frac{\sqrt{2}}{\sqrt{\frac{1}{2}\|w\|^2+C\sum_i\xi_i+\sum_k\sigma_k d_k}}。 \tag{2}$$

为实现间隔约束,首先选择一个基本特征作为参考,其权重 d_s 一般设置为0.5,间隔值为 m_s,其他基本特征的分离间隔设为 m_k。利用 m_s 与 m_k 的边际比率,可以将第 k 个基本特征的基本权值约束在上限 B_k^U 与下限 B_k^L 之间,在进行多核学习时更新该约束(式(3))限制权值。而对于上限 B_k^U 和下限 B_k^L 可以通过公式(4)进行计算。对于不同模态的权重采用如式(5)的梯度下降策略进行更新。

$$B_k^L\leqslant d_k\leqslant B_k^U; \tag{3}$$

$$\begin{cases}B_k^L=\left(\dfrac{m_k}{m_s}\right)^n\cdot d_s,\\ B_k^U=\left(\dfrac{m_k}{m_s}\right)^n\cdot d_s\cdot(1+\delta);\end{cases} \tag{4}$$

$$d_k=d_k-\frac{\partial f}{\partial d_k}。 \tag{5}$$

式(4)中通过 n 控制了上下限对于间隔的依赖程度。随着 n 增加,B_k^L 和 B_k^U 的值对 m_k 和 m_s 之间的比值更加敏感,而当 n 较小时上下限就不受到间隔比值的影响,而是 δ 对于上下限的范围进行控制。经过多次对比实验,当 n 设置为1.5、δ 设置为1时模型表现效果能够达到最好的水平。

2.1.2 引入间隔约束的多核学习

高斯核函数(RBF)由于其在图像领域的出色的表现,使其被大量推广到其他领域。高斯核函数RBF可用式(6)表示。

$$K(x_i,x_j)=\exp(-\gamma\sum_{q=1}^{D}(x_{i,q}-x_{j,q})^2), \tag{6}$$

式中:D 为样本特征维数;x_i 和 x_j 分别表示第 i 个样本和第 j 个样本;$x_{i,q}$ 和 $x_{j,q}$ 是特征向量中的第 q 个元素;γ 是RBF核参数,它能够确定从低维特征空间到高维空间的映射的维度大小。

为便于空间维度的转换,特征向量首先会进行归一化到[0,1)之间,当 γ 值在其他参数不变的情况下以一个固定值增加时,式(6)的值将会减小。

根据MKL算法融合的依据,对于不同模态而言其特征存在于不同的样本空间之中,因此对于不同模态必定会有不同的参数。因此,MKL算法在不同模态的融合之中无法发挥最优的作用,也无法对于不同模态之间的特征做出很好的判别。

归纳来说,在MKL中,无法将所有的基本特征作为判别特征,只能选择那些最具辨识能力的基本特征。因此,MKL对于不同模态的所有类型特征无法充分利用。

由于无法满足多模态的需求,因此,提出将维度参数 γ 进行RBF核函数标准化,维度标准化的RBF核函数可表示为:

$$K(x_i,x_j)=\exp(-\frac{\gamma}{D}\sum_{q=1}^{D}(x_{i,q}-x_{j,q})^2)。 \tag{7}$$

通过除以特征维度 D 进行标准化,该步骤能够消除特征维度 D 对 γ 选择的影响,使不同模态的所有基本特征获得类似的特征维度参数值,且MDMKL算法能够发挥出不同模态在基本特征上的判别能力。

对于特征向量 x_i 在高维空间的组合特征为 $\varphi(x_i)$,其核函数组合如式(8)所示。

$$K_{opt}=\sum_k d_k\cdot K_k, \tag{8}$$

式中：d_k 一般设定为 0.5，K_k 代表第 k 个核函数，损失函数 f 为计算方便常常采用极大极小对偶化进行解决，如式(9)所示。

$$\begin{aligned}\underset{\partial_i}{\mathrm{Max}} f_D &= \sum_i \partial_i - \frac{1}{2}\sum_{i,j}\partial_i\partial_j y_i y_j K_{\mathrm{opt}}(x_i,x_j) + \sum_k \sigma_k d_k,\\ &\text{s.t. } 0 \leqslant \partial_i \leqslant C;\ \sum_i \partial_i y_i = 0,\end{aligned} \tag{9}$$

