

文章编号:1007-2780(2018)08-0690-07

基于 Mask R-CNN 的 ORB 去误匹配方法

张 博^{1,2}, 韩广良^{1*}

(1. 中国科学院 长春光学精密机械与物理研究所, 长春 130033)

(2. 中国科学院大学, 北京 100049)

摘要:为了提高多目标图像的 ORB 匹配的正确率,提出一种基于 Mask R-CNN 的图像 ORB 去除误匹配方法,该算法首先通过 Faster R-CNN 方法对图像进行识别,运用区域推荐网络得到矩形框标注的感兴趣区域和类别标签,该步骤可以得到感兴趣区域的预测类别和坐标信息,并且通过全卷积网络卷积层进行像素级别校正,得到像素级别的目标所属类别,然后进行目标分割。最后在原有 ORB 特征点匹配基础上,剔除两幅图像中相同目标分割区域以外的误匹配点。为了验证该方法的有效性,对传统 ORB 匹配与基于本文方法的 ORB 匹配进行了仿真实验。改进后的算法,使得在多目标环境下的目标的匹配精度提高了约 18.6%,结果表明,本文算法较传统的 ORB 匹配算法的精度有一定提高。

关键词:ORB 匹配;Mask R-CNN 算法,误匹配

中图分类号:TP394.1;TH691.9 **文献标识码:**A **doi:**10.3788/YJYXS20183308.0690

ORB removal mis-matching method based on Mask R-CNN

ZHANG Bo^{1,2}, HAN Guang-liang^{1*}

(1. *Changchun Institute of Optics, Fine Mechanics and Physics,*
Chinese Academy of Sciences, Changchun 130033, China;

2. *University of Chinese Academy of Sciences, Beijing 100049, China*)

Abstract: This paper presents an elimination image ORB mis-matching method based on Mask R-CNN in order to improve the accuracy of multi-target image matching ORB. First of all, the algorithm identifies the image by faster R-CNN method, and it uses Region Proposal Network to obtain the interest area and category labels which are marked by rectangular box. We can get the forecast category and coordinate information of the interest area by this step. And then we correct the pixel level by the convolution layer of the Fully Convolutional Networks, and get the category of target belonging to the pixel level and segment target. Finally, the mis-matching points outside the same target segmentation area are eliminated in the two images on the basis of the original ORB matching. At the same time, some simulation experiments which use the traditional ORB matching method and the ORB matching based on this paper is carried out to verify the effectiveness of this method. The results show that the algorithm is much more accurate than that of the traditional ORB matching algorithm. The improved

收稿日期:2018-03-09; **修订日期:**2018-06-06.

基金项目:国家自然科学基金(No.61602432)

Supported by National Natural Science Foundation of China(No.61602432)

* 通信联系人, E-mail: hangl@ciomp.ac.cn

algorithm makes the matching accuracy in the multi-target environment be increased by 18.6%.

Key words: ORB matching; Mask R-CNN algorithm; Mis-matching

1 引言

随着机器视觉与自主导航的发展,机器视觉中的及时定位与建图技术变得愈发重要,对其中关键技术特征点的匹配精度也提高了要求。但机器视觉中的图像复杂度越来越高,对特征点的匹配带来了挑战,伴随着机器学习的发展,目标的识别和分割技术也愈发成熟,对传统的基于特征点的匹配方法 ORB^[1]、FAST^[2] 算子等都可以进行优化。在多目标的定位中,ORB 是 FAST 特征点与 BRIEF^[3] 描述子结合的一种匹配方法,具有良好的方向不变性和尺度不变性,ORB 匹配方法由于运算速度较快、匹配精度较高得到了较好的应用,但传统的 ORB 匹配方法在目标定位中也会受到目标外未知环境的干扰,难以得到良好的匹配精度,采用传统的 RANSAC^[4] 方法计算量较大,通过给定的迭代次数,找出最优解,对参数的设置会极大的影响精度和计算复杂度,难以得到较好的匹配效果,并且对特定目标的匹配达不到所要求。在发展过程中,许多学者提出了图像匹配方法,例如张迁提出的基于 SUSAN^[5] 算子的配准方法,徐丽燕对多光谱图像基于特征点的匹配^[6] 方法。目前最热门的也是结合深度学习方法进行匹配,随着深度学习的发展,目标的识别与分割技术也得到了极大的提升,应用也愈加广泛。

