

# 基于深度学习的盲文自动识别系统

李荣瑞 施 霖

(昆明理工大学 信息工程与自动化学院, 云南 昆明 650500)

**摘 要** 传统盲文识别采用手工特征点比对的方法, 只能识别特定条件下采集的盲文图片, 系统的实用性很低。文中构建卷积神经网络自动学习盲文特征, 消除了手工提取特征这一弊端; 并提出一种基于灰度投影的方法对盲文图片进行校正和分割。结果表明, 系统对各类盲文测试集的总体识别率达到 98.62%, 在传统盲文识别方法无法工作的多种噪声条件下均可取得较高的识别率, 大幅提高了盲文识别系统的准确率和实用性。

**关键词** 深度学习; 盲文识别; 灰度投影; 卷积神经网络; 特征学习

中图分类号 TP391 文献标识码 A 文章编号 1007 - 7820(2018)09 - 045 - 05

## Braille Automatic Recognition System Based on Deep Learning

LI Rongrui, SHI Lin

(School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

**Abstract** The traditional braille recognition adopts the method of manual feature point matching and it can only identify braille images collected under certain conditions, the system is of low utility. In this paper, convolution neural network is constructed to learn the braille feature automatically and eliminate the drawbacks of manual extraction; a method based on gray projection is proposed to correct and segment the braille images. The results shows that the overall recognition rate of the system for all kinds of braille test sets is 98.62%, the traditional braille recognition method can not work under a variety of noise conditions, but with this method in the paper, the system can achieve very high recognition rate and it highly improved the accuracy and practicability of braille recognition system.

**Keywords** deep learning; braille recognition; gray projection; convolution neural network; feature extraction

盲文识别研究对于盲人特殊教育事业和生活具有重要的意义<sup>[1-2]</sup>。近年来, 国内外学者在盲文识别方面做了大量研究, 国外 Yasuhiko Ogawat 等<sup>[3-5]</sup>进行了基于数字图像处理的盲文识别等研究; 国内的周小燕、周春耕等学者<sup>[6-9]</sup>做了许多盲文汉语转换方面的工作。但是这些研究中还存在问题: (1) 使用条件要求严苛, 光照、角度的轻微变化会严重影响识别准确率; (2) 图片预处理过程复杂, 步骤繁琐; (3) 特征点由研究人员手工提取, 费时费力且无法保证准确性。

自 2006 年以来, 深度学习作为机器学习领域中对图像等进行特征学习的一种新方法, 在图像识别方面取得了巨大成功<sup>[10]</sup>。卷积神经网络 (Convolutional Neural Network, CNN) 是深度学习中广为应用的经典

模型, 例如 Lenet-5<sup>[11]</sup>, Cifar-10<sup>[12]</sup> 等模型, 其强大的特征学习能力使得图像识别研究取得了突破性的进展。

本文利用灰度投影方法来自动校正和分割盲文; 构建卷积神经网络 (CNN) 模型, 利用大量叠加噪声的盲文图片训练网络得到的 CNN 模型连接 Softmax 分类器作为盲文识别部分。最终的盲文自动识别系统具有识别率高、抗干扰性强等特点, 在传统盲文识别研究中无法工作的多种环境下, 系统可良好工作, 突破传统盲文识别系统对环境的严苛要求。

## 1 盲文图像自动校正与分割

### 1.1 二值图像处理

作盲文图像的灰度投影图要先将盲文图像转换为二值图像。盲文图像盲点与背景的灰度差异很大且灰度值比较集中, 这有利于通过阈值分割方法进行二值处理, 图 1 分别为盲文图像和利用大津法 (OTSU 法)<sup>[13]</sup>得到的此盲文图像的二值图像。

收稿日期: 2017-08-31

基金项目: 国家自然科学基金 (61368005)

作者简介: 李荣瑞 (1992 -), 男, 硕士研究生。研究方向: 图像处理, 深度学习。施霖 (1972 -), 男, 博士, 副教授。研究方向: 计算机视觉, 计算机图形学。

