

# 大视场域的目标检测与识别算法综述

李唐薇<sup>1,2</sup>, 童官军<sup>1\*</sup>, 李宝清<sup>1</sup>, 卢晓洋<sup>1</sup>

<sup>1</sup>中国科学院上海微系统与信息技术研究所微系统技术重点实验室, 上海 201800;

<sup>2</sup>中国科学院大学, 北京 100049

**摘要** 目标精确感知与识别为信息化战争提供了一个重要的技术增长点,全景视觉传感设备因其拥有大视场(LFOV)范围优势而逐渐被应用于安防及军事领域中的目标检测与识别任务。首先从相机成像模型、图像成像质量以及目标物体的非对称性三个方面对存在的困难以及挑战进行阐述。基于是否进行畸变校正预处理,将近年来LFOV域的目标检测与识别算法分为基于畸变校正的目标检测与识别算法和基于原始LFOV图像的目标检测与识别算法两类,并针对这两类算法进行了全面梳理和总结,对当前LFOV域的目标检测与识别各类算法的统一性和差异性进行思考分析,探讨其未来发展趋势。

**关键词** 图像处理; 目标检测与识别; 大视场图像; 畸变校正

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP57.120002

## Review on Object Detection and Recognition in Large Field of View

Li Tangwei<sup>1,2</sup>, Tong Guanjun<sup>1\*</sup>, Li Baoqing<sup>1</sup>, Lu Xiaoyang<sup>1</sup>

<sup>1</sup>Key Laboratory of Science and Technology on Micro-System, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 201800, China;

<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract** Precise object detection and recognition plays an important role in information-based warfare. Due to panoramic vision sensors' large field of view (LFOV), they are gradually applied to security and military areas. In this paper, first, the difficulties and challenges in the development of object detection and recognition in LFOV are presented from three aspects: camera imaging model, image imaging quality, and asymmetry of object. Then, based on whether the distortion correction preprocessing is carried out or not, the object detection and recognition algorithms in LFOV are classified into two categories: distortion correction based algorithms and original LFOV image based algorithms. These two kinds of algorithms are comprehensively combed and summarized. Finally, the paper analyzes the unity and difference of various algorithms for object detection and recognition in LFOV and discusses their future development trend.

**Key words** image processing; object detection and recognition; large field of view images; distortion correction

**OCIS codes** 100.2000; 100.3008; 040.1880; 150.1488

## 1 引言

图像目标检测与识别技术通常可在视觉传感器获取的图像中找到所需要检测与识别的目标,获得其存在的空间位置信息,标识出其所覆盖的范围,同时给出目标类别<sup>[1-2]</sup>。相比于红外传感设备、雷达等,视觉传感器具有获取信息丰富、采样周期短、受

磁场或传感器相互间干扰影响较小、重量轻、能耗小、使用方便等优势,因此在军事侦察、反恐处突等领域中备受青睐。然而现阶段市面上占有量较高的传统视觉传感器(如红外视频图像传感器、可见光视频图像传感器)所获得图像的可视范围有限,在90°左右,在应用过程中具有极大的局限性。全景视觉传感器则较好地弥补了这一缺陷,其获取图像信息

收稿日期: 2019-07-19; 修回日期: 2019-10-10; 录用日期: 2019-10-31

基金项目: 国家重点研发计划公共安全专项

\* E-mail: tongguanjun@mail.sim.ac.cn

的视野范围可以达到  $360^\circ$ , 更甚者到达  $720^\circ$ , 在大视场(LFOV)应用领域展现出得天独厚的优势。另一方面, 全景视觉传感器自身的成像特点造成目标在图像中存在扭曲失真、形变模型难以用简单的数学模型构建<sup>[3-5]</sup>等问题, 这给大视场域的目标检测与识别增加了新的挑战。

本文对目前基于大视场域的目标检测与识别算法进行研究, 首先从相机成像模型、图像成像质量以及目标物体的非对称性三个方面对大视场域的目标检测与识别研究过程中存在的难点和挑战进行阐述。然后以是否需要畸变校正作为预处理手段对目前大视场域的目标检测与识别算法进行分类梳理总结。在此基础上, 从各类算法之间的统一性与差异性两个方面对大视场域的目标检测与识别算法进行思考, 对其未来发展趋势进行进一步讨论。

## 2 大视场域的目标检测与识别难点与挑战

目标检测与识别是计算机视觉中的基本问题之

一, 而大视场域的目标检测与识别是该领域中一个极具挑战性的分支, 其难度远大于常规的目标检测与识别。虽然目前已经诞生了许多效果非常好的方法和模型<sup>[6-8]</sup>, 但直接将其迁移到 LFOV 图像下的目标检测与识别任务中, 效果并不理想。除常规的目标检测与识别的难点与挑战外<sup>[9]</sup>, 还存在如下三个方面的难点与挑战。

### 1) 相机成像模型难以准确建模

与常规相机的线性模型相比, 全景相机通过引入透镜<sup>[10]</sup>扩大视场范围, 因此常用非线性模型来描述。在实际研究应用中, 全景相机模型通常只考虑切向畸变和径向畸变两种透镜畸变。切向畸变主要源自制作工艺上的缺陷, 如在制作工艺上透镜本身与相机传感器平面(成像平面)或图像平面不平行放置。径向畸变则来自于透镜的形状, 主要包括桶形畸变和枕形畸变两种。畸变过程在图像中表现为像素点空间变换, 同类型的不同镜头畸变情况不同, 如图 1 所示。因而这类非一致畸变难以用统一的数学模型进行建模。



图 1 目标的非一致畸变

Fig. 1 Inconsistent distortion of object

### 2) 图像质量下降

虽然 LFOV 图像具有视场大的优势, 但其焦距较短, 呈现的视场范围越大, 所需的焦距就越小, 这

使得距离镜头较远处目标的成像质量下降, 导致目标存在残缺、模糊等现象, 如图 2 所示。



图 2 目标在 LFOV 图像中的残缺、模糊情况

Fig. 2 Incomplete and blurring object in LFOV images

对于小目标而言, 根据国际光学工程学会(SPIE)对小目标的定义<sup>[11]</sup>, 即对于一幅  $256 \text{ pixel} \times 256 \text{ pixel}$  的图像, 目标尺寸所占像素不超过整幅图像的  $0.12\%$  (约  $80 \text{ pixel}$ ), 此类目标

(图 3)在 LFOV 图像成像过程中部分信息可能会因位置挤压而造成像素点过少, 使得目标特征过少甚至丢失。

在实际成像过程中, 图像中的目标区域还会受



图3 小目标在 LFOV 图像中的失真情况

Fig. 3 Distortion of small object in LFOV images

到噪声的影响,上述这些因素对检测与识别任务来说均较为敏感。除此之外,在非受控自然场景中,背景的复杂性会干扰目标物体,这对目标检测与识别算法提出更高要求。

### 3) 目标物体的非对称性

目标物体在不同位置的畸变表现为图像中像素

点在空间尺度上的不规则分布。根据透镜畸变程度分布,在其原始位置的一定范围内,光轴中心的畸变为 0,畸变沿着镜头半径方向从中心到边缘越来越严重,目标物体的非对称性更加明显,如图 4 所示。同时与常规图像相比,目标物体丢失了常规图像下的位移、旋转、尺度不变性的特点。



图4 目标在不同 LFOV 图像中的非对称性情况

Fig. 4 Asymmetry of object in different LFOV images

上述这些问题都给大视场域的目标检测与识别算法带来了极大的挑战,因此如何在大视场域中提出准确性更高、鲁棒性更好的目标检测与识别算法无疑有着重要的研究价值。

## 3 大视场域的目标检测与识别算法

越来越多的科研工作者聚焦于大视场域的目标检测与识别,并将其应用于众多科研领域。国外对此研究最早可以追溯至 20 世纪 70 年代。美国宾夕法尼亚大学 Rees<sup>[12]</sup>于 1970 年针对塔楼安防首次提出了使用双曲面反射镜来实现全景视觉的方法。在 20 世纪 90 年代末,美国里海大学实验室利用全景视觉传感器针对目标检测与跟踪进行了研究<sup>[13]</sup>。2004 年美国卡内基梅隆大学的研究员借用多幅图像拼接的思想构建了一个由 5 个摄像机组成的高分辨全景视觉系统(CAMEO),该系统可以实现周围环境的动态目标检测<sup>[14]</sup>。国内针对大视场背景条件下的相关研究起步相对较晚。四川大学曾吉勇等<sup>[15]</sup>于 2003 年针对全景视觉感知技术开展了探索和研究,建立的折反射光学全景成像理论为大视场域的目标检测与识别的进一步探索提供了基础。此