考虑到 Mask R-CNN^[7-8] 方法运算速度快,识别精度与目标切割效果较好,具有保持特定目标良好的轮廓信息的作用,因此本文提出了基于 Mask R-CNN 的 ORB 匹配方法,对匹配进行约束。本文先对多目标图像进行识别,得到感兴趣区域的预测类别和坐标信息,并且通过全卷积网络卷积层进行像素级别校正,得到像素级别的目标所属类别,然后进行目标分割。得到分割后的轮廓信息,进行图片的类别匹配,通过得到的轮廓信息,剔除匹配目标对儿的轮廓外的误匹配。在保证准确率的同时获得效果较好的匹配^[9] 点。

2 ORB 匹配方法

2.1 FAST 特征点的提取

ORB 是基于 FAST 特征点和 BRIEF 描述子的特征提取算法,(1)通过粗提取提取大量 FAST 特征点,并通过检测特征点周围像素值,剔除不符合条件的特征点。(2)将特征点输入决策树处理,提取最优的 FAST 特征点。(3)通过非极大值抑制的方法去除局部特征点密集。(4)通过高斯金字塔保证其尺度不变性。(5)通过矩的方法计算质心,保证方向不变型,定义如公式(1):

$$m_{pq} = \sum_{x,y \in r} x^p y^q I(x,y), \quad (1)$$

其中: $I(x,y)$ 为图像灰度,该矩的质心如式(2):

$$C = (\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}), \quad (2)$$

则向量的角度即为该特征点的方向,如式(3):

$$\theta = \arctan(\frac{m_{01}}{m_{00}} / \frac{m_{10}}{m_{00}}) = \arctan(m_{01}/m_{10}), \quad (3)$$

2.2 BRIEF 描述子的计算

BRIEF 描述子是平滑图像后,在 FAST 特征点周围选择一个区域,并选择 n 个点对进行像素值的比较,会得到一串二进制结果,如公式(4):

$$\tau(p;x,y) = \begin{cases} 1 & p(x) < p(y) \\ 0 & \text{其他} \end{cases}, \quad (4)$$

所以 BRIEF 描述子可用二进制的向量表式,如公式(5):

$$f_n(p) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(p;x_i,y_i), \quad (5)$$

即 BRIEF 描述子匹配可以通过汉明距离匹配。

3 改进的 ORB 匹配方法

3.1 Mask R-CNN 目标识别

任意的输入图像,首先通过 CNN(卷积神经网络)进行特征提取,然后通过区域推荐网络产生评分后的推荐区域。模拟这个过程通过全连接卷积神经网络,因为要与 Fast R-CNN^[10] 目标检测网络共享计算,为了产生推荐区域,要在卷积特征

图上进行网格滑动, $n \times n$ 的网格与输入的卷积特征图全连接, 并把滑动窗口映射到低维的向量, 通常可以映射为 256 维或者 512 维。该向量被投入到两个全连接层进行分类和边框回归, 当选择 n 为 3 时输入图像的有效区域最大。该微型网络结构在图 1 中表示出来, 因为该微型网络以滑动窗口方式进行计算, 全连接层在所有的空间位置共享权值。该结构先通过 $n \times n$ 的卷积层进行转换后由 1×1 的滑动卷积层再进行处理。

