

红外图像中基于似物性与稀疏编码的行人检测

魏丽¹, 丁萌^{2,3}, 曾丽君¹

(1.南京航空航天大学金城学院, 江苏南京 210016; 2.南京航空航天大学民航学院, 江苏南京 210016;

3.光电控制技术重点实验室, 河南洛阳 471023)

摘要: 行人检测是计算机视觉的经典问题。针对红外图像中的行人检测问题, 提出了一种基于似物性和稀疏编码及空间金字塔特征提取的行人检测方法。首先, 针对红外图像的特点, 利用基于频域残差的显著性分析方法得到红外图像的显著图, 在此基础上提出了一种似物性计算方法, 进而得到不同区域的似物度得分, 并根据得分提取出感兴趣区域; 其次, 以尺度不变特征转换为基础, 将稀疏编码和空间金字塔算法应用于非监督特征学习实现对感兴趣区域的特征提取; 最后, 利用线性支持向量机构建分类器实现对图像中每个感兴趣区域的行人检测。实验结果验证了本文提出的感兴趣区域提取算法和针对单幅红外图像行人检测算法的有效性。

关键词: 红外图像; 行人检测; 似物性; 频域残差; 稀疏编码; 空间金字塔

中图分类号: TP391.4 文献标识码: A 文章编号: 1001-8891(2016)09-0752-06

Pedestrian Detection Based on Objectness and Sparse Coding in a Single Infrared Image

WEI Li¹, DING Meng², ZENG Lijun¹

(1. Jincheng College, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China;

2. College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

3. Science and Technology on Electro-optic Control Laboratory, Luoyang 471023, China)

Abstract: Pedestrian detection is a classic issue of computer vision. For the pedestrian detection problems in a single infrared image, this paper proposes a pedestrian detection method based on objectness, sparse coding and spatial pyramid matching. The algorithm can be divided into three phases. Firstly, the saliency map is computed based on spectral residual, and the paper presents an objectness score computation based on saliency map and selects regions of interest according to the score of different sub-windows. Secondly, scale-invariant feature transform, sparse coding and spatial pyramid matching are used to extract the feature vectors of the regions of interest. Finally, linear support vector machine is used to build a classifier and detect pedestrian in each region of interest. The experimental results verify the effectiveness of objectness score computation and the proposed algorithm for infrared images.

Key words: infrared image, pedestrian detection, objectness, spectral residual, sparse coding, spatial pyramid matching

0 前言

行人检测 (pedestrian detection) 一直是计算机视觉、模式识别领域研究的热点之一^[1]。从单幅图像中实现对行人进行检测的算法主要由 2 部分组成: 感兴趣区域 (region of interest, ROI) 的获取和识别算法设

计。

ROI 获取也称为目标候选区域确定, 其主要目的是从图像中首先确定可能包括行人的区域, 确定 ROI 区域就可以避免了识别算法对整个图像区域的遍历, 即传统意义上的滑窗法 (slide window), 这可以降低行人检测算法的运算量, 提高算法的实时性。现有的

收稿日期: 2016-01-21; 修订日期: 2016-05-10。

作者简介: 魏丽: (1982-), 女, 江苏镇江人, 讲师, 硕士, 红外图像分析、信号处理。

通讯作者: 丁萌: (1981-), 男, 江苏仪征人, 副教授, 博士, 计算机视觉与模式识别。

基金项目: 航空科学基金(20155152041), 国家自然科学基金(61203170), 中国博士后基金特别资助(2013T60539), 中央高校基本科研业务费(NS2016061)。

ROI获取方法总体上可以分为静态背景下基于序列图像的方法和动态背景下基于单幅图像的方法,本文所研究的主要是动态背景下基于单幅红外图像的行人检测方法。近年来,基于显著性(saliency)分析的ROI区域获取方法成为研究热点^[2-4],由此实现了行人检测从滑窗到选择窗(Selected Window)转变,通过对显著图窗口的似物性得分(objectness score)的计算^[5-6],确定窗口中的存在目标的可能性,由此实现对窗口的选择,达到ROI区域提取的目的。