后,国防科学技术大学陈立栋等<sup>[16]</sup>、浙江大学赵琪等<sup>[17]</sup>、上海交通大学图像处理与模式识别研究所李立鸿等<sup>[18]</sup>都逐渐投入大量的科研精力,并取得了一定的研究成果。

近年来,大视场域的目标检测与识别的理论和算法层出不穷,本部分试图依据检测与识别流程中是否对 LFOV 图像进行畸变校正预处理来梳理其发展脉络,将其中的工作分为两大类进行综述:一类是先对 LFOV 图像进行畸变校正的预处理操作,再采用常规的目标检测与识别算法;另一类是直接在 LFOV 图像上完成目标检测与识别。整体结构如图 5 所示,其中目标检测为目标识别提供了目标的位置信息,目标识别在目标检测的基础上完成对此目标的分类。以下将详细介绍这两类大视场域的目标检测与识别算法。

### 3.1 基于畸变校正的目标检测与识别算法

与常规平面图像相比,虽然通过全景视觉传感器获取的 LFOV 图像视场信息丰富,但所获取的 LFOV 图像存在畸变、局部非对称性以及光照非对称性等不符合人眼视觉习惯的情况,这不利于目标检测与识别。早期研究人员对于由相机本身的物理

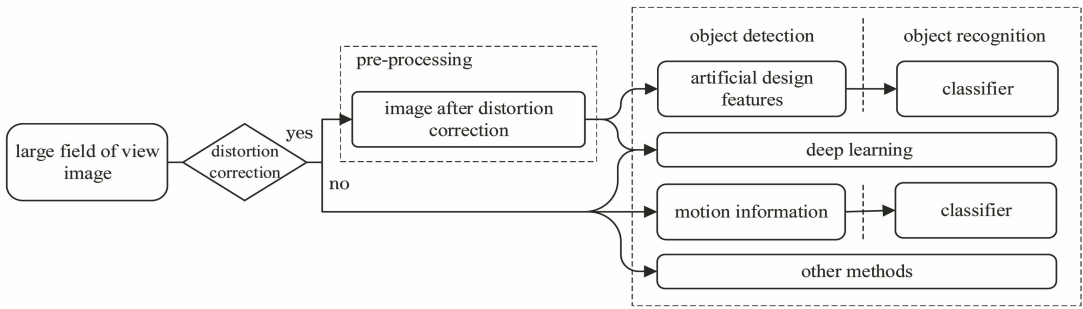


图5 大视场域目标检测与识别算法的分类

Fig. 5 Classification of object detection and recognition in large field of view

特性而造成的大视场图像的畸变,通过相机标定<sup>[19]</sup>估计镜头内外参数的方法进行畸变消除。此后延续这一思路,研究人员又提出了多项式拟合、经纬度等畸变校正算法<sup>[20-21]</sup>,先将 LFOV 图像校正成一个或多个平面图像,再利用正常视角下的目标检测与识别算法完成检测与识别任务。如图 6 所示,此类算法一般分为如下 4 个阶段:1)将获取图像进行畸变校正,减少输入图像的几何失真;2)在此图像上检测出目标物体可能的候选区域;3)对这类候选区域进行特征提取操作;4)将提取的特征送入已训练完成

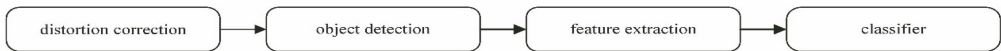


图6 基于畸变校正的目标检测与识别流程图

Fig. 6 Flow chart of object detection and recognition based on distortion correction

### 3.1.1 基于人工设计特征的目标检测与识别算法

基于人工设计特征的目标检测与识别算法依赖于人类对目标自身特征的认识,通过对这种认识的建模,“显式”完成目标特征的提取过程。此类算法通常被用于检测识别几类特定的目标,如行人、车辆等。

Jeong 等<sup>[22]</sup>提出在 LFOV 图像中利用方向梯度直方图(HOG)特征完成车辆检测,首先,通过构建一种畸变模型将畸变图像转换为平面图像,弥补因全景视觉传感器获取大视场而造成的图像失真,而后利用滤波器去除图像的噪声并将其转换为灰度图像,最后利用 HOG 特征<sup>[23]</sup>在此灰度图像上准确检测识别车辆。Silberstein 等<sup>[24]</sup>利用 Caltech 相机标定工具完成 LFOV 图像的畸变校正工作,而后通过构建图像金字塔,在此基础上采用加速特征融合(AFS)完成多尺度行人检测,此方法在每帧图像上的 avgMRF(平均丢失率)为 10.3%。为进一步提高检测与识别的准确性,Levi 和 Silberstein<sup>[25]</sup>在文献[17]的基础上将 AFS 过程进一步细分为粗检测以及细检测两部分,此方法在对特征提取上进行了改

的分类器中,完成目标识别任务。

在图像畸变校正后,依据特征提取阶段方法的不同对这类算法进行进一步划分:1)采用传统的手动设计方式提取人工设计特征;2)采用深度学习的方法得到 LFOV 图像中的目标特征。这两种特征提取方式虽都依据生物神经理论,但也存在不同:手动设计方式提取的特征是基于人类的视觉信息设计的,而基于深度学习的特征提取方式侧重于如何通过深度学习学习样本数据之间潜在的特征信息。

进,先初步筛选出特征,再在初步筛选的基础上进一步对特征进行筛选,排除一些非目标区域,将每帧图像上的 avgMRF 降低至 3.5%,但由于上述方法均忽略消除径向畸变后出现的严重透视畸变,因此 Bertozzi 等<sup>[26]</sup>在图像预处理中采用了一种新的畸变校正方法——圆柱投影来校正 LFOV 图像并采用 Soft-Cascade 和聚集通道特征(ACF)<sup>[27]</sup>完成行人检测与识别,FFPI(false positive per image,每张图错误正例数目)为  $10^{-1}$  时丢失率为 0.35。除上述减少径向畸变的方法外,Suhr 等<sup>[28]</sup>提出使用 Mercator 投影的方法来减少行人因径向畸变导致的形状变化,通过利用 HOG 特征分别检测图像下半部分或上半部分区域内是否存在脚或头来确认是否为行人,这种方法每次检测一部分的局部信息,再对局部信息进行融合以确定目标,最终报警的准确率达 97.5%,比基于 DPM<sup>[29]</sup>的方法提高了 8.3%,可以较好地应用于车辆倒车系统中。

表 1 对上述综述的各算法从检测对象、预处理、特征提取、评价指标(丢失率、报警率)等方面进行了比较。

表 1 各算法的性能比较  
Table 1 Comparison of performance of different algorithms

Paper	Object	Pre-process	Feature extraction	Evaluation	Year
Undistortion using FOV					
Jeong <i>et al.</i> <sup>[22]</sup>	Vehicle	$r_u = \frac{\tan(r_d w)}{2 \tan\left(\frac{w}{2}\right)}$	HOG		2016
Silberstein <i>et al.</i> <sup>[24]</sup>	Pedestrian	Camera calibration	AFS	10.3% (avgMRF)	2014
Levi and Silberstein <sup>[25]</sup>	Pedestrian	Camera calibration	AFS-Multi-Cue	3.5% (avgMRF)	2015
Image un-warping					
Bertozzi <i>et al.</i> <sup>[26]</sup>	Pedestrian	$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} c_u + \text{atan2}(x, z) \cdot f_\theta \\ c_v + k_v \cdot \frac{y}{\sqrt{x^2 + y^2}} \end{bmatrix}$	Soft-Cascade and ACF	0.35 (FPPI is $10^{-1}$ ), 0.59 (FPPI is $10^{-2}$ ), 0.75 (FPPI is $10^{-3}$ )	2015
Image warping					
Suhr <i>et al.</i> <sup>[28]</sup>	Pedestrian	$\theta = \arctan(x/f)$ $\phi = \arctan(y/\sqrt{x^2 + y^2})$ $x' = s\theta$ $y' = s \ln(\sec \phi + \tan \phi)$	HOG and TER-based classifier	97.3% (TWA)	2017