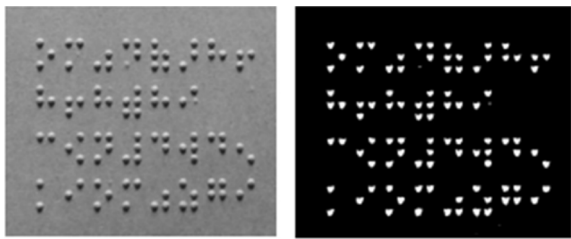


图1 盲文图像二值化处理

## 1.2 图像校正

如图2所示,将采集的盲文图像缩放为边长为10 cm的正方形图像,放置在建立的直角坐标系中,使图像4个端点分别落在坐标点(5,5)、(5,15)、(15,15)、(15,5)上,图像中心点在坐标(10,10)上。将 $x$ 轴上投影的采集范围定为(0,0)到(20,0)区间,将 $y$ 轴的采集范围定为(0,0)到(0,20),这样可保证图片在旋转不同角度时,其二值投影图像都被完整采集。

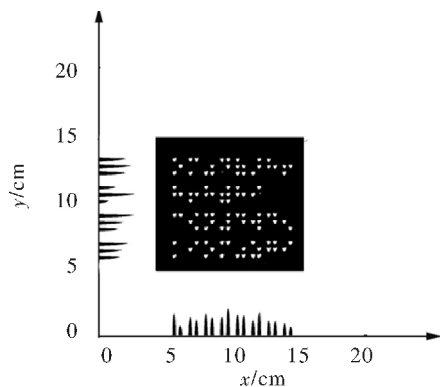


图2 二值图像坐标投影

设定顺时针方向为正方向,将盲文偏转角度依次为 $\alpha = 0^\circ, 6^\circ, 12^\circ, 18^\circ, 24^\circ$ 。将其二值图像向 $x$ 轴投影得到竖直灰度投影图,图3为以 $x$ 轴为底、高度为2的同一区域的灰度对比图。结果显示,盲文方向偏转角度越大,投影图中灰度条越集中于底部。

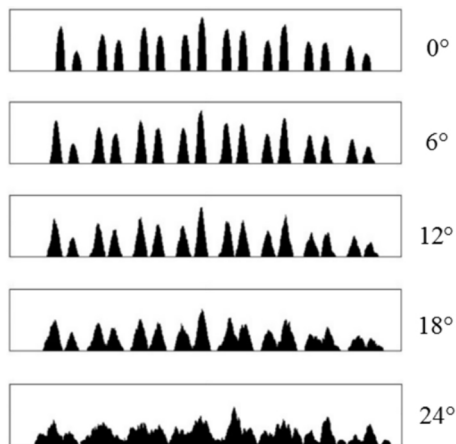


图3 灰度投影对比图

灰度条的集中程度利用如下方法测量:在竖直灰

度投影图中,将 $x$ 轴、 $y$ 轴、直线 $x=20$ 、切线 $y=0.5$ 四条直线围住的矩形部分定义为 $T$ 区域,其中(0,0.5)坐标点的选择是根据50张不同的盲文图像确定的最佳分割点。假设 $T$ 区域内像素点总数为 $M$ ,灰度为0(黑色部分)的像素点数量为 $M_0$ ,则 $T$ 区域内灰度条像素点所占比例 $p$ 的值为

$$p = \frac{M_0}{M} \quad (1)$$

对于同一幅盲文图片偏转不同角度的数张灰度投影图, $T$ 区域大小相同,像素点总数 $M$ 不变。偏转角度越大, $T$ 区域内黑色像素点数量 $M_0$ 越大,从而 $p$ 越大。