式中：∂_i 是拉格朗日乘子，$\sum_k \sigma_k d_k$ 是一个常数，之后采用梯度下降算法更新权重 d_k，如式(6)所示，且由间隔约束进行限制。

计算出特征的最终权重后，采用最优核对支持向量机分类器进行训练，分类器如式(10)所示。

$$g(x) = \sum_i \partial_i y_i \sum_k d_k K_k(\boldsymbol{s}_i, x) + b, \tag{10}$$

式中：$\boldsymbol{s}_i$ 是支持向量，利用一对一的方式进行多分类实现身份识别。

2.2　决策融合算法

对于决策融合而言，不会在特征层面上采用特征融合的方式融合，而是对于不同模态分配不同的分类器，将分类的输出结果作为评分。具体而言，分类器的输出结果将会转化为一个样本的可能概率值，通过对每个分类器结果分配不同权值后再进行加权，最终将选择概率最大的标签值作为分类的结果输出。

针对于智慧园区中，用于身份识别任务，可以使用模态分别包括图像、文本和音频模态。因此，不同模态的特征向量分别为 x_1、x_2 和 x_3，与此对应 λ_1、λ_2 和 λ_3 分别代表了图像、文本和视频模态的分类器，对于不同模态的估计概率可以用 $p(w'_k \mid x_i, \lambda_i)$，$i=1,2,3$ 表示，其加权公式如式(11)所示。

$$\begin{aligned}p(w_k \mid x_1, x_2, x_3) &= \sum_{i=1}^{3} [p(w'_k \mid x_i, \lambda_i)]\, w_i, k = 1,2\ldots,n,\\ w^* &= \max_k p(w_k \mid x_i),\ k = 1,2\ldots,n,\end{aligned} \tag{11}$$

式中：w_i 表示不同模态的权重值；n 表示总的样本类别数量即园区中人的数量；w'_k 代表待融合的第 k 个样本的标签值；w_k 代表融合后的第 k 个样本标签的概率；参数 w^* 表示最终估计样本标签概率。

由于 MDMKL 时间复杂度较高，仅单独使用特征融合会导致模型的整体时间表现比较差。因此提出一个采用特征融合与决策融合的集成新框架 MMMKL，如图 4 所示。其中 M_1，M_2，M_3 表示不同模态，其中模态 M_1 和 M_2 采用 MDMKL 算法进行特征融合，并将结果利用决策融合算法与模态 M_3 进行融合。

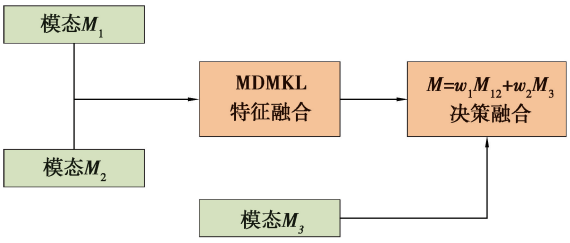


图 4　多模态混合融合框架 MMMKL

Fig. 4　Hybrid multi-mode fusion framework

2.3　算法设计

MMMKL 的流程如图 5 所示，该模型通过引入 SVM 思想中的间隔约束条件提不同模态特征的身份信息，解决了传统的 MKL 算法对于不同模态的特征提取算法没有较强鉴别能力的问题，同时解决了模型过于复杂难以直接求解的问题。采用将问题转化为对偶问题来简化求解过程，在对偶问题求解过程中，为获取一个固定的特征权重值，利用梯度下降法获取最优值。为避免模态间的特征过多导致难以进行区分，为判别能力差的特征分配一个较小的权值，并合并权值较小的权值达到模态鉴别能力最大化。

w_1 和 w_2 对应着融合后的特征 M_{12} 与待融合特征 M_3 的权值，其中 $w_1+w_2=1$。w_1 的值会根据迭代计算由 0.1 开始每次增加 0.1，直到达到 0.9，对应的 w_2 值由 0.9 减小到 0.1，通过训练可以达到最佳参数结果。