不同于一般的行人检测算法所针对可见光图像,本文所研究的红外图像其在色彩、纹理、边缘信息方面的丰富程度都远远不及前者。因此,现有的许多基于色彩、边缘等的显著性分析方法无法有效在红外图像中加以利用。为此,针对红外图像中的ROI区域提取问题提出了一种基于红外图像频域残差的ROI区域提取算法。

识别算法一般由2个部分组成,即特征提取与分类器设计。在行人检测的识别算法中,最为经典方法是HOG+SVM的方法^[7],即利用方向梯度直方图(histogram of gradient, HOG)进行特征提取,利用支持向量机(support vector machine, SVM)进行分类器训练。在HOG特征的基础上,近年来又出现了共生梯度方向直方图特征、二阶梯度方向直方图、局部二值模式(LBP)以及cell-LBP特征等。在分类器算法方面,又先后出现了最小二乘SVM、Adaboosting等^[8-12]。

近年来,随着深度学习(deep learning)、稀疏编码(sparse coding)等方法出现,传统的特征提取方法逐渐被取代。ScSPM(sparse coding spatial pyramid matching)就是建立在稀疏编码和特征池化基础上的特征提取方法^[13-14]。其将传统的特征提取升华为无监督特征学习,在相同分类器的前提下,利用ScSPM特征学习获取的特征向量所得到的识别准确率有显著提高。由于本文所研究的对象是红外图像,因此传统的基于可见光图像所开发的特征提取方法在识别过程中的效果不是十分理想,为此,本文在SCSPM架构的基础上,利用非监督特征学习实现行人的特征提取,在此基础上利用SVM实现行人检测。

本文针对单幅红外图像提出了一种基于窗口似物性得分和非监督特征学习的行人检测方法。该方法首先利用频域显著性分析方法计算窗口似物性得分,根据得分获取图像中的ROI区域,在此基础上利用非监督特征学习方法对ROI区域进行特征提取,进而利用SVM分类器实现ROI区域的行人检测。

1 ROI区域获取

相比于可见光图像,红外图像的纹理、色彩及边缘信息的丰富程度较低,因此现有的许多基于纹理、边缘等信息的显著性分析方法都无法有效地用于红外图像。为此,利用基于频域残差(spectral residual, SR)的显著性分析方法实现对ROI区域的提取。基于SR的显著性分析方法是由Hou等人首先提出的,SR的基本步骤如下^[2]:

输入: 红外图像X

输出: 显著图S

步骤1: 计算X的傅里叶变换F;

步骤2: 计算傅里叶变换的对数幅值: $L = \lg(|F|)$;

步骤3: 求出傅里叶变换的相位角: $P = \text{Angle}(F)$;

步骤4: 对对数幅值进行均值滤波处理: $L' = \text{AverageFilter}(L)$;

步骤5: 计算频谱残差: $R = L - L'$;

步骤6: 生成残差傅里叶变换: $F' = \text{Exp}(R) * \text{Exp}(iP)$;

步骤7: 利用傅里叶反变换计算显著性图: $S = \text{InverseFourierTransform}(F')$.

在利用SR算法计算出红外图像X对应的频域残差显著图S前需要对图像进行成比例缩小(缩小系数 δ),这样做的目的是为了确保对显著度局部峰值对目标区域的集中。在得到显著图S后,通过设置阈值threshold,得到显著图对应的二值图T,将二值图T的每个像素与显著图S的对应像素相乘得到合成图M。在此基础上,计算不同窗口区域的似物性得分score:

$$\text{score} = \frac{\alpha \times \beta}{a \times b} \quad (1)$$

式中: α 表示二值图T的相应窗口范围内的所有点的像素值之和; β 表示合成图M的相应窗口范围内的所有点的像素值之和。

在行人检测中确定搜索窗口区域的大小范围为从原始图像高度30%~80%作为搜索窗口的高度变化范围,搜索窗口的宽度为对应高度的50%。图1为ROI区域确定方法流程图。图2为ROI区域确定过程中得到的中间结果和最终结果,图2(e)中的数字为score值(为了显示方便仅取score值的整数部分)。