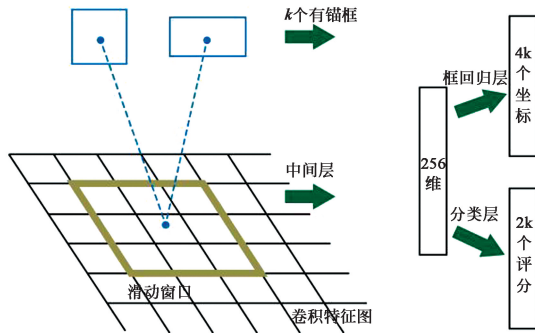


图 1 区域推荐网络

Fig.1 Region proposal network

在每个滑动窗口的位置上, 我们同时预测 k 个推荐区域, 框回归层会产生 k 个框的 $4k$ 个坐标编码。对于每个推荐区域统计的目标或者是非目标的可能性框识别层会产生 $2k$ 个评分。 k 个推荐被参数化后被称为锚。每个锚都集中在识别问题中的滑动窗口中间, 通常使用 3 个尺度和 3 个长宽比, 在每个滑动窗口位置产生 $k=9$ 个锚。对于尺寸大小 $W \times H$ (通常小于 2400), 一般有 WHk 个锚。该方法的重要特性是具有平移不变性的, 并且在计算锚和计算相关与锚的函数中都具有该特性。作为对比, 多框运用 K-means 方法产生锚就不具备平移不变性, 如果一幅图像中的目标目标位置变化, 推荐目标也应该变化, 函数应该能够预测任意位置的推荐目标。

为了训练区域推荐网络, 对无论是目标的还是非目标的每个锚设定一个二进制标签, 我们会给两种锚设定为正标签, (1) 有最高交并比 (交并比可以理解为模型产生的目标窗口与原标记窗口的交叠率, 即检测结果的交集比上并集) 的锚, (2) 与任何实际框交并比都高于 0.7 的锚。也要注意到一个实际框可以为多个锚点设定正标签。反而, 对与任何实际框位置的交并比低于 0.3 的锚

设定为负标签。但锚的标签设定对目标的权值训练不产生影响。

通过这些定义, 最小化 Fast R-CNN 的多任务目标损失函数, 对于目标的损失函数定义如公式 (6):

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{\text{reg}}(t_i, t_i^*), \quad (6)$$

其中: i 是少量训练样本中锚的索引, p_i 是锚 i 被预测为目标概率, 如果锚的标签设定为正则实际标签概率 p_i^* 为 1, 反之标签为负则概率为 0. t_i 是代表了预测框的 4 个参数化坐标的向量。 t_i^* 是设定为正标签的实际框。分类损失函数 L_{cls} 是有目标和无目标的对数损失函数。对于回归损失函数, 使用函数 $L_{\text{reg}}(t_i, t_i^*) = R(t_i - t_i^*)$ 其中 R 为鲁棒性较好的损失函数, $p_i^* L_{\text{reg}}$ 表示回归损失由正标签的锚激励和负标签的锚归零。输出的分类和框回归层各自包括 $\{p_i\}$ 和 $\{t_i\}$, 这两个变量由 N_{cls} 、 N_{reg} 和一个平衡权重参数 λ 规范化。对于回归, 采用如公式 (7) 和 (8) 的 4 组坐标参数化。

$$t_x = (x - x_a) / \omega_a, t_y = (y - y_a) / h_a, \\ t_w = \log(\omega / \omega_a), t_h = \log(h / h_a), \quad (7)$$

$$t_x^* = (x^* - x_a) / \omega_a, t_y^* = (y^* - y_a) / h_a, \\ t_w^* = \log(\omega^* / \omega_a), t_h^* = \log(h^* / h_a), \quad (8)$$