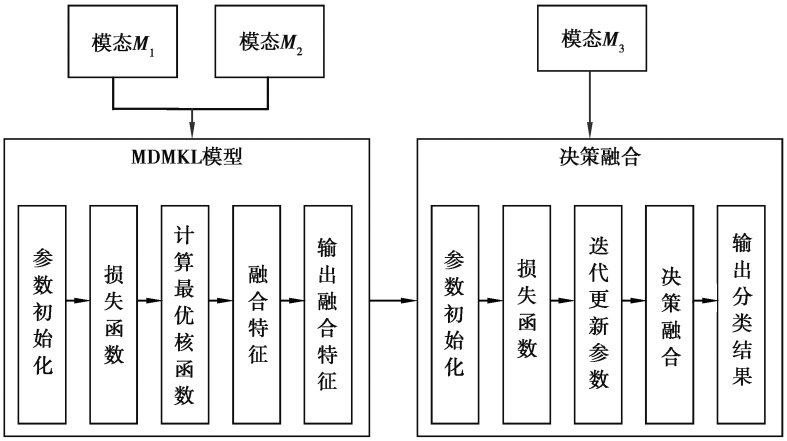


图5 MMMKL模型实现流程图

Fig. 5 The flow diagram of model MMMKL

3 实验分析

3.1 实验数据集

为了检测基于多模态的身份识别技术在智慧园区中应用的效率以及普遍性，首先选取了在中国模式识别与计算机视觉大会(PRCV2018)中，爱奇艺公开最大的明星视频数据集(IQY-VID)。该数据集被广泛用于“多模态视频任务识别挑战赛”，其中包含了4 934个人物，视频共有565 372条片段，并且被随机分为训练集219 677条，验证集172 860条，测试集172 835条，数据示例图如图6所示。同时也选择了真实智慧园区中曹家滩视频数据集，该数据集存储于智能监控系统之中，通过专业设备采集，其中包含个体的视频数据。采集的视频数据共有80 000条片段，其中随机选取了50%作为训练集，30%作为验证集，其余的为测试集。

图6 数据示例图

Fig. 6 Data sample graph

3.2 MDMKL模型实验

在MDMKL模型实验中，采用IQY-VID数据集以及视频中的文本信息。对于视频，获取图像模型并进行灰度变化获取图像模态矩阵，同时获取音频，采用重采样获取音频文件的关键特征；对于文本信息，通过计算词频获取特征向量。将身份识别的准确率作为实验的评价指标。在经过特征工程相关处理之后，使用MDMKL模型对图像、文本、音频特征进行特征融合。图7为不同融合算法进行多模态数据特征融合，并进行身份识别的实验结果对比。

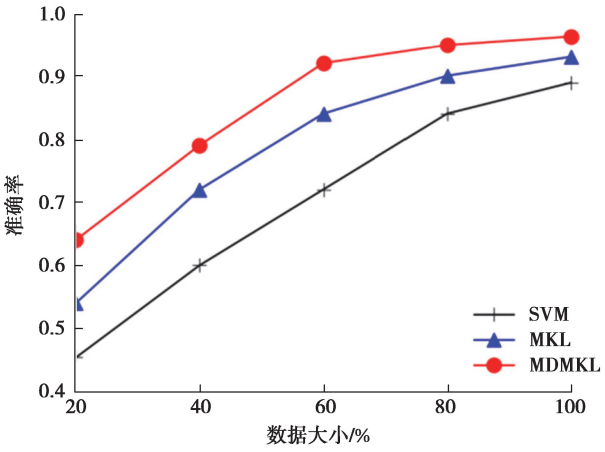


图 7　不同融合算法的实验对比图

Fig. 7　Experimental comparison of different fusion algorithms

从图 7 中可以看出，随着数据量的增加，所有的模型在融合之后都有准确率的提升，在数据量增长的初始阶段分类结果准确率提升幅度较大。对比传统的 SVM 和 MKL 融合算法，引入间隔的多核学习算法无论是最后的表现效果还是其准确率的增长效果都有更好的表现。具体而言，MDMKL 算法由于融合了多个模态而且能够提供区分性强的特征使得模型的准确率最高能达到 97.25%，而 SVM 和 MKL 融合算法表现最好的准确率分别为 88.90%和 94.34%。因此 MDMKL 模型要优于其他对比模型。