2 行人特征提取

以尺度不变特征(scale-invariant feature transform, SIFT)为基础利用稀疏编码在SPM架构下建立目标的稀疏特征是由Yu Kai等^[13]人最早提出的,该特征是建立非监督特征学习的基础上,相较于传统的SIFT、HOG、LBP等特征提取方法,其得到的特征向量其对于不同目标可分性越高,可以利用简单的分类

器完成不同类型的分类。本文所采用的特征提取方法如下：

输入：待提取特征的图像块，即单个 ROI 区域
输出：特征向量

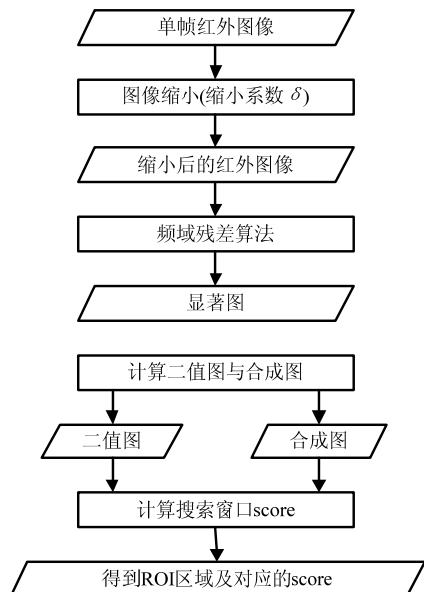


图 1 ROI 获取算法流程

Fig.1 Flowchart of ROI detection algorithm

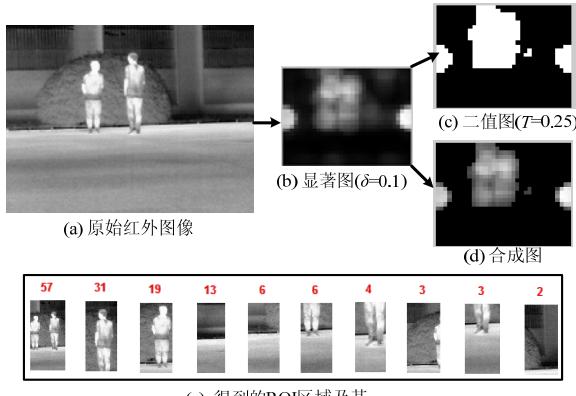


图 2 ROI 获取结果

Fig.2 Result of ROI detection

步骤 1：将图像块大小缩放为固定大小；本文将所有图像块固定为 128×64 pixel。

步骤 2：将图像块分成若干子区域，每个子区域的大小为 $g_1 \times g_1$ ；其中相邻两个子区域之间的重叠度为 g_2 ，在本文中 $g_1=16$, $g_2=10$ ，由此共得到 $19 \times 9=171$ 个子区域（如图 3）。

步骤 3：对每个子区域提取一个 SIFT 特征，即一个 128 维向量，因此对一个 128×64 pixel 图像块，共得到 171 个 128 维特征向量。

步骤 4：利用稀疏字典计算每个特征向量的稀疏表示；在本文中选择的字典规模为 128×1024 ，即每个特征向量的稀疏表示为一个 1024 维的稀疏向量。

步骤 5：对图像块进行金字塔表示，将根据不同的划分尺度 τ ，对二维平面图像进行不同尺度的划分。当划分尺度 $\tau=0$ 时，不对图像进行划分，即原始图像，把它看作金字塔的第 1 层。当 $\tau=1$ 时，将图像平均分成 4 个部分，作为金字塔的第 2 层。依次类推，对于划分尺度 τ ，将图像平均分成 2^{τ} 份，对应金字塔的第 $\tau+1$ 层；在本文中 $\tau=3$ 层 ($\tau=0,1,2$)，可得到 $1+4+16=21$ 个图像子块。

步骤 6：对每个图像子块中对应的稀疏向量进行特征池化（pooling）。本文利用最大值池化（max-pooling）进行特征融合，该方法最大的优点是对局部噪声有鲁棒性。

步骤 7：特征合并完成目标特征提取，通过最大值特征池化共得到 21 个 1024 维融合特征向量，将这 21 个特征首尾相接得到一个长度为 $1024 \times 21=21504$ 维的特征向量，完成目标的特征提取。