其中: x, y, w 和 h 表示框的中心坐标和宽度与高度。变量 x, x_a, x^* 分别代表预测框, 锚框和真实框, 这也可以理解为从锚框到附近真实框的回归。该方法通过基于特征图获得框回归。框的回归是基于任意尺寸区域的特征池化基础上展现的, 并且回归权重可以在所有区域上共享, 极大地提高运算效率。在公式中 (1) 中, 在特征图上用于回归的特征在空间上具有相同的尺寸。考虑一组 k 个回归框的不同尺寸, 每个独立的回归之间不共享权重, 并且有独自的尺度和宽高比。尽管特征有固定的尺寸仍然可能预测到不同框的尺寸。

区域推荐网络是由全连接网络实现的, 可以由端到端的反向传播算法和随机梯度下降法训练, 遵循以图像为中心的抽样策略来训练网络, 每个小组训练集都来自包含有正样本和负样本的单个图片, 可能对所有锚的损失函数进行优化, 但也可能由于负样本主导而产生偏差。相反, 在图像中随机抽取 256 个锚计算小组的损失函数, 其中

正标签和负标签的锚比例为 1 : 1; 如果一幅图片中少于 128 个正样本, 则用负样本填充。

从 0 均值并且标准差为 0.01 的高斯分布随机抽取权重初始化新的无权重层。其他的层(例如共享权重卷积层)由 ImageNet 分类数据集训练模型初始化。对于 60k 个小组训练采用 0.002 的学习率。

已经描述完成如何训练区域推荐网络, 但是并没有考虑使用这些推荐的目标检测卷积神经网络。对于检测网络, 采用的是与 Fast R-CNN 权值共享的区域推荐网络。但训练确实互相独立的, 会以不同的方式修正权重。所以需要一种方法共享他们的卷积层, 但不能通过定义同时包含两种网络的结构达到共享, 然后进行反向传播算法优化达到目的。原因是 Fast R-CNN 训练决定于固定目标推荐, 如果改变推荐机制会具有不明确的先验。通过 ImageNet 训练来初始化模型训练 RPN, 由 ImageNet 训练初始化的模型, 由区域推荐进行端到端的微调, 第二步运用之前产生的推荐通过 Fast R-CNN 分割检测网络的训练, 也是由 ImageNet 训练模型初始化。第三步通过目标检测网络初始化区域推荐训练网络, 但仅是共享卷积层和微调的区域推荐网络。现在两个网络共享卷积层。最后, 将共享卷积层固定, 如此, 两个网络共享卷积层组成一个独特的网络。

3.2 已识别目标轮廓特征提取

对于像素级别的切割, 是与上文框回归和目标识别并行的运算, 训练过程中会对感兴趣区域定义多任务损失函数如公式(9):

$$L = L_{cls} + L_{box} + L_{mask}, \quad (9)$$

对于每个感兴趣区域的输出是分辨率为 $m \times m$ 的 K 个二进制分割编码。应用逐像素的 sigmoid 函数, 定义平均二进制交叉熵损失函数 L_{mask} , 感兴趣区域与真实类相关, L_{mask} 只由第 k 个轮廓信息定义。网络对于每个类产生的分割信息的损失函数 L_{mask} 类之间不产生竞争, 依靠专门的分类层预测类别标签用于选择输出分割信息。不同于常见的全连接层语义分割方法, 通常使用的是逐像素的 softmax 和多项交叉损失函数, 会产生类别竞争。但通过逐像素的 sigmoid 和二进制损失函数不会产生类别竞争。这个公式也是好的实例分割结果的关键。

对输入目标进行轮廓编码, 因此, 不同于通过

全连接层将类别标签或者框信息处理成短的向量信息, 提取轮廓的空间结构可以由逐像素的卷积得到。特殊的, 对于使用全连接层的感兴趣区域预测到的分割信息, 这就不用将 $m \times m$ 的空间布局分割信息拆分成缺少空间维度信息的向量。全连接层会使用较少的参数并且也更准确。