盲文图像的角度校正过程如图4所示:将原始盲文图片的二值图片偏转不同角度 $\alpha$ 并分别向 $x$ 轴投影,计算其 $T$ 区域内黑色占比 $p$ 。选中其中最小的值 $p_x$ ,此值对应的旋转角度 $\alpha_x$ 即为校正原始盲文图片所需要旋转的角度。校正后的图片中盲文角度误差小于 $\alpha$ 步长的 $1/2$ 。若要求更高的校正精度,可减小 $\alpha$ 每次增加的步长。此处取偏转角度最大为 $30^\circ$ 是因为正常盲文图片的采集偏转角度不会超出此范围。

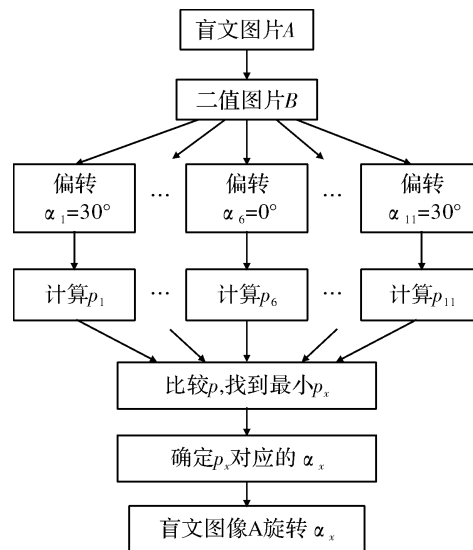


图4 图像校正流程图

取50张盲文图片对此校正方法进行实验,结果显示50张盲文图像的校正成功率为100%。

## 1.3 盲文图像分割

在识别盲文古籍、作业这种含有大量盲文字符的图片时,经过角度校正后,其灰度图投影在 $x$ 轴、 $y$ 轴上形成数条相互独立的灰度条。在竖直投影图中,灰度条是由同一列盲点在 $x$ 轴的灰度投影叠加而成,即灰度条在 $x$ 轴上的坐标为这一列盲点的 $x$ 坐标;同理,在水平投影图中,灰度条是由同一行盲点在 $y$ 轴的灰度投影叠加而成,灰度条在 $y$ 轴上的坐标即为这一行盲点的 $y$ 坐标。盲文的盲方为 $3 \times 2$ 的矩形点阵,一行

盲文在  $y$  轴的投影为 3 条灰度条, 一列盲文在  $x$  轴的投影为 2 条灰度条。根据这种特点, 在  $y$  轴上每隔两条灰度条做一条平行分割线, 在  $x$  轴上每隔两条灰度条做一条竖直分割线, 横竖分割线共同作用可将盲文图片分割。

## 2 盲文识别部分

### 2.1 数据集的设计与制作

由于没有完整的盲文数据库, 实验所需要的数据集从实际采集的盲文图片中截取盲文字符制作完成。本次实验数据集包含 63 类盲文字符和空白盲文图片一类共 64 类, 每类盲文字符包含 20 张图片, 此时数据集图片数量为:  $64 \times 20 = 1280$  张。考虑到实际使用中的几种影响因素: 光照变化、图像倾斜、盲点偏移等, 为增强识别系统的鲁棒性和抗干扰性, 将这些影响因素加入到数据集中。对每张图片作添加噪声处理 9 次, 对应 9 种图像变化处理: 光照强度 +15, +30, -15, -30 共 4 种, 角度左倾  $6^\circ$ , 右倾  $6^\circ$  两种, 字符平移缩放 3 种。图 5 为一张盲文图片经过噪声处理形成的 10 张图片。此时, 1260 张原始采集盲文字符图片经过 9 种噪声处理, 数据集图像数量变为:  $1280 + 1280 \times 9 = 12800$  张, 每类盲文字符包含图片 200 张。数据集中 80% 作为训练集, 20% 作为测试集。