研究人员发现 LFOV 图像边缘部分的信息在畸变校正过程中丢失,于红绯等<sup>[30]</sup>提出的三平面校正方法可以通过将单幅图像校正成多幅平面图像来弥补成像视角的损失,但是这种做法使算法的复杂度明显增加。这种复杂度明显增加的原因在于:该方法针对整幅图像完成畸变校正的预处理工作,但

由于最终目的只需要检测出目标物体,因此当目标物体在图像中占据像素比较小时,对整幅图像进行畸变校正处理,就会存在大量冗余的计算,影响检测速度,对此,研究人员提出在不损失精度的情况下提高检测速度的方法(表 2)。

表 2 各算法的性能比较  
Table 2 Comparison of performance of different algorithms

Paper	Object	Pre-process	Feature extraction	Evaluation	Year
Transformation					
Martinez <i>et al.</i> <sup>[31]</sup>	Human beings	$\phi = c/R_{\text{med}}$ $R_{\text{med}} = (R_{\text{min}} + R_{\text{max}})/2.0$ $c_0 = c_1 + (R_{\text{min}} + r) \sin \phi$ $r_0 = r_1 + (R_{\text{min}} + r) \cos \phi$	Viola-Jones classifier	Time reduced from 600-700 ms to 10-15 ms	2010
Ding <i>et al.</i> <sup>[32]</sup>	Motion object	$\bar{x} = x + \{k_1 x(x^2 + y^2) + s_1(x^2 + y^2) + [p_1(3x^2 + y^2) + 2p_1 xy]\}$ $\bar{y} = y + \{k_2 x(x^2 + y^2) + s_2(x^2 + y^2) + [p_2(3x^2 + y^2) + 2p_2 xy]\}$	Based on texture and color feature	5× faster	2016

Martinez 等<sup>[31]</sup>提出了一种基于全景系统的人员检测系统,该系统首先通过背景减法进行运动目标检测,将场景中所有的运动元素都视为潜在的目标,然后对此部分区域进行畸变校正,由于检测出的运动目标并非全部都是人,因此对校正后的区域使用 Viola-Jones 分类器<sup>[33]</sup>进行人脸检测,从而确认哪些候选区域是所需要关注的目标——人。在监控

领域中,检测与识别算法对算法的实时性提出了更高的要求,但畸变校正过程较为复杂,除不断改进畸变校正算法<sup>[34-35]</sup>外,丁莹等<sup>[32]</sup>采用与文献[31]相类似的思想,针对大视场背景下的运动目标提出了局部视角无畸变的加速算法,对视场内运动目标所在的局部视角图像进行畸变校正和识别等操作。由于仅对整张图像中目标区域进行畸变校正处理,因此

计算数据量大幅减少,缩短了图像的处理时间,在满足识别准确性的前提下,提速5倍以上,实现了实时在线处理。

### 3.1.2 基于深度特征表示的目标检测与识别算法

基于深度特征表示的目标检测与识别算法通过输入大量校正后的图像自动学习隐含在样本中更为本质的特征,提高检测与识别的准确性、鲁棒性。

Yoshimi 和 Takahashi<sup>[36]</sup>提出一种适用于纠正垂直长物体的变形(站立的人或行人)的畸变校正方法——水平全景镶嵌(HPM),之后将处理完成的图片作为 Faster R-CNN (faster region convolutional neural network)网络的输入,最终得到目标检测识别结果,此方法较基于 DPM 的方法在 LAMR(log-average miss rate)上有明显的下降,降低了11.58%。蔡成涛等<sup>[37]</sup>根据全景视觉系统特殊的成像原理建立全景系统的成像参数模型,根据此模型将得到 LFOV 图像的柱面展开图,同时根据目标的

长宽比对 YOLO 的网络结构进行修改,使其适应目标在展开图的形状,其在全景目标检测速率上已经达到 30.89 frame/s,远远满足实时性要求,识别准确率也已经达到 70%以上。徐佳等<sup>[38]</sup>通过球面投影校正的原理得到每张鱼眼图像的 9 个不同方向覆盖 90°视角的平面透视图,然后运用 LCNN(look-up-based convolutional neural network)<sup>[39]</sup>识别校正后的 LFOV 图像中的人脸,此方法在人脸识别准确率方面有显著提升且模型计算量减少。除特定人员目标外,邓军等<sup>[40]</sup>提出一种基于深度学习的火焰识别方法,此方法先将鱼眼图像投影到球面模型上,再对球面模型上获取火焰的候选区域进行畸变校正,将校正后的候选区域输入 CNN(convolutional neural networks)中完成最终的火焰识别任务,在较大视场的情况下具有较低的误检率和漏检率。表 3 从预处理方法、特征提取网络、检测的时间、检测的准确率方面对上述算法进行总结对比。

表 3 各算法的性能比较

Table 3 Comparison of performance of different algorithms

Paper	Object	Pre-process	Feature extraction	Evaluation	Date
Yoshimi <i>et al.</i> <sup>[36]</sup>	Pedestrian	HPM	Faster R-CNN	LAMR: 24.5%; $E=48.65\%$	2017
Cai <i>et al.</i> <sup>[37]</sup>	VOC Pascal	Cylindrical unwarping Correction	YOLO	Detection rate: 30.89 frame · s <sup>-1</sup> Accuracy rate: 72%	2018
Xu <i>et al.</i> <sup>[38]</sup>	Face	Spherical projection	LCNN	97.3% (TWA)	2018
Deng <i>et al.</i> <sup>[40]</sup>	Fire	Spherical projection	CNN		2017

虽然深度学习可以自适应地学习隐含在样本中更为本质的特征,但 Yang 等<sup>[41]</sup>将 LFOV 图像进行等距投影后校正图像作为网络输入,通过比较不同检测与识别算法的结果,得出此类算法若不解决失真问题仍然会产生相应的精度损失。

### 3.1.3 基于畸变校正的目标检测与识别算法的对比分析

基于畸变校正的目标检测与识别算法在 LFOV 图像的预处理阶段加入畸变校正,然后基于校正后的图像完成目标检测与识别任务。由于普通图像下的目标检测与识别算法无法很好地检测与识别 LFOV 图像中的目标,因此研究人员提出利用畸变校正模型最大程度地将 LFOV 图像还原成普通图像,弱化目标在 LFOV 图像中的畸变特性,之后直接利用普通图像下的目标检测与识别算法完成大视场域的目标检测与识别任务。在从 LFOV 图像校正成平面图像的过程中图像质量显著下降,成像视角信息损失,尤其是图像边缘的信息损失严重,这是影响后续目标检测与识别算法精确度以及检测速度的

重要因素之一,这些区域收集了大部分包含目标的场景信息,因此畸变校正成为此类算法能否准确检测与识别目标的关键所在,从表 1、表 2 的畸变校正预处理部分可以发现,研究人员从最初的相机标定方法逐步向基于模型的非量测标定算法改进,大多数研究都关注如何在畸变校正预处理部分进行改进。传统的相机标定方法不但可以用于任意的相机模型,而且参数估计准确,但是其存在如下一些问题:1)依赖于棋盘等特定的标定物;2)计算过程较为复杂;3)难以用于实时监控场景。可以看出,此类方法依赖一定的先验信息,且需要多张视角图来确定相机的内外参数,而基于模型的非量测标定算法则解决了上述问题,也越来越受到研究人员的重视。表 4 对现阶段的大视场域中所用的畸变校正方法进行了整理总结。