逐像素处理的方法要求感兴趣区域的特征是小特征图, 并且要对像素的空间位置一致性进行精准的校正。这使得接下来的区域分割预测起到关键作用。

提取感兴趣区域的小特征图(例如 7×7), 区域池化是标准操作。区域池化首先将一系列感兴趣区域量化为分离的特征图, 被细分为空间二进制文件, 最后特征值会由最大池化操作进行聚合。

小部分的转换虽然可能不会影响分类, 具有鲁棒性, 但对预测像素级分割信息具有负面效果。所以用区域校正层来移除这种负面效果, 采用的方法是不对感兴趣区域进行量化分离。使用双线性插值的方法在区域 4 个规则采样位置提取特征值, 然后进行最大值或平均值的聚合。

3.3 目标的 ORB 特征点匹配

算法整体流程图如图 2 所示, 选择合适的双目相机图像, 进行卷积处理获取目标的类别标签和识别框, 再通过区域推荐网络卷积处理得到目标像素级分割信息。然后对图像进行基于本文方法的去误匹配。

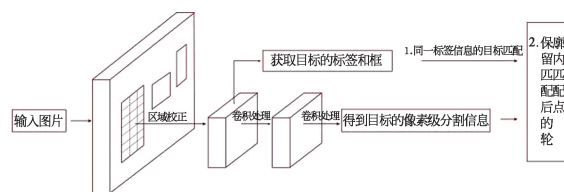


图2 去误匹配算法图

Fig.2 Removal of mis-matching algorithm

具体的算法流程与实施细则如下: 选取双目相机图像, 通过 Mask R-CNN 算法获得识别后目标的轮廓信息与标签信息, 并对其进行序列化处理, 转化为二进制数据流, 使用 python 中的标准化模块 pickle 库存储, 避免了大量图片的二次存储。

对获取的类别标签信息与轮廓信息的 pickle 文件进行反序列化, 提取关键信息, 对双目图片进行 ORB 特征点匹配, 通过 Opencv 对两幅图片中的所有特征点坐标信息和匹配点数目提取。设第

i 个特征点在第 n ($1 \leq n \leq 2$) 张图坐标为 (x_i^n, y_i^n) , 第 n 张图的第 m 类的轮廓区域为 S_{nm} ($1 \leq n \leq 2$)。

对第一幅图像中的特征点进行遍历, 通过 numpy 库对特征点在第一张图像中的坐标进行判定, 判定其是否在图片中已经识别出的某一类的轮廓内, 如公式(10):

$$\begin{cases} (x_i^1, y_i^1) \in S_{1m} & \text{第 } i \text{ 个点在 } m \text{ 类轮廓内} \\ (x_i^1, y_i^1) \notin S_{1m} & \text{第 } i \text{ 个点不在 } m \text{ 类轮廓内} \end{cases} \quad (10)$$

若在轮廓内则对其在第二幅图像中匹配点坐标进行判定, 若在同一类的轮廓内, 则通过 Opencv 保留匹配, 画出匹配线。

针对如下两种情况对匹配进行去误处理。
(1) 第一幅图像的特征点不在识别出的 m 个类别轮廓内, 由于不符合识别目标的匹配, 所以去除匹配点。
(2) 特征点在第一幅图像中的已识别轮廓区域, 但不在第二幅图相中与第一幅图像配准类别的轮廓区域内。如公式(11)所示:

$$\begin{cases} (x_i^1, y_i^1) \notin S_{1m} \\ (x_i^1, y_i^1) \in S_{1m} \cap (x_i^2, y_i^2) \notin S_{2m} \end{cases} \quad \text{去除匹配点} \quad (11)$$

在算法的实施中, 类的匹配方法采用哈希表方法, 时间复杂度为 $O(1)$, 可以较快地进行类别的匹配。

4 实验结果及分析

为了验证方法有效性, 选取主频 2.2 GHz 的 i7 处理器, 选取 VMware 虚拟机的 64 位 Ubuntu14.04 系统, 基于 python 语言 PyCharm 为实验平台进行实验。采用的数据集为 KITTI 双目数据集与自然场景数据, 以下为进行的多组实验中的 3 组。