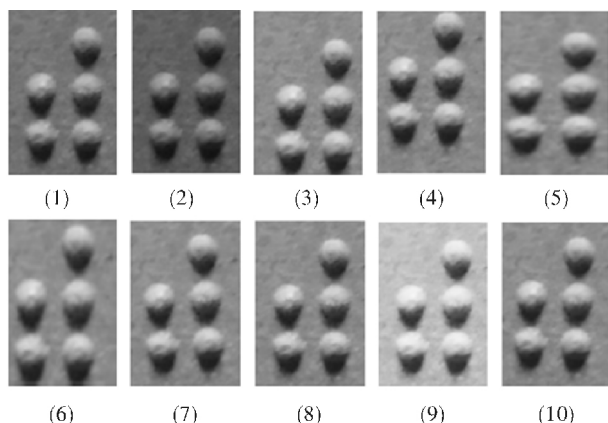


图 5 噪声处理图片示例

### 2.2 卷积神经网络的设计

CNN 的运算包含前向传播和反向传播<sup>[14]</sup>。在前向传播中, 上一层的输出作为这一层的输入, 整个过程通过激活函数逐层传递。反向传播过程中通过计算实际输出与理想输出的均方误差, 对权值参数进行优化。信息正向传播和误差反向传播交替进行, 直到满足要求或达到最大迭代次数<sup>[15]</sup>。CNN 结构过大或过小过拟合或欠拟合, 数据集太小或不合理会导致网络模型失去泛化能力, 初始权值选取不合理会导致训练陷入局部最优解<sup>[16]</sup>。

由于盲文字符相对比较简单, CNN 中卷积层的层数初步定为 2~3 层, 每一层卷积层后面接下采样层, 下采样层可降低特征图的维度, 减少运算参数, 提高运算时间; 另一个作用是增强系统平移、伸缩不变性。取盲文图片对两种卷积层模型进行测试, 结果显示, 两层卷积层模型相比于 3 层卷积层模型, 运算时间减少 35%, 运算数据量减少 43%, 而识别精度仅减少 1%。综合运算速度、内存和准确率等因素, 文中将卷积层数目定为 2 层, 每一层卷积层后连接一层下采样层以减少运算参数。

全连接层为两层, 试验不同节点数量 CNN 模型的识别效果, 当第一层全连接层节点数为 300, 第二层全连接层节点数为 64 时, CNN 模型的识别准确率最高。本文实验构建了多种 CNN 模型, 试验了不同层数和节点数的组合, 在保证识别准确率的同时考虑识别速度等因素, 最终确定的 CNN 模型结构如图 6 所示, 网络结构包括: 输入层, 2 层卷积层 (Conv), 2 层下采样层 (Pool), 1 层激活层 (Relu), 2 层全连接层 (Ip)。CNN 模型末端接一层 Softmax 分类层, 右侧数字表示一张高、宽为 28 像素的三通道图片在网络中经过各层运算后的数据变化过程。

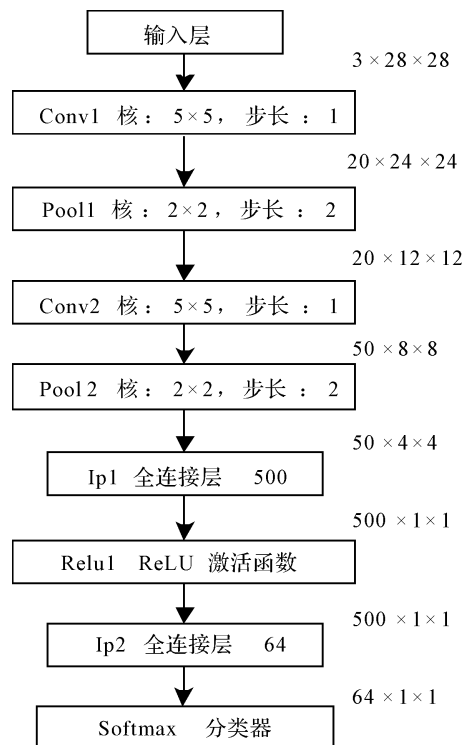


图 6 卷积神经网络结构

常见的激活函数有 sigmoid 函数, tanh 函数, ReLU 函数, 此处选择稀疏性更好的 ReLU 函数<sup>[17]</sup>。ReLU 的算式为