基于人工设计特征的目标检测与识别算法的核心在于如何依据先验信息高效准确地提取特征以及构建合理的分类器,这种方法具有设计简单、易于实现、可解释性好等优势,但由于人为设计的特征常常是针对特定的几类目标而设计的,针对性较强,因此

表4 各类畸变校正算法的性能比较

Table 4 Comparison of performance of different algorithms

Method	Advantage	Disadvantage
Camera calibration	Higher precision No restrictions on camera type	Complexity Need to know a certain size of calibration object
Non-metric calibration	Simple Few parameters	Worse stability Not suitable for high-precision system

不能很好地表达大量、多类目标,具有较大的局限性。而基于深度特征表示的目标检测与识别算法通过学习大量样本自动提取能够有效反映检测目标本质的特征,构建更加精确的目标特征提取模型。相比于人为设计特征而言,此方法在一定程度上减少了不具有明显代表性或无法用模型、数学公式进行准确表述的特征。在实际应用中,若要满足实时性的要求则对硬件要求较高,移植到一般的嵌入式设备难度较大,因此,研究人员仍然需要在保证算法的鲁棒性、准确性的前提下,精简算法,提高实时性。

### 3.2 基于原始 LFOV 图像的目标检测与识别算法

全景视觉传感器因种类繁多、成像结构复杂导致形变未知或畸变特性无法准确用模型表达。研究人员发现目前已有的畸变校正过程过于繁琐,校正后的图像存在信息丢失的问题。除此之外,畸变校正过程往往需要一定的先验信息来加以约束,但是实际场景中,这些先验信息并不能准确地表达实际的畸变特性,因此提出直接基于原始 LFOV 图像完成目标检测与识别任务。根据不同的侧重点将此类方法进行归类,可分为如下三类:1)基于深度特征表示的目标检测与识别算法;2)基于运动信息的目标检测与识别算法;3)其他目标检测与识别算法。

#### 3.2.1 基于深度特征表示的目标检测与识别算法

得益于 CNN 在目标检测与识别算法上的快速发展,基于深度特征表示的目标检测与识别在近年来成为研究热点。其基本思想是通过构建深度网络学习图像在不同卷积层上的特征表达,最终得到目标从底层到高层的一个深度特征描述。Deng 等<sup>[42]</sup>第一次将 LFOV 图像直接作为网络输入完成多类目标检测与识别任务,mAP 较 DPM 有明显的提升,从 29.4% 提升至 68.7%,且各个类别的 AP 也均有明显提升。王志旋等<sup>[43]</sup>同样利用 Faster R-CNN 实现了全景图像中路灯底座的检测,研究结果均表明较传统的方法,深度学习在完成多类目标检测与识别任务中展现了强有力的优势。但 LFOV 图像本身存在畸变失真问题,而预训练模型是基于常规

视角训练的,故其提取的特征与真实目标在 LFOV 图像下的特征会存在较为明显的区别。若直接用此训练好的模型进行迁移学习,虽检测结果较传统方法有一定提升,但对于检测平面视角小的目标而言检测效果会有所下降。因此,Fu 等<sup>[44]</sup>提出了通过模拟桶形畸变自制大视场域的真实人脸数据集来训练 CNN 网络,实验结果表明准确性有明显提升。这是因为 CNN 能够从大量的训练样本中学习到这些大视场域的人脸特征。而在实际应用中,此方法并不适用,这是由于缺少大量公开的数据集,故需要通过自行标注数据来完成学习任务,该过程相当消耗人力物力。

普通相机拍摄的图像由于不存在畸变,因此在整个图像上应用相同形状的卷积核对目标物体进行特征提取时,同类目标保持平移、旋转、尺度不变性。而 LFOV 图像存在畸变且不同位置的畸变情况不一样,此时同类目标特征丢失了其在常规图像下的平移、旋转、尺度不变性的特性,若仍采用相同的卷积核完成对整幅图的卷积操作,会造成卷积操作不均匀,无法更好地提取有意义的特征。因此,文献<sup>[45-46]</sup>提出了一种替代方案:将几何变换的不变性直接编码到 CNN 中,结果表明这样的方法在减少模型参数的数量以及所需的训练样本数量方面非常有效。基于此思想,Su 等<sup>[47-48]</sup>在 LFOV 图像的每个纬度使用畸变感知核解决大视场域的目标检测与识别问题。与此相类似,Coors 等<sup>[49]</sup>通过将失真不变性编码到 CNN 的体系结构中,提出了利用 CNN 处理 LFOV 图像的新框架——SphereNet。SphereNet 根据球形图像表示的几何形状调整卷积滤波器的采样网格位置,从而避免失真。Lee 等<sup>[50]</sup>提出一种新的卷积层以及池化层的方法,利用球面多面体来表示 LFOV 视图,该方法最小化了球面上空间分辨率的方差,通过将此卷积层和池化层添加到现有的网络中实现大视场域的目标检测与识别任务。此类方法在训练过程中完全抛弃了畸变校正这一步预处理操作,直接使用大视场域下具有畸变的图像进行训练。

### 3.2.2 基于运动信息的目标检测与识别算法

基于运动信息的目标检测与识别算法通常基于视频序列图像,相比于静态图像,视频序列图像在提取目标的时候具有明显的优势,可以利用视频序列图像中目标的运动特性对图像中变化的区域进行提取;通过对该区域图像的进一步处理(噪声点消除、连通域处理等)得到较为完整的目标信息,但在目标识别阶段,全景相机在成像过程中不同位置的畸变程度不同,导致目标物体在运动过程中,本身特征随运动过程而发生变化,不再具有普通视场下的平移不变性,因此如何选定特征和分类器来识别该目标成为研究人员研究的一个重要方向。此类算法思路简单,但由于特征和分类器的设计难度较大,因此目前只能检测出视频序列中行人和车辆等特定类型的运动目标。

在目标区域提取阶段,常用的方法有光流法<sup>[51-52]</sup>、背景减法<sup>[53-54]</sup>、帧间差分法<sup>[55]</sup>以及其他的一些组合法。文献<sup>[56]</sup>将常用的光流方法,如 Anandan 等<sup>[57]</sup>所提方法、Horn 和 Schunck<sup>[58]</sup>所提方法、Lucas 和 Kanade<sup>[59]</sup>所提方法和 Fleet 和 Jepson<sup>[60]</sup>所提方法,在 LFOV 图像上进行比较,得出如下结论:由于刚体运动这一假设在大视场域下不成立,故 Anandan 等<sup>[57]</sup>所提方法效果较差;Horn 和 Schunck<sup>[58]</sup>所提方法和 Lucas 和 Kanade<sup>[59]</sup>所提方法因匹配过程中存在较多误匹配点导致结果不理想;Fleet 和 Jepson<sup>[60]</sup>所提方法通过计算视频相位的梯度来获取相位的轮廓信息,由于运动目标变化过程中对比度与尺度会发生相应变化,因此这种基于相位的方法较基于幅度的方法有更好的鲁棒性,但计算量过大、时间复杂度较高。文献<sup>[61]</sup>分析了在 LFOV 图像下的光流场分割的一些特性,基于这些特性 Herceg 等<sup>[62]</sup>根据基于移动机器人的测距法搜索估计的光流场与由 LFOV 图像序列计算的光流场两者之间的差异,提出了一种适用于 LFOV 图像的运动目标检测算法,最终检测识别准确率达到 97.19%。根据 LFOV 图像目标在不同位置畸变程度不同的特点,文献<sup>[63]</sup>提出了依据图像中心与光学图像中心的距离来设置不同的卷积核,完成形态学操作,以消除运动目标内部出现空洞的问题。同样考虑到图像的畸变特性,吴健辉等<sup>[64]</sup>引入运动点团模式,从像素级层面解决了 LFOV 图像中因为畸变以及各个不同位置分辨率不同带来的目标大小、形状的改变问题。该方法相比前述的几种方法对图像边缘的大畸变、低分辨率目标有更好的稳定性和准确性,消除了噪点和干扰点对畸变的微小运动目