图 3 图像块划分

Fig.3 Image patch division

3 分类器设计

分类器一直以来都是模式识别研究的重点，模式识别理论诞生于 20 世纪 60 年代，在 20 世纪 70 年代成为一门独立的学科体系。模式识别方法大体上可以分为 4 类：统计模式识别、句法模式识别、模糊模式识别、神经网络模式识别。前两者是较为古老的模式识别方法，当前基于神经网络的模式识别是较常用的方法，其中以小样本学习理论与支持向量机成为新的研究与应用热点。本文利用 SVM 为为基础设计分类器。

基于机器学习的识别算法由 2 个部分组成：第一是离线的学习（learning）过程，就是将包含行人的 ROI 区域作为正样本，不包含行人的作为负样本，并预先抽取出来作为训练样本，利用监督学习的方法训练分类器，所有的训练原本都需要首先进行尺寸上的统一，在此基础上提取出特征向量及其对应的标签（label），完成训练后得到表征分类器的参数，由于这一训练过

程是预先完成的,因此称为离线训练过程。第二是在线识别(recognition)的过程,利用离线训练得到的分类器,在线对提取出的ROI区域进行特征提取,将其特征向量送入分类器,得到其对应的识别标签,根据标签就可以判断每个ROI区域中是否包含行人。

在利用SVM训练分类器过程中,对训练样本中负样本的选择参考该样本的似物性得分,对于似物性较高的负样本,在随机训练过程中的选中概率较高,该类负样本的用于训练的重复度也较高。这是因为ROI获取过程中得到这类负样本的概率较高,故分类器应该能够重点剔除这类负样本。

4 实验分析

4.1 ROI区域获取算法实验验证

图4为3幅行人检测的红外图像,分别拍摄于室内弱光环境、室外白天环境与室外夜晚环境,从图中可见,人体温度与周围环境温度之间有着明显区别。图5为利用本文提出的ROI获取算法得到的3幅图像的ROI区域,由图可见提出的方法均能将行人包含在ROI区域中,由此证明了本算法的有效性。同时,图5中每个ROI区域上方对应的数字为其score值(为显示方便取整值),从获取的score值可见,行人目标的显著性明显,由此证明了采用频域残差算法的有效性。

4.2 识别算法实验验证与比较

使用相同的SVM分类器,针对两种不同的特征提取方法,对本文所使用的特征提取方法和常用的HOG法针对红外图像中的行人识别的准确率进行比较。所选择的训练样本和测试样本完全相同,通过选择训练样本的数量的不同,得到的识别结果如表1所示。在实验中,共有正样本318个,负样本685个,分4组进行实验,每组训练样本的数量分别为:第1组正样本100个,负样本200个;第2组正样本150个,负样本300个;第3组正样本200个,负样本400个;第4组正样本250个,负样本500个。所有剩余样本作为测试样本,每组样本随机抽取10次,得到的识别结果如表1所示。图6为每组样本数量的平均识别率,从该图可见,本文利用的基于ScSPM的分类器的识别率明显高于基于HOG的分类器的识别率。此外,训练样本数量的增加也有助于提高分类器