图 3 为数据的原始双目图像, 大小为 1241×376 。

图 4 为经过 Mask R-CNN 处理后得到轮廓信息图, 经过 KITTI^[11] 数据集的测试, 汽车的识别准确率可以达到 90% 以上。

图 5 为常规 ORB 的匹配效果图, 其中汽车轮廓内存在误匹配, 对识别目标的匹配准确率有较大影响。

图 6 为基于 Mask R-CNN 的 ORB 去误匹配方法的实验图, 采用突出目标匹配的效果图, 去除



图 3 原始双目图片
Fig.3 Original stereo image



图 4 经过 Mask R-CNN 处理后得到轮廓信息图
Fig.4 Outline image after Mask R-CNN processing

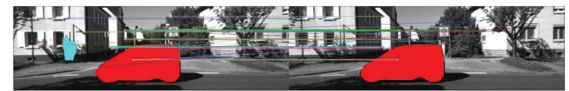


图 5 ORB 匹配图
Fig.5 ORB matching image



图 6 本文的匹配效果图
Fig.6 Image of the matching algorithm in the paper

了汽车目标与目标外的匹配, 经过 KITTI 数据集测试, 对目标的匹配精度得到提高。

图 7、8、9 为选取的第二组实验, 经过本文算法处理保留了目标人和车的正确匹配, 提高了匹配精度。

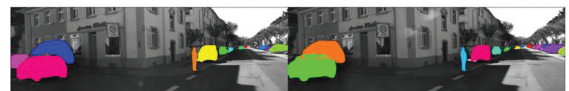


图 7 经过 Mask R-CNN 处理后得到轮廓信息图
Fig.7 Outline image after Mask R-CNN processing

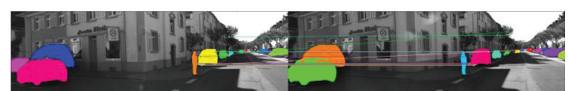


图 8 ORB 匹配图
Fig.8 ORB matching image



图 9 本文的匹配效果图

Fig.9 Image of the matching algorithm in the paper

图 10、11、12 为选取的第三组实验,选取的环境为自然场景,目标为人和狗,经过本文算法处理保留了目标人和人的正确匹配。

由数据集结果进行多组匹配精度的统计计算,设定 TM 为目标的正确匹配数, FM 为目标



图 10 原始双目图片

Fig.10 Original stereo image

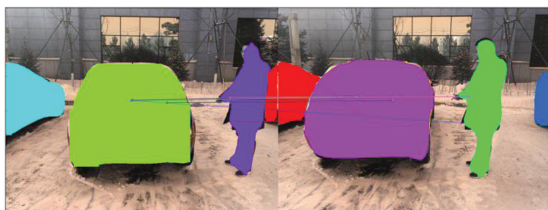


图 11 ORB 匹配图

Fig.11 ORB matching image

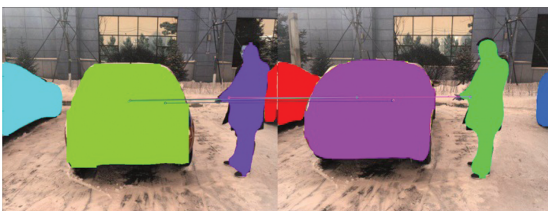


图 12 本文的匹配效果图

Fig.12 Image of the matching algorithm in the paper

的错误匹配数, $R = TM / (TM + FM)$ 为匹配正确率,由表 1 可以看出,本文的算法在匹配精度上有较好的提高,在目标较为清晰的情况下,对目标匹配精度提高做平均计算,精度提高约为 18.6%,说明本文算法的有效性。