对 IQY-VID 数据集进行实验时，同时进行了算法的收敛性实验，即验证迭代次数的增加对损失函数 f 值下降的影响。在图 8 中，比较了传统的 MKL 方法与引入间隔约束后的收敛性能，其中红色线代表的 MDMKL 能够迅速达到收敛，而传统的 MKL 算法需要经过 7～8 次迭代才能达到收敛，说明 MDMKL 模型的收敛速度更快。

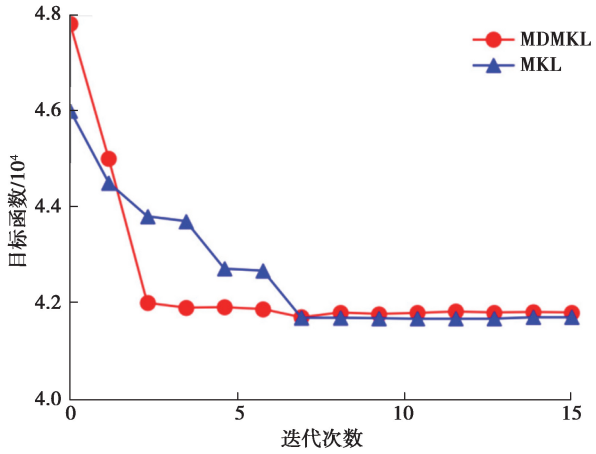


图 8　目标收敛性与迭代次数对比图

Fig. 8　Comparison of target convergence and iteration times

为了对比不同模态融合对最终结果的影响，表 1 展示了实验结果。对于单独一个模态的识别，采用 MDMKL 方法在不同单模态中提取特征之后得出的结果再做识别能够取得较好效果。在进行模态融合之后，识别的效果也都有了显著提高，其中表现最好的是 3 种模态进行特征组合之后的结果，能达到 97.25%的准确率。而基于图像与文本模态的融合也能达到较好的效果，这反映了图像与文本提供的特征对于一个人身份的识别产生的作用较大。可见，MDMKL 能够有效地提取不同模态的特征，并进行融合识别。

表1 单模态和多模特征级融合对比(曹家滩)

Table 1 Single mode and multi-modal fusion comparison (Caojiatan)

数据来源	准确率/%						
	音频A	图像P	文本T	P+T融合	P+A融合	A+T融合	P+T+A融合
文献[16]	74.22	76.38	79.77	85.46	83.69	84.12	88.60
CRMK[17]	79.14	94.50	74.49	95.75	83.85	95.38	96.12
MDMKL	76.30	95.25	81.67	96.97	91.84	96.56	97.25

为验证MDMKL模型的可移植性,采用曹家滩数据集进行实验验证,同样对于视频数据进行处理,对于文本数据采用员工登录系统的文本数据。将所有的特征向量进行特征维度上的合并,并使用MDMKL模型进行特征融合。表2为单模态与多模态特征融合对比的实验结果,同样也显示了在单一模态之中,图像模态由于能够区分身份的特征较多,能够有较高的准确率。而多模态融合识别结果表明图像模态与其他模态进行特征融合之后的结果能取得较好的效果。当同时使用图像、音频和文本的三种模态的特征并进行特征融合之后,能够达到较好的表现效果。在这个实验中,模型的表现效果比爱奇异视频提供的数据表现得更好,这可能是由于爱奇艺数据中的文本信息无法提供稳定的特征,同时图像数据受到视频的分辨率与是否有干扰等影响。

表2 单模态和多模特征级融合对比(爱奇艺)

Table 2 Single mode and multi-modal fusion comparison(IQY)

模态	准确率/%
音频模态A	59.91
文本模态T	72.11
图像模态P	75.83
P+T模态融合	81.22
P+A模态融合	76.53
A+T模态融合	76.31
P+T+A模态融合[16]	74.67
P+T+A模态融合(CRMKL)[18]	72.22
P+T+A模态融合(MDMKL)	84.46

3.3 多模态融合(MMMKL)实验

曹家滩样本数据包括图像P、音频A与T文本3种模态,采用特征融合与决策融合,3种模态的融合方式包括PT+A,PA+T以及AT+P。实验结果如图9所示。其中横坐标代表了在进行特征融合之后在进行决策融合的权重,其表示为$wM_{12}+(1-w)M_3$。可见,随着w的逐渐增大,模型识别准确率逐渐提升,并在权重值$w=0.6$附近时,各个模态融合的表现效果达到较好的效果。从图9中可以看出图像与音频信息进行特征融合之后,再与文本特征进行决策融合达到的识别准确度是最高的,在$w=0.6$时达到97.37%的准确率。这体现了图像信息中无法识别的信息,可以通过音频信息补全,并通过文本特征进一步确认。