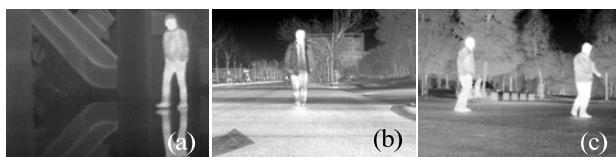


图4 用于行人检测的红外图像

Fig.4 Infrared images for pedestrian detection

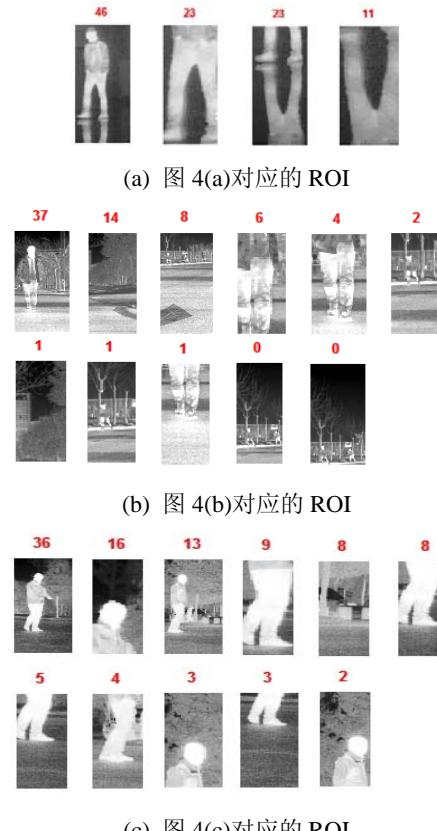


图5 ROI区域获取 Fig.5 ROI detection

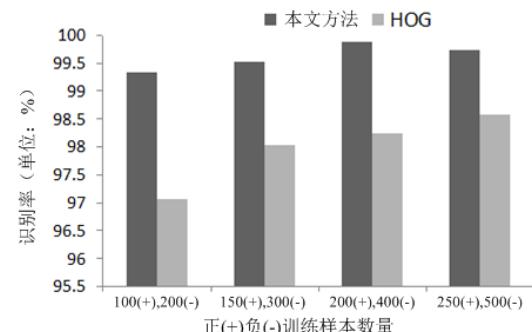


图6 不同训练样本数量的平均识别率

Fig.6 Average recognition rates based on different number of training samples
的识别率。

4.3 行人检测实验验证

利用第4组训练样本得到的分类器对图2与图5的ROI区域进行识别,得到的行人检测的结果如图7所示,由图7可见,本文提出的方法可以有效地从红外图像中完成行人检测。

5 结论

针对单幅红外图像,研究了一种基于似物性得分和稀疏编码空间金字塔特征提取的行人检测方法,在红外图像频域显著性分析的基础上提取出ROI区域,利用非监督特征学习与SVM实现对图像中的行人检

测,验证了该算法针对单幅红外图像行人检测的有效性。本文的主要特点如下:

1) 根据红外图像特点,利用频域残差算法获取红外图像的显著图;

2) 提出了一种基于频域残差显著图的窗口似物性得分计算方法;

3) 将窗口似物性得分作为得到选择训练样本负样本的重要参考,有针对性地进行分类器训练。

表1 本文方法和HOG法对红外图像中的行人识别的准确率比较

Table 1 Recognition rates based on the first set of training samples

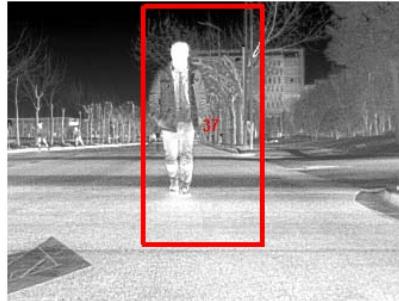
组别	提取方法	识别准确率									
		99.4382	98.6470	100	99.2088	98.9795	99.7706	99.6675	99.2320	99.0826	99.3351
第1组	本文方法	97.0415	96.2504	97.9821	97.6265	95.9179	96.0442	96.0210	97.1910	98.5671	98.0852
	HOG	99.8701	99.5725	99.3128	98.9773	99.8701	99.8701	98.6797	99.7024	99.5725	99.8701
第2组	本文方法	98.7175	99.1829	96.8398	97.0996	98.4199	98.1602	97.5271	98.1223	98.6797	97.6569
	HOG	99.5763	99.8246	100	100	100	99.5763	100	100	100	99.8246
第3组	本文方法	98.4508	98.2025	98.9771	98.5534	97.3550	97.1067	98.1296	98.8017	98.8745	97.9542
	HOG	99.2647	99.2647	100	100	100	99.7297	98.9944	100	100	100
第4组	本文方法	98.7242	98.9944	98.4539	98.2591	98.9944	98.7242	98.5294	98.2591	99.1892	97.7186
	HOG										



(a) 图2 行人检测结果



(b) 图5(a)行人检测结果



(c) 图5(b)行人检测结果



(d) 图5(c)行人检测结果

图7 行人检测结果

Fig.7 Results of pedestrian detection

参考文献:

[1] Dollar P, Wojek C, Schiele B, et al. Pedestrian detection: An evaluation of the state of the art[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(4): 743-761.