$$f(x) = \max(0, x) \quad (2)$$

若通过卷积计算得到的值大于 0, 则此值保持不

变;若卷积计算的值小于0,则将此值赋值为0。

本次实验共有64类盲文字符,因此分类层我们选择适合解决多分类问题的Softmax函数。假设输入Softmax函数 $\zeta$ 的向量 $z$ 的维度为 $c$ ,那么Softmax函数的数据也是一个 $c$ 维度的向量 $y$ ,里面的值为1或0。Softmax函数是一个归一化的指数函数,定义为

$$y_c = \zeta(z)_c = \frac{e^{z_c}}{\sum_{d=1}^c e^{z_d}} \quad c = 1, 2, \dots, C \quad (3)$$

其中分母起到了正则项的作用,使得

$$\sum_{c=1}^c y_c = 1 \quad (4)$$

Softmax函数放在神经网络的末端起着分类的作用,用 $c$ 个神经元可以表示函数中的值。对于给定的输入 $z$ ,可以得到 $t = c, c = 1, 2, \dots, C$ 中每个分类的概率为

$$\begin{bmatrix} P(t=1|z) \\ \vdots \\ P(t=C|z) \end{bmatrix} = \begin{bmatrix} \zeta(z)_1 \\ \vdots \\ \zeta(z)_c \end{bmatrix} = \frac{1}{\sum_{d=1}^c e^{z_d}} \begin{bmatrix} e^{z_1} \\ \vdots \\ e^{z_c} \end{bmatrix} \quad (5)$$

其中 $P(t=C|z)$ 为给定输入 $z$ 时,该输入数据是 $C$ 分类的概率。

### 2.3 训练网络模型

将数据集的盲文图片导入构建的深度卷积神经网络中,设定网络训练迭代次数为1000次,图7为训练过程中准确率随迭代次数增加的变化曲线图。

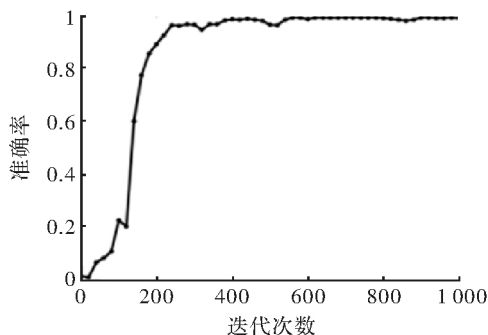


图7 准确率曲线

如图所示,CNN模型经过迭代训练逐渐收敛,网络误差减小,准确率升高。迭代训练1000次后,CNN模型对测试集的识别准确率为99.41%,这表明构建的卷积神经网络通过1000次的迭代训练已经从数据集中学习到各类盲文字符的深层特征并保存在CNN模型中。

### 3 实验结果

在标准采集环境和多种影响环境下制作盲文字符图片共2928张,输入CNN模型进行识别,总体识别率为98.57%,不同噪声影响下盲文识别率如表1所示。

表1 识别结果

噪声种类	图片数量	准确率
标准环境	568	100%
光照	672	100%
平移缩放	384	96.88%
角度	512	99.61%
多种噪声叠加	792	96.46%

实验最后检测构建的神经网络模型对于实际应用中盲文识别的准确率,在日常使用环境下取50张不同语句盲文图片对本文研究的基于深度学习的盲文自动识别系统进行从输入盲文图片到输出汉语拼音的整体测试。50张盲文图片中盲文字符总数(包含空白字符)为2746个,每张图片包含的盲文字符数量(包含空白字符)在24~120个之间,经实验统计,系统成功识别其中2708个,自动识别系统对盲文图片中字符的识别成功率为98.62%。