在视频与文本进行特征融合识别之后,再进行音频特征识别结果[19]的决策融合中,其表现效果在$w=0.7$时达到最好,准确率为92.5%。这可能是由于决策融合对于整体框架中无法达到较好的效果,获取的音频特征存在噪声数据,影响了整体表现效果。

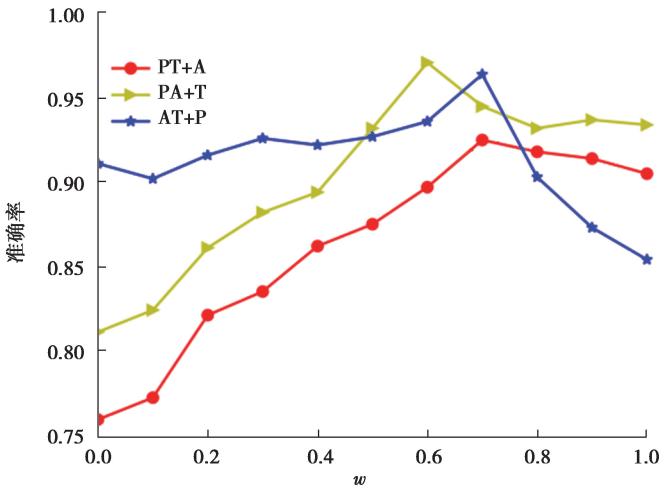


图 9　曹家滩视频数据集的多模态实验结果

Fig. 9　Multimodal experiment results based on Caojiatan dataset

在对本文所提出的模型框架的实验中，表 3 显示了曹家滩智慧园区历史视频数据的多模态特征融合与决策融合实验结果，在该数据下首先对图像特征与音频特征进行特征融合，再对文本信息进行决策融合达到了最好的表现效果。表 4 显示了爱奇艺视频数据集的特征融合与决策融合实验结果，与曹家滩视频数据集实验结果类似，当使用图像与音频模态作为特征融合时，模型的准确率能到达最高水平。

表 3　基于曹家滩视频集数据的多模态融合

Table 3　Hybrid fusion based on Caojiatan

融合顺序	准确率/%
图像→文本→音频	96.47
图像→音频→文本	97.17
文本→音频→图像	92.45

表 4　基于爱奇艺数据的多模态融合

Table 4　Hybrid fusion based on IQY

融合顺序	准确率/%
图像→文本→音频	79.24
图像→音频→文本	85.23
文本→音频→图像	77.50

4　结　语

智慧园区离不开新型技术的支持，针对于曹家滩智慧园区的技术发展，提出了一种基于多模态的身份识别技术，能有效地解决智慧园区中对于不同园区的分级管理问题。由于现有的人脸识别只是单纯考虑到图像这一种模态对于身份识别的影响，基于图像、音频、文本 3 种模态提出了一种 MDMKL 模型，有效地提升了身份识别的效率与准确性，提高了受监控区域的安全性。同时由于采用的是非接触的信息采集方式，设备本身安装方便、性能可靠，能够显著提升园区的管控水平和事件处理速度。在园区的智慧管理建设过程中，做出针对性的技术升级和创新，提高园区信息化技术水平，也充分利用技术创新带动产业创新，打开了园区智慧化管理新局面。

参考文献：

[1] 王瑞江，朱晓霞，张莉莉，等. 面向新材料产业园的智慧管理系统研究[J]. 科技风，2020(18)：159-160，162.
Wang R J，Zhu X X，Zhang L L，et al. Research on IMS for new material industry park[J]. Technology Wind，2020(18)：159-160，162.(in Chinese)

[2] 许斌，苏家兴，郭栋，等. 智慧园区信息基础设施规划思路研究[J]. 城市住宅，2020，27(3)：129-130.
Xu B，Su J X，Guo D，et al. Research on planning ideas of information infrastructure in smart parks [J]. City & House，