[2] Hou X D, Zhang L. Saliency detection: a spectral residual approach[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA: IEEE, 2007: 1-8.

[3] Achanta R, Hemami S, Estrada F, et al. Frequency-tuned salient region detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA: IEEE, 2009: 1597-1604.

[4] Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA: IEEE, 2010: 2376-2383.

[5] Alexe B, Deselaers T, Ferrari V. Measuring the objectness of image windows[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(11): 2189-2202.

[6] 柯洪昌, 孙宏彬. 图像序列的显著性目标区域检测方法[J]. 中国光学, 2015, 8(5): 768-774.
KE Hong-chang, SUN Hong-bin. A saliency target area detection method of image sequence[J]. *Chinese Optics*, 2015, 8(5): 768-774.

[7] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, USA: IEEE, 2005(1): 886-893.

[8] Geronimo D, Lopez A M, Sappa A D, et al. Survey on pedestrian detection for advanced driver assistance systems[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(7): 1239-1258.

[9] Dollar P, Wojek C, Schiele B, et al. Pedestrian detection: A benchmark[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, USA: IEEE, 2009: 304-311.

[10] 张春风, 宋加涛, 王万良. 行人检测技术研究综述[J]. 电视技术, 2014, 38(3): 157-161.
ZHANG Chunfeng, SONG Jiatao, WANG Wanliang. Survey on pedestrian detection technology[J]. *Video Engineering*, 2014, 38(3): 157-161.

[11] 许腾, 黄铁军, 田永鸿. 车载视觉系统中的行人检测技术综述[J]. 中国图象图形学报, 2013, 18(4): 359-367.

[12] XU Teng, HUANG Tiejun, TIAN Yonghong. Survey on pedestrian detection technology for on-board vision systems[J]. *Journal of Image and Graphics*, 2013, 18(4): 359-367.

[13] 侯旺, 孙晓亮, 尚洋, 等. 红外弱小目标检测技术研究现状与发展趋势[J]. 红外技术, 2015, 37(1): 1-10.
HOU Wang, SUN Xiao-liang, SHANG Yang, et al. Present state and perspectives of small infrared targets detection technology[J]. *Infrared Technology*, 2015, 37(1): 1-10.

[14] Yang J C, Yu K, Gong Y H, et al. Linear spatial pyramid matching using sparse coding for image classification[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, USA: IEEE, 2009: 1-8.

[15] Bao C Q, He L T, Wang Y L. Linear spatial pyramid matching using non-convex and non-negative sparse coding for image classification[C]//*IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, China: IEEE, 2015: 186-190.

(上接第 741 页)

[2] 杨彦辉, 谢启明, 岳清, 等. 锗单晶二维六面扫描转鼓的数控精密成形[J]. 新技术新工艺, 2010(12): 94-96.
YANG Yanhui, XIE Qiming, YUE Qing, et al. CNC precision shaping for two-dimension polygon scanner of Ge crystal[J]. *New Technology & New Process*, 2010(12): 94-96.

[3] 金宁, 曹凌, 崔勤. 红外六面折射转鼓扫描轨迹的仿真分析[J]. 红外技术, 2007, 29(5): 249-252.
JIN Nin, CAO Ling, CUI Qin. Analysis for scanning footprint of infrared 6-face rotating drum system[J]. *Infrared Technology*, 2007, 29(5): 249-252.

[4] 郑子文, 翁雷, 陈善勇, 等. 高精度六面体垂直度误差测量的一种新方法[J]. 光学技术, 2008, 34(3): 391-394.
ZHENG Ziwen, WENG Lei, CHEN Shanyong, et al. A novel method for perpendicularity measurement of high precision hexahedron[J]. *Optical Technique*, 2008, 34(3): 391-394.

[5] 徐凤茹, 林玉池, 赵美蓉, 等. 基于视觉跟踪式自准直仪的十字线提取技术[J]. 激光与红外, 2011, 41(11): 1211-1214.
XU Fengru, LIN Yuchi, ZHAO Meirong, et al. Reticle extraction technology based on visual tracking autocollimator[J]. *Laser & Infrared*, 2011, 41(11): 1211-1214.