标的影响。与上述在基于单大视角相机拍摄的图像下完成检测与识别任务的思路不同,近年来,研究人员通过多相机融合技术来提高检测识别的准确性。吴健辉等<sup>[65]</sup>基于文献<sup>[64]</sup>提出构建一个由鱼眼相机和 PTZ 相机组成的主从目标监控系统,利用文献<sup>[64]</sup>中的方法实现目标检测,同时计算出目标对应的 PTZ 参数,通过 PTZ 相机的指向性成像很好地解决了 LFOV 图像边缘分辨率低和畸变严重的问题,目标检测率为 92% 以上,平均捕获误差率为 1.6% 左右。沿用此思想 Baris 等<sup>[66]</sup>采用背景减法完成 LFOV 图像中的目标检测,当目标进入特定范围后调用 PTZ 相机并依据形状的特征完成目标识别任务,检测准确率高达 99%。

### 3.2.3 其他的目标检测与识别算法

除基于运动信息和基于深度学习两种思路外,张力丹等<sup>[67]</sup>将显著性信息引入大视场域的目标检测与识别算法中,通过构建多幅鱼眼图像下不同畸变程度的显著性信息来增加显著性检测约束条件。全局关联性信息约束弥补了单张图像显著性信息较少的缺陷,最终其检测识别的准确度、查全率及  $F$  度量值均达到 80% 以上。

另外还有一些针对特定目标、特定场景提出的大视场域的目标检测与识别算法,设计过程中更多地结合应用场景、目标外形、成像原理等特定信息。如 Saito 等<sup>[68]</sup>针对安装于天花板的监控鱼眼摄像头提出以身体轮廓和头肩轮廓的概率特征作为人体特征,通过核岭回归(KRR)建立概率外观模型,并使用最大后验(MAP)完成大视场域的人员检测。Cinaroglu 等<sup>[69]</sup>首次依据 LFOV 图像本身特征,结合黎曼矩阵完成梯度修正,形成非矩形滑动窗口,显著提高 HOG 特征的自适应性。Wang 等<sup>[70]</sup>同样依据目标在 LFOV 图像中不同位置的形状变化,设计不同的检测掩码,通过与上述这些掩码进行匹配完成目标检测与识别任务。

### 3.2.4 基于原始 LFOV 图像的目标检测与识别算法的对比分析

本部分依据检测与识别方法的侧重点对此类算法予以总结。表 5 从每个类别中选择有代表性的方法进行综合对比分析。

基于运动信息的方法通过对背景信息进行估计,构建出背景模型与时间之间隐含的关联关系,利用当前帧与所建背景模型进行对比作差可间接分离出运动前景,最后通过前景分割得到目标。此类方法一般具有实现简单、运算效率较高等优点,常用于



表 5 各算法的性能比较

Table 5 Comparison of performance of different algorithms

Paper	Object	Method	Evaluation	Year
Deng <i>et al.</i> <sup>[42]</sup>	20 class	Faster R-CNN	mAP: 68.7%	2017
Coors <i>et al.</i> <sup>[49]</sup>	Flying cars	Encode the invariance of geometric transformation directly into CNN	mAP: 50.18%	2018
Herceg <i>et al.</i> <sup>[62]</sup>	Motion object	Corner detection, optical flow	Accuracy: 97.19%	2011
Wu <i>et al.</i> <sup>[65]</sup>	Motion object	Moving blob method; PTZ shots image	Accuracy: 92%	2017
Zhang <i>et al.</i> <sup>[67]</sup>	Salient object	Co-saliency detection algorithm	Precision: 0.82; recall: 0.75; $F_1$ : 0.81	2017
Cinaroglu <i>et al.</i> <sup>[69]</sup>	People	Omnidirectional sliding window Modified HOG+SVM	SVM scores: 2.94	2014
Wang <i>et al.</i> <sup>[70]</sup>	People	Template-based	MOTA: 0.85	2017

视频下目标检测与识别任务,但适用范围较为狭窄,通常只能检测与识别运动的目标,检测过程中易受到光照变化、阴影、动态背景等因素的影响,对目标造成误检或漏检,从而影响识别准确率。而基于深度特征表示的方法则是将显式的结构先验嵌入到深度学习模型中,通过设计不同的采样卷积,使机器自动学习刻画样本的本质特征,学习隐含在数据内部的关系,最终完成目标检测与识别任务。此类方法克服了基于运动信息的方法只能检测运动目标的这一局限,增加了目标检测与识别的目标类别,但是此类方法需要依赖大量的学习样本并且样本的覆盖面要广。在实际应用中,此类方法由于网络中参数量较大会占用大量的存储空间,在移动端的应用很受限制。虽然可以通过将上述两种方法进行融合来解决上述存在的一些弊端,但是由于 LFOV 图像与普通平面图像之间的一些映射关系难以刻画,因此如何较为准确地刻画目标在 LFOV 图像中的一些特征是研究者们今后仍需要解决的问题。

## 4 大视场域的目标检测与识别算法的思考

目前大视场领域的目标检测与识别算法已经取得了一系列进展,本文接下来从算法的统一性和差异性两个方面对大视场域的目标检测与识别算法进行讨论,探讨其未来的发展方向。

### 4.1 大视场域的目标检测与识别各类算法的统一性和差异性

#### 1) 统一性

大视场域的目标检测与识别作为目标检测与识别的一个重要分支,纵观近年来的各类相关文章可以发现基于畸变校正的目标检测与识别的算法和基

于原始 LFOV 图像的目标检测与识别这两类算法在研究思路上有统一性(图 5),均包含了如下的两种研究思路:

(1)专注于构建目标本身结构信息的传统目标检测与识别方法。利用目标自身结构信息在设计特征时具有较好的解释性以及较高的准确性,但是这些结构信息通常是针对某一类或者某几类的目标来设计的,尤其是当多类目标的结构相类似或者同一类目标具有多种不同类型的结构时,这都会导致此类算法的鲁棒性较差。虽然可以通过增加结构信息来提高算法的鲁棒性,但是可能会因信息冗余而降低识别的精度,影响检测与识别速度。

(2)专注于通过深度学习构建目标从底层至高层语义的目标检测与识别方法。深度学习是直接从图像像素中学习得到与任务相关的特征描述,这弥补了基于手工设计的底层特征描述子的固有缺陷,即因对目标的描述依赖于人类的固有的一些“显式”知识而造成对目标的刻画不够本质,这往往会过早地丢失一些潜在有用信息。深度学习类似于根据已知的输入输出信息,自动求解输入与输出之间的映射关系的过程,但对于深度特征的选择缺少强有力的理论解释。

#### 2) 差异性

LFOV 图像与常规图像相比,最大不同在于目标物体在 LFOV 图像的成像过程中产生畸变、几何位置信息发生偏移等问题,因此如何利用畸变信息成为大视场领域的目标检测与识别算法亟待解决的关键问题。从处理流程(图 5)可清晰看出,检测过程中是否采用畸变校正算法是这两类算法之间最大的不同。基于畸变校正的目标检测与识别算法是在预处理阶段利用畸变信息构建畸变校正模型,而基于原始 LFOV 图像的目标检测与识别算法则是在

特征提取阶段利用畸变信息直接提取原始 LFOV 图像中物体的形变特征。二者各有优劣,前者通过畸变校正模型将 LFOV 图像转换成平面图像(全局信息),这就使得在检测与识别阶段其与常规图像的目标检测与识别模型并无太大差异,因此当目标形变特征信息(局部信息)无法准确表达时,仍可以通过全局信息检测识别出此目标,但检测与识别的准确性主要依赖于畸变校正模型的构建,需要更多地考虑全景视觉传感器成像的原理,但畸变校正后的图像都存在丢失边缘信息的问题;后者充分利用了 LFOV 图像具有大视场的优势,保留图像在畸变校正过程中丢失的边缘信息,通过对目标在大视场下