2020，27(3)：129-130.(in Chinese)

[3] 谢易辰. 智慧园区建设方案略谈[J]. 技术与市场，2019，26(12)：65-67.
Xie Y C. Brief discussion on the construction plan of smart parks[J]. Technology and Market，2019，26(12)：65-67.(in Chinese)

[4] 邓雄，王洪春，赵立军，等. 人脸识别活体检测研究方法综述[J]. 计算机应用研究，2020，37(9)：2579-2585.
Deng X，Wang H C，Zhao L J，et al. Survey on face anti-spoofing in face recognition[J]. Application Research of Computers，2020，37(9)：2579-2585.(in Chinese)

[5] 刘亮. 基于PCA和LDA改进算法的人脸识别技术研究[J]. 无线互联科技，2019，16(17)：110-111.
Liu L. Research on face recognition technology based on PCA and LDA improved algorithm[J]. Wireless Internet Technology，2019，16(17)：110-111.(in Chinese)

[6] 杨巨成，刘娜，房珊珊，等. 基于深度学习的人脸识别方法研究综述[J]. 天津科技大学学报，2016，31(6)：1-10.
Yang J C，Liu N，Fang S S，et al. Review of face recognition methods based on deep learning[J]. Journal of Tianjin University of Science & Technology，2016，31(6)：1-10.(in Chinese)

[7] Hall D L，Llinas J. An introduction to multisensor data fusion[J]. Proceedings of the IEEE，1997，85(1)：6-23.

[8] 刘奕. 多模生物特征融合关键技术研究[D]. 济南：山东大学，2017.
Liu Y. The research of key technologies based on multi-model biometric feature fusion[D]. Jinan：Shandong University，2017.(in Chinese)

[9] 陈振浚，张小凤，方宇杰，等. 基于深度传感的人脸识别算法研究与实现[J]. 自动化应用，2020(8)：80-82.
Chen Z J，Zhang X F，Fang Y J，et al. Research and implementation of face recognition algorithm based on depth sensing[J]. Automation Application，2020(8)：80-82.(in Chinese)

[10] Fang Y K，Deng W H，Du J P，et al. Identity-aware CycleGAN for face photo-sketch synthesis and recognition[J]. Pattern Recognition，2020，102：107249.

[11] Sheng B，Li P，Gao C H，et al. Deep neural representation guided face sketch synthesis[J]. IEEE Transactions on Visualization and Computer Graphics，2019，25(12)：3216-3230.

[12] Mullah H U，Deka B，Prasad A V V. Fast multi-spectral image super-resolution via sparse representation[J]. IET Image Processing，2020，14(12)：2833-2844.

[13] 霍静. 面向异构人脸识别的跨模态度量学习研究[D]. 南京：南京大学，2017.
Huo J. Cross-modal metric learning for heterogeneous face recognition[D]. Nanjing：Nanjing University，2017.(in Chinese)

[14] Setumin S，Suandi S A. Difference of Gaussian oriented gradient histogram for face sketch to photo matching[J]. IEEE Access，2018，6：39344-39352.

[15] 吴凯亮. 人脸识别在智慧园区中的应用[J]. 电视技术，2019，43(21)：70-74.
Wu K L. Application of face recognition in smart park[J]. Video Engineering，2019，43(21)：70-74.(in Chinese)

[16] Yuille A L，Hallinan P W，Cohen D S. Feature extraction from faces using deformable templates[J]. International Journal of Computer Vision，1992，8(2)：99-111.

[17] Poria S，Chaturvedi I，Cambria E，et al. Convolutional MKL based multimodal emotion recognition and sentiment analysis[C]// 2016 IEEE 16th International Conference on Data Mining (ICDM)，December 12-15，2016，Barcelona，Spain. IEEE，2016：439-448.

[18] Liu Y L，Shi P P，Peng B，et al. iQIYI celebrity video identification challenge[C]// Proceedings of the 27th ACM International Conference on Multimedia，Nice，France. New York，NY，USA：ACM，2019：2516-2520.

[19] Nagrani A，Chung J S，Zisserman A. VoxCeleb：a large-scale speaker identification dataset[C]// Interspeech 2017，August 20-24，2017，Stockholm，Sweden. ISCA，2017：2616-2620.

（编辑 郑 洁）