表 1 ORB 匹配方法与本文方法比较结果

Tab.1 Comparison between the ORB matching method and the method in this paper

	数据	匹配方法	
		本文算法	ORB 匹配方法
KITTI 单目标	TM	3	3
	FM	0	1
	R	1	0.75
KITTI 多目标	TM	4	4
	FM	4	5
	R	0.5	0.44
采集图像	TM	4	4
	FM	0	3
	R	1	0.75

5 结 论

本文根据图片处理环境日益复杂,对匹配精度要求的提高,提出了基于 Mask R-CNN 的 ORB 去误匹配方法,并介绍了卷积神经网络系统的结构和工作原理。研究了 ORB 在双目 KITTI 数据集的匹配。最后,通过卷积神经网络进行目标匹配以及获得的轮廓信息,进行去除误匹配算法。实验结果证明:本文算法较传统的 ORB 匹配算法对目标的匹配精度提高了约 18.6% 以上。

参 考 文 献:

- [1] RUBLEE E, RABAU D V, KONOLIGE K, et al. ORB: an efficient alternative to SIFT or SURF [C]//*Proceedings of 2011 International Conference on Computer Vision*. Barcelona, Spain: IEEE, 2011: 2564-2571.
- [2] ROSTEN E, DRUMMOND T. Machine learning for high-speed corner detection [C]//*Proceedings of the 9th European Conference on Computer Vision*. Graz, Austria: Springer, 2006: 430-443.
- [3] CALONDER M, LEPETIT V, STRECHA C, et al. BRIEF: binary robust independent elementary features [C]//*Proceedings of the 11th European Conference on Computer Vision*. Heraklion, Crete, Greece: Springer, 2010: 778-792.

- [4] FISCHLER M A, BOLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [J]. *Communications of the ACM*, 1981, 24(6): 381-395.
- [5] SMITH S M, BRADY J M. SUSAN—a new approach to low level image processing [J]. *International Journal of Computer Vision*, 1997, 23(1): 45-78.
- [6] 张二磊, 马骏, 王晓田. 一种改进的 SURF 彩色遥感图像配准算法[J]. 液晶与显示, 2017, 32(2): 144-152.
ZHANG E L, MA J, WANG X T. Improved SURF algorithm for color remote sensing image registration [J]. *Chinese Journal of Liquid Crystals and Displays*, 2017, 32(2): 144-152. (in Chinese)
- [7] 杨梦铎, 李凡长, 张莉. 李群机器学习十年研究进展[J]. 计算机学报, 2015, 38(7): 1337-1356.
YANG M D, LI F C, ZHANG L. Advances in the study of lie group machine learning in recent ten years [J]. *Chinese Journal of Computers*, 2015, 38(7): 1337-1356. (in Chinese)
- [8] HE K M, GKIOXARI G, DOLLÁR P, *et al.* Mask R-CNN [C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE, 2017: 2980-2988.
- [9] 吴晓波, 安文斗, 杨钢. 图像测量系统中的误差分析及提高测量精度的途径[J]. 光学 精密工程, 1997, 5(1): 133-141.
WU X B, AN W D, YANG G. Error analysing and approaches of improving measureing precision in image measuring system [J]. *Optics and Precision Engineering*, 1997, 5(1): 133-141. (in Chinese)
- [10] GIRSHICK R. Fast R-CNN [C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago, Chile: IEEE, 2015: 1440-1448.
- [11] GEIGER A, LENZ P, STILLER C, *et al.* Vision meets robotics: the KITTI dataset [J]. *The International Journal of Robotics Research*, 2013, 32(11): 1231-1237.

作者简介: 张博(1991—), 男, 吉林辽源人, 硕士研究生, 2014 年于吉林大学获得学士学位, 主要从事机器视觉方面的研究。E-mail: 1220792913@qq.com

韩广良(1968—), 男, 山东嘉祥人, 博士, 研究员, 现为中科院长春光机所图像室主任, 主要从事目标识别、目标跟踪、机器视觉方面的研究。E-mail: hangl@ciomp.ac.cn