的形变特点进行分析,将畸变信息融入在检测与识别的过程中,但同一类物体的特征因在 LFOV 图像中所处位置不同而不同,使得这种局部形变特征难以准确表达。由此可见,两类算法之间虽存在较大差异,但仍有互补性。

#### 4.2 大视场域的目标检测与识别的发展方向

图 7 展示了近年来的大视场域的目标检测与识别算法的主要发展脉络以及这些关键技术的基本思想。与此同时,结合当前大视场域的目标检测与识别研究的现状和难点以及算法的统一性和差异性,本文从如下 4 个方面提出了大视场域的目标检测与识别的发展方向。

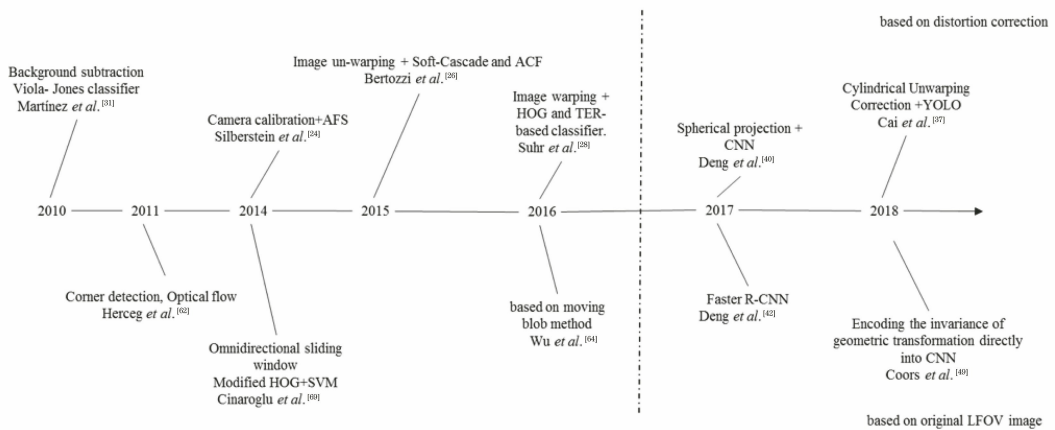


图 7 大视场域的目标检测与识别算法的主要发展脉络

Fig. 7 Road map of object detection and recognition in large field of view

##### 1) 畸变特征的精确提取

畸变校正过程本身被认为是一个病态求逆问题。LFOV 图像在成像过程中,通过一定的映射函数完成世界坐标到图像坐标的映射,但在映射过程中,像素点会存在重叠,造成部分信息丢失,尤其是图像边缘部分信息,这就导致在畸变校正后的重构图像中一些目标的特征信息往往在校正过程中难以重建。畸变校正只能根据一定的先验知识施加约束,根据上述映射函数的逆变换完成校正过程。但在真实环境中,这些约束往往不能被满足,尤其当受到设备本身的影响,即因制作工艺、制作环境、制作设备等不同导致同一类全景视觉传感器采集到的图像在视场范围、物体畸变程度等方面存在一些潜在、未知且随机变化的差异,这些未知差异难以用准确的数学解析式准确表达。而直接对原始图像进行处理,避免了畸变校正过程中细节信息难以重建的问题,能够充分保留 LFOV 图像中目标的畸变特征,正逐渐成为近年来的研究热点。

##### 2) 基于弱监督学习的样本数据挖掘

通过深度学习构建目标从底层至高层语义的目标检测与识别方法是大视场域的目标检测与识别算法中一种重要的研究思路。而深度学习模型的好坏依赖于标记数据集的数据规模、多样性等。实际研究中发现目前缺少公开的 LFOV 图像的目标检测与识别数据集。研究人员虽利用自制数据集在特定的图像测试集中得到较好的效果,但自制数据集难以覆盖实际情况,在实际系统运行过程中很难保证较高的检测与识别的准确率。若直接对实际数据进行标注,则标注成本过大。而弱监督学习结合了主动学习以及半监督学习的主要思想,较好地解决了这一数据标注匮乏的问题,仅从少量标注的数据以及大量没有标签的图像数据中训练一个满足准确率和性能的目标检测模型<sup>[71-72]</sup>,对大视场域的目标检测与识别算法在降低人工标注成本和提高检测灵活性方面无疑有着重要的意义。

##### 3) 混合深度学习

当目标特征易于抽象时,传统的目标检测与识

别算法有较好的表现且易部署于端;若场景特征较为复杂、难以抽象时,则需利用深度学习对整张图像进行特征提取,但这对计算资源要求较高。而混合深度学习<sup>[73]</sup>是深度学习方法与传统方法相结合的一种方法,尝试利用传统的目标检测与识别算法对目标区域进行初步筛选,在此基础上利用深度学习完成目标检测与识别,这为大视场域的目标检测与识别算法提供了一种新的解决方案。

#### 4) 深度强化学习

在 LFOV 图像中快速精确地检测和识别目标是大视场域的目标检测与识别算法的关键。而深度强化学习充分结合深度学习与强化学习的优势,相较于直接定位目标位置存在定位精度不高、定位框难以准确标定目标物体等问题,深度强化学习中的定位智能体不断地做出动作决策,迭代地定位到待检测目标<sup>[74]</sup>,实现对环境的自适应控制,这为大视场域的目标检测与识别算法提供了新的方式。

## 5 结束语

大视场域的目标检测与识别作为目标检测与识别的一个重要分支,以其视场范围大等优势逐渐被应用于各个领域。以是否对 LFOV 图像进行畸变校正为划分依据,将近年来大视场域的目标检测与识别算法分为基于畸变校正的目标检测与识别算法和基于原始 LFOV 图像的目标检测与识别算法,全面梳理和总结了这两类算法。在此基础上,对这两类算法进行对比分析,并进一步讨论了其未来发展趋势。在不同的应用环境下,研究相对精确和鲁棒且满足实时性要求的检测与识别算法必然是一个重要技术增长点。

### 参 考 文 献

- [1] Everingham M, van Gool L, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [2] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge [J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [3] Saez A, Bergasa L M, Romeral E, et al. CNN-based fisheye image real-time semantic segmentation[C] // 2018 IEEE Intelligent Vehicles Symposium (IV), June 26-30, 2018. Changshu. IEEE, 2018: 1039-1044.
- [4] Zhang M, Yao J, Xia M H, et al. Line-based multi-label energy optimization for fisheye image rectification and calibration [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015. Boston, MA, USA. IEEE, 2015: 4137-4145.
- [5] Yin X Q, Wang X C, Yu J, et al. FishEyeRecNet: A multi-context collaborative deep network for fisheye image rectification [M] // Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 475-490.
- [6] Ju M R, Luo H B, Wang Z B, et al. Improved YOLO V3 algorithm and its application in small target detection[J]. Acta Optica Sinica, 2019, 39(7): 0715004.  
鞠默然, 罗海波, 王仲博, 等. 改进的 YOLO V<sub>3</sub> 算法及其在小目标检测中的应用[J]. 光学学报, 2019, 39(7): 0715004.
- [7] Wang J Q, Chen K, Yang S, et al. Region proposal by guided anchoring [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019. Long Beach, CA, USA. IEEE, 2019: 2965-2974.
- [8] Cai Z W, Vasconcelos N. Cascade R-CNN: delving into high quality object detection[C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018. Salt Lake City, UT. IEEE, 2018: 6154-6162.
- [9] Liu L, Ouyang W L, Wang X G, et al. Deep learning for generic object detection: a survey[EB/OL]. (2019-08-22) [2019-10-10]. <https://arxiv.org/abs/1809.02165>[LinkOut]
- [10] Li X J, Feng D W, Xiang Y. Refractive-diffractive hybrid panoramic annular optical system design[J]. Laser & Optoelectronics Progress, 2019, 56(18): 182201.  
李显杰, 冯大伟, 向阳. 折衍混合环带式全景光学系统设计[J]. 激光与光电子学进展, 2019, 56(18): 182201.
- [11] Zhang W, Cong M Y, Wang L P. Algorithms for optical weak small targets detection and tracking: review [C] // International Conference on Neural Networks and Signal Processing, 2003. Proceedings of the 2003, December 14-17, 2003. Nanjing. IEEE, 2003: 643-647.
- [12] Rees D W. Panoramic television viewing system: US3505465[P]. 1970-04-07.
- [13] Boulton T, Micheals R, Erkan A, et al. Frame-rate