相比于传统盲文识别方法,本文方法具有以下优势:(1) 本文方法避免了传统方法中繁琐的预处理过程;(2) 传统方法需要手工定义特征点,费时费力且不能保证特征的有效性,本文方法通过CNN自动进行训练,得到的模型更能刻画盲文的本质特征;(3) 传统方法只用来识别特定光照和角度的盲文图片,且识别率不足96%,本文方法将此识别率提升为100%;(4) 传统方法无法对平移、伸缩、角度倾斜等噪声条件下的盲文图片进行识别,而本文方法对此类噪声图片的识别率高达98.2%。

### 4 结束语

本文研究的基于深度学习的盲文自动识别系统相比于传统的盲文识别研究,突破了传统方法中对使用环境的限制,在传统盲文识别装置不能正常工作的多种环境下仍能达到很高识别率,提高了盲文识别系统的适用性、实用性和准确率。文中的识别系统准确率达到98.62%,若要进一步提高系统的识别准确率,可通过两种方法进行提升:(1) 减小图像校正部分的旋转角度步长 $\alpha$ ;(2) 丰富识别部分训练集的噪声变化图片,增加训练集图片数量,增加训练迭代次数。

### 参考文献

- [1] 尹佳. 纸介盲文自动识别系统关键技术研究[D]. 长春: 长春理工大学, 2011.
- [2] 尹佳,李杰,王丽荣. 盲文自动识别方法研究[J]. 长春大学学报, 2010(8): 54-56.
- [3] Ogawa Y, Kikuchi Y, Ohtake N. Braille reader: old Braille text for young [C]. New York: International Conference on Computers for Handicapped Persons, Springer-Verlag, 1994.

- [4] Antonacopoulos A, Bridson D. A robust braille recognition system [C]. Florence: Document Analysis Systems VI, International Workshop, Proceedings DBLP 2004.
- [5] Shahbazkia H R, Silva T T, Rui M G. Automatic braille code translation system [C]. Berlin: Progress in Pattern Recognition, Image Analysis and Applications, Springer 2005.
- [6] Jiang M, Zhu X, Gielen G, et al. Braille to print translations for Chinese [J]. Information & Software Technology 2002, 44 (2): 91-100.
- [7] 江铭虎, 朱小燕. 基于多种知识的盲文翻译的研究 [J]. 清华大学学报: 自然科学版 2000, 40(9): 69-73.
- [8] 庄丽, 包塔, 朱小燕. 盲人用计算机软件系统中的语音和自然语言处理技术 [C]. 武汉: 全国人机语音通讯学术会议 2003.
- [9] 周春耕, 张秉权, 黄河燕. 基于混合语言模型的盲文机器翻译系统的研究与实现 [J]. 计算机工程与应用 2003, 39 (4): 127-130.
- [10] 余凯, 贾磊, 陈雨强, 等. 深度学习的昨天、今天和明天 [J]. 计算机研究与发展 2013, 50(9): 1799-1804.
- [11] LéCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [12] Krizhevsky A. Convolutional deep belief networks on CIFAR-10 [J]. IEEE Transactions on Image 2010(9): 86-89.
- [13] Otsu N. A threshold selection method from Gray-level [J]. IEEE Transactions on Systems Man & Cybernetics, 1979, 9 (1): 62-66.
- [14] 李斯凡, 高法钦. 基于卷积神经网络的手写数字识别 [J]. 浙江理工大学学报: 自然科学版 2017, 37(3): 438-443.
- [15] 高原. 基于 BP 神经网络的文本验证码破解 [J]. 电子科技 2012, 25(7): 37-42.
- [16] 王闪闪. 基于群智能算法的神经网络建模研究 [J]. 电子科技 2017, 30(4): 56-59.
- [17] 曾维亮, 林志贤, 陈永洒. 基于卷积神经网络智能冰箱果蔬图像识别的研究 [J]. 微型机与应用 2017, 36(8): 56-59.