- multi-body tracking for surveillance[C]//Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999: 1106-1112.
- [14] Rybski P E, de la Torre F, Patil R, et al. CAMEO: camera assisted meeting event observer[C]//IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004, April 26-May 1, 2004. New Orleans, LA, USA. IEEE, 2004.
- [15] Zeng J Y, Su X Y. Catadioptric omnidirectional system with undistorted imaging for horizontal scene[J]. *Acta Optica Sinica*, 2003, 23(5): 636-640.  
曾吉勇, 苏显渝. 水平场景无畸变的折反射全景成像系统[J]. *光学学报*, 2003, 23(5): 636-640.
- [16] Chen L D, Wang W, Zhang M J, et al. Design analysis of a complementary double-mirror catadioptric omnidirectional imaging system[J]. *Acta Optica Sinica*, 2010, 30(12): 3487-3494.  
陈立栋, 王伟, 张茂军, 等. 双反射面互补结构的折反射全向成像系统设计[J]. *光学学报*, 2010, 30(12): 3487-3494.
- [17] Zhao Q, Shi X, Gong W, et al. Large field-of-view and deep tissue optical micro-imaging based on parallel wavefront correction algorithm[J]. *Chinese Journal of Lasers*, 2018, 45(12): 1207001.  
赵琪, 石鑫, 龚薇, 等. 基于并行波前校正算法的大视场深穿透光学显微成像[J]. *中国激光*, 2018, 45(12): 1207001.
- [18] Li L H, Shi P F, Zhao Q F. A multi-resolution based image registration method for panorama[J]. *Journal of Image and Graphics*, 2006, 11(9): 1271-1275.  
李立鸿, 施鹏飞, 赵群飞. 基于多分辨率技术的快速全景图图像匹配算法[J]. *中国图象图形学报*, 2006, 11(9): 1271-1275.
- [19] Zhang Z. Flexible camera calibration by viewing a plane from unknown orientations[C]//Proceedings of the Seventh IEEE International Conference on Computer Vision, September 20-27, New York: IEEE, 1999: 666-673.
- [20] Wu J, Wang L R, Huang M Y, et al. High precision calibration of fisheye camera with single image under multiple geometric constraints [J]. *Acta Optica Sinica*, 2018, 38(11): 1115001.  
吴军, 王玲容, 黄明益, 等. 多几何约束下的鱼眼相机单像高精度标定[J]. *光学学报*, 2018, 38(11): 1115001.
- [21] He Y, Wang L X, Cai Y, et al. Research progress and prospect of catadioptric panoramic system [J]. *Chinese Optics*, 2017, 10(5): 681-698.
- [22] Jeong J S, Kim H T, Kim B, et al. Wide rear vehicle recognition using a fisheye lens camera image[C]//2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), October 25-28, 2016. Jeju, South Korea. IEEE, 2016: 691-693.
- [23] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), June 20-25, 2005, San Diego, CA, USA. IEEE, 2005: 886-893.
- [24] Silberstein S, Levi D, Kogan V, et al. Vision-based pedestrian detection for rear-view cameras[C]//2014 IEEE Intelligent Vehicles Symposium Proceedings, June 8-11, 2014. MI, USA. IEEE, 2014: 853-860.
- [25] Levi D, Silberstein S. Tracking and motion cues for rear-view pedestrian detection[C]//2015 IEEE 18th International Conference on Intelligent Transportation Systems, September 15-18, 2015. Gran Canaria, Spain. IEEE, 2015: 664-671.
- [26] Bertozzi M, Castangia L, Cattani S, et al. 360° detection and tracking algorithm of both pedestrian and vehicle using fisheye images [C]//2015 IEEE Intelligent Vehicles Symposium (IV), June 28-July 1, 2015. Seoul, South Korea. IEEE, 2015: 132-137.
- [27] Yang B, Yan J J, Lei Z, et al. Aggregate channel features for multi-view face detection [C]//IEEE International Joint Conference on Biometrics, September 29-October 2, 2014. Clearwater, FL, USA. IEEE, 2014: 1-8.
- [28] Suhr J K, Jung H G. Rearview camera-based backover warning system exploiting a combination of pose-specific pedestrian recognitions [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018, 19(4): 1122-1129.
- [29] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model [C]//2008 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008. Anchorage, AK, USA. IEEE, 2008: 1-8.
- [30] Yu H F, Liu W, Yuan H, et al. Moving object detection based on sub-block motion compensation [J]. *Acta Electronica Sinica*, 2017, 45(1): 173-180.  
于红绯, 刘威, 袁淮, 等. 基于子块运动补偿的运动目标检测[J]. *电子学报*, 2017, 45(1): 173-180.
- [31] Martinez E, del Pobil A P. A panoramic vision system for human-robot interaction [C]//2010 5th

- ACM/IEEE International Conference on Human-Robot Interaction (HRI), March 2-5, 2010. Osaka. IEEE, 2010: 171-172.
- [32] Ding Y, Quan W, Han C. A motion detection accelerated algorithm based on distortion correcting of local field angle [J]. *Science Technology and Engineering*, 2016, 16(6): 70-75.  
丁莹, 权巍, 韩成. 局部视角无畸变的运动目标检测加速算法[J]. *科学技术与工程*, 2016, 16(6): 70-75.
- [33] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C] // *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, December 8-14, San Diego, CA, USA, 2001: 511-518.
- [34] Zhang B F, Ma Z J, Zhu J C, et al. Implementation of fisheye image real-time correction based on DSP [J]. *Laser & Infrared*, 2015, 45(2): 181-184.  
张宝峰, 马志军, 朱均超, 等. 基于 DSP 的鱼眼图像实时校正系统的实现[J]. *激光与红外*, 2015, 45(2): 181-184.
- [35] Wu Z J, Wu Q Y, Zhang B C. A new calibration method for fisheye lens based on spherical model[J]. *Chinese Journal of Lasers*, 2015, 42(5): 0508006.  
吴泽俊, 吴庆阳, 张佰春. 一种新的基于球面模型的鱼眼镜头标定方法[J]. *中国激光*, 2015, 42(5): 0508006.
- [36] Yoshimi K, Takahashi K. Warping approach for rearview pedestrian detection with fish eye cameras [C] // *2017 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, June 27-28, 2017. Vienna, Austria. IEEE, 2017: 121-126.
- [37] Cai C T, Wu K J, Liu Q F, et al. Panoramic multi-object real-time detection based on improved YOLO algorithm [J]. *Computer Engineering and Design*, 2018, 39(10): 3259-3264, 3271.  
蔡成涛, 吴科君, 刘秋飞, 等. 基于改进 YOLO 算法的全景多目标实时检测[J]. *计算机工程与设计*, 2018, 39(10): 3259-3264, 3271.
- [38] Xu J, Yang H B, Song Y, et al. A face recognition technology based on fisheye camera [J]. *Information & Communications*, 2018, 31(1): 131-132.  
徐佳, 杨鸿波, 宋阳, 等. 基于鱼眼摄像头的一种人脸识别技术[J]. *信息通信*, 2018, 31(1): 131-132.
- [39] Bagherinezhad H, Rastegari M, Farhadi A. LCNN: lookup-based convolutional neural network[C] // *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 21-26, 2017. Honolulu, HI. IEEE, 2017: 7120-7129.
- [40] Deng J, Qin X B, Wang W F. A method of flame recognition based on deep learning: CN106845410A [P]. 2017-06-13.  
邓军, 秦学斌, 王伟峰. 一种基于深度学习模型的火焰识别方法: CN106845410A [P]. 2017-06-13.
- [41] Yang W Y, Qian Y L, Kamarainen J K, et al. Object detection in equirectangular panorama [C] // *2018 24th International Conference on Pattern Recognition (ICPR)*, August 20-24, 2018. Beijing. IEEE, 2018: 2190-2195.
- [42] Deng F C, Zhu X R, Ren J M. Object detection on panoramic images based on deep learning [C] // *2017 3rd International Conference on Control, Automation and Robotics (ICCAR)*, April 24-26, 2017. Nagoya, Japan. IEEE, 2017: 375-380.
- [43] Wang Z X, Zhong R F, Xie D H. Automatically measuring the coordinates of streetlights in vehicle-borne spherical images [J]. *Journal of Image and Graphics*, 2018, 23(9): 1371-1381.  
王志旋, 钟若飞, 谢东海. 球面全景影像自动测量路灯坐标的方法[J]. *中国图象图形学报*, 2018, 23(9): 1371-1381.
- [44] Fu J L, Alvar S R, Bajic I V, et al. Fddb-360: face detection in 360-degree fisheye images [EB/OL]. (2019-02-07)[2019-07-19]. <https://arxiv.org/abs/1902.02777v1>.
- [45] Cohen T, Welling M. Group equivariant convolutional networks [C] // *International Conference on Machine Learning*, June 19-24, 2016. New York. IEEE, 2016: 2990-2999.
- [46] Worrall D E, Garbin S J, Turmukhambetov D, et al. Harmonic networks: deep translation and rotation equivariance [C] // *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 21-26, 2017. Honolulu, HI. IEEE, 2017: 5028-5037.
- [47] Su Y C, Grauman K. Learning spherical convolution for fast features from 360 imagery [EB/OL]. (2018-11-07)[2019-07-19]. <https://arxiv.org/abs/1708.00919>.
- [48] Tateno K, Navab N, Tombari F. Distortion-aware convolutional filters for dense prediction in panoramic images [M] // *Computer Vision-ECCV 2018*. Cham: Springer International Publishing, 2018: 732-750.
- [49] Coors B, Condurache A P, Geiger A. SphereNet: learning spherical representations for detection and