(上接第 44 页)

目标实时转动效果。

表 1 视频图像跟踪测试

帧数	时间/s
0~55	1.83
55~96	1.36
96~135	1.3

在视频跟踪系统中对跟踪目标进行定位和跟踪时,当目标物体的形态产生变化,将无法准确跟踪到目标。当跟踪目标被遮挡,形心法将无法准确计算目标形心坐标,因此,本文所实现的目标定位、跟踪方法适用于目标物形态不变情况下的视频跟踪。

## 4 结束语

本文研究了机器视觉技术下的 FPGA 视频跟踪问题,设计了 FPGA 的实时视频采集显示系统,研究实现了实时视频目标检测、定位及跟踪算法。根据跟踪目标参数及运动轨迹,计算云台步进电机控制脉冲信号,并通过 FPGA 输出该控制信号,云台电机受到控制信号驱动后带动摄像头转动。将视频跟踪系统中的各个子系统在 FPGA 平台上进行了集成与测试。测试结果表明,所设计的 FPGA 平台视频跟踪系统在实时性与准确性等方面都有较好的表现,与传统视频跟踪系统相比,该系统具有实时刷新率高、精确跟踪目标等优势。但本文所提出的算法仅适用于目标形态不变情况下的视频跟踪,可从运动补偿角度考虑。

## 参考文献

- [1] 龚声民. 运动目标识别与跟踪系统研究 [D]. 哈尔滨: 哈尔滨工业大学 2014.
- [2] 冯亮. 机器视觉技术的视频跟踪系统研究及 FPGA 实现

[D]. 贵阳: 贵州大学 2015.

- [3] 王四伟. 基于 SOPC 的指纹采集与处理系统的研究 [D]. 武汉: 武汉理工大学 2009.
- [4] Li Kaiyu, Zhang Huanchun, Jing Yazhi. FPGA based fast and high-quality rotation of image [J]. Journal of Image and Graphics 2004, 9(3): 285-289.
- [5] 刘紫燕, 冯亮, 祁佳. 一种基于 FPGA 的实时视频跟踪系统硬件平台设计 [J]. 传感器与微系统 2014, 33(7): 98-102.
- [6] Yuan Pao Hsu, Hsiao Chun Miao, Ching Chih Tsai. FPGA implementation of a real-time image tracking system [C]. Korea: SICE Annual Conference 2010.
- [7] Wu-Chih Hu. Adaptive template block-based block matching for object tracking [J]. IEEE Transactions on Intelligent Systems Design and Applications 2008(1): 61-64.
- [8] 吴长江, 赵不赓, 郑博, 等. 基于 FPGA 的动态目标跟踪系统设计 [J]. 电子技术应用 2010, 36(3): 45-50.
- [9] 高仕博, 程咏梅, 肖利平, 等. 面向目标检测的稀疏表示方法研究进展 [J]. 电子学报 2015, 43(2): 320-332.
- [10] 袁国武. 智能视频监控中的运动目标检测和跟踪算法研究 [D]. 昆明: 云南大学 2012.
- [11] 李亮, 罗毅. 帧间差分法在视频监控中的应用研究 [J]. 四川理工学院学报: 自科版 2015, 28(6): 58-62.
- [12] 沈瑜, 王新新. 基于背景减法和帧间差分法的视频运动目标检测方法 [J]. 自动化与仪器仪表 2017(4): 122-124.
- [13] Sharifi M, Fathy M, Tayefeh Mahmoudi M. A classified and comparative study of edge detection algorithms [C]. Iran: Proceedings of the International Conference on Information Technology: Coding and Computing 2002.
- [14] 张丽红, 凌朝东. 基于 FPGA 的 Sobel 边缘检测应用 [J]. 电子技术应用 2011, 37(8): 26-29.
- [15] 孙涛. 基于数学形态学和小波变换的图像边缘检测的研究 [D]. 长沙: 中南大学 2010.