- classification in omnidirectional images [M] // Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 525-541.
- [50] Lee Y, Jeong J, Yun J, et al. SpherePHD: applying CNNs on a spherical PolyHeDron representation of 360° images [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019. Long Beach, CA, USA. IEEE, 2019: 9173-9181.
- [51] Horn B K P, Schunck B G. Determining optical flow [J]. Artificial Intelligence, 1981, 17(1/2/3): 185-203.
- [52] Barron J L, Fleet D J, Beauchemin S S. Performance of optical flow techniques[J]. International Journal of Computer Vision, 1994, 12(1): 43-77.
- [53] Haritaoglu I, Harwood D, Davis L S. W<sup>4</sup>: real-time surveillance of people and their activities [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 809-830.
- [54] McKenna S J, Jabri S, Duric Z, et al. Tracking groups of people [J]. Computer Vision and Image Understanding, 2000, 80(1): 42-56.
- [55] Collins R T, Lipton A J, Kanade T, et al. A system for video surveillance and monitoring [R]. Carnegie Mellon University, 2000: 1-68.
- [56] Stratmann I. Omnidirectional imaging and optical flow [C] // Proceedings of the IEEE Workshop on Omnidirectional Vision 2002. Held in conjunction with ECCV '02, June 2, 2002, Copenhagen, Denmark. IEEE Comput. Soc, 2002: 104-111.
- [57] Anandan P. A computational framework and an algorithm for the measurement of visual motion [J]. International Journal of Computer Vision, 1989, 2(3): 283-310.
- [58] Horn B K P, Schunck B G. Determining optical flow [J]. Artificial Intelligence, 1981, 17(1/2/3): 185-203.
- [59] Lucas B, Kanade T. An iterative technique of image registration and its application to stereo [EB/OL]. (1981-01-01) [2019-07-19] [https://www.researchgate.net/publication/239061443\\_An\\_iterative\\_technique\\_of\\_image\\_registration\\_and\\_its\\_application\\_to\\_stereo](https://www.researchgate.net/publication/239061443_An_iterative_technique_of_image_registration_and_its_application_to_stereo).
- [60] Fleet D J, Jepson A D. Computation of component image velocity from local phase information [J]. International Journal of Computer Vision, 1990, 5(1): 77-104.
- [61] Šamija H, Marković I, Petrović I. Optical flow field segmentation in an omnidirectional camera image based on known camera motion [C] // Proceedings of the 34th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2011), May 23-27, 2011, Opatija, Hrvatska, 2011, 165-169.
- [62] Herceg D, Marković I, Petrović I. Real-time detection of moving objects by a mobile robot with an omnidirectional camera [C] // 2011 7th International Symposium on Image and Signal Processing and Analysis (ISPA), September 4-6, 2011, Dubrovnik, Croatia, IEEE, 2011: 289-294.
- [63] Jin G F. Pedestrian detection and tracking of fisheye images based on multi-feature fusion [D]. Xi'an: Xidian University, 2017: 23-28.  
金敢峰. 基于多特征融合的鱼眼图像行人检测与跟踪 [D]. 西安: 西安电子科技大学, 2017: 23-28.
- [64] Wu J H, Shang C, Zhang G Y, et al. Study of moving objects detection in fisheye image based on moving blob method [J]. Computer Engineering and Applications, 2016, 52(24): 158-161, 177.  
吴健辉, 商橙, 张国云, 等. 基于运动点团的鱼眼图像中多目标检测方法研究 [J]. 计算机工程与应用, 2016, 52(24): 158-161, 177.
- [65] Wu J H, Shang C, Zhang G Y, et al. A master-slave object surveillance system based on fisheye camera and PTZ camera [J]. Computer Engineering and Science, 2017, 39(3): 540-546.  
吴健辉, 商橙, 张国云, 等. 鱼眼相机与 PTZ 相机相结合的主从目标监控系统 [J]. 计算机工程与科学, 2017, 39(3): 540-546.
- [66] Baris I, Bastanlar Y. Classification and tracking of traffic scene objects with hybrid camera systems [C] // 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), October 16-19, 2017. Yokohama. IEEE, 2017: 1-6.
- [67] Zhang L D, Zhu J C, Feng W J. Fish-eye lens image co-saliency detection algorithm [J]. Journal of Tianjin University of Technology, 2017, 33(4): 41-45.  
张力丹, 朱均超, 冯为嘉. 鱼眼图像协同性目标检测方法 [J]. 天津理工大学学报, 2017, 33(4): 41-45.
- [68] Saito M, Kitaguchi K, Kimura G, et al. People detection and tracking from fish-eye image based on probabilistic appearance model [C] // SICE Annual Conference 2011, September 13-18, 2011, Tokyo, Japan, IEEE, 2011: 435-440.
- [69] Cinaroglu I, Bastanlar Y. A direct approach for

- object detection with catadioptric omnidirectional cameras [J]. *Signal, Image and Video Processing*, 2016, 10(2): 413-420.
- [70] Wang T, Chang C W, Wu Y S. Template-based people detection using a single downward-viewing fisheye camera [C] // *2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, November 6-9, 2017. Xiamen, China. IEEE, 2017: 719-723.
- [71] Zeng Y, Zhuge Y Z, Lu H C, et al. Multi-source weak supervision for saliency detection [C] // *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 15-20, 2019. Long Beach, CA, USA. IEEE, 2019: 6074-6083.
- [72] Vardazaryan A, Mutter D, Marescaux J, et al. Weakly-supervised learning for tool localization in laparoscopic videos [M] // *Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*. Cham: Springer International Publishing, 2018: 169-179.
- [73] Zhuang Z, Zhang G H, Dong W, et al. Intelligent fault detection of high-speed railway turnout based on hybrid deep learning [M] // *AI 2018: Advances in Artificial Intelligence*. Cham: Springer International Publishing, 2018: 98-103.
- [74] Li Y, Fu K, Sun H, et al. An aircraft detection framework based on reinforcement learning and convolutional neural networks in remote sensing images [J]. *Remote Sensing*, 2018, 10(2): 243.