

妇科医学显微图像中乳酸杆菌的检测和统计分析

郭玉坤^{1,2} 马丽文^{1,2} 李金屏^{1,2}

(1. 济南大学信息科学与工程学院, 济南, 250022; 2. 山东省网络环境智能计算技术重点实验室, 济南, 250022)

摘要: 针对妇科分泌物光学显微图像成分复杂的特点, 提出了一种快速有效的检测并统计乳酸杆菌数量的算法。首先利用窗口 Laws 能量方法鉴别背景和背景之外其他成分的纹理特征, 从而保留乳酸杆菌所在的背景区域; 然后利用 Laws 能量对乳酸杆菌进行预分割, 并计算出分割区域内的平均灰度, 再结合背景的灰度值精确分割; 根据乳酸杆菌细胞的成像特点, 提取目标区域的长宽比、面积大小等特征滤除区域中的杂质; 最后对乳酸杆菌的数量进行统计分析。实验结果表明, 该方法能够有效检测复杂显微图像中的乳酸杆菌目标, 并获得较好的统计分析结果。

关键词: 图像处理; 医学显微图像; 乳酸杆菌; Laws 能量

中图分类号: TP391 **文献标志码:** A

Detection and Statistical Analysis of Lactobacillus in Gynecological Medical Micrographs

Guo Yukun^{1,2}, Ma Liwen^{1,2}, Li Jinping^{1,2}

(1. School of Information Science and Engineering, University of Jinan, Jinan, 250022, China; 2. Shandong Provincial Key Laboratory of Network Based Intelligent Computing, Jinan, 250022, China)

Abstract: Since the characteristics that optical microscopic images of gynecological secretions are very complex, an effective algorithm to detect and collect the number of lactobacillus is proposed. Firstly, the windowed Laws energy is used to identify the texture features of the background and other components, thus keeping the background region with lactobacillus. Secondly, the Laws energy is utilized to pre-segment the lactobacillus, and the average gray of the pre segmentation regions is calculated. Thirdly, with the combination of the background grey, the accurate segment is carried out. On the basis of the imaging characteristics of the lactobacillus, the values of area, length-width ratio and duty-ratio of target area are extracted to remove impurities. Finally, the statistical analysis is carried out on the number of lactobacillus. Experimental results demonstrate that the proposed method can effectively detect lactobacillus in complex gynecological medical micrographs and offer well-analyzed results.

Key words: image processing; medical micrograph; lactobacillus; Laws energy

引言

目前国内外关于医学图像分析的研究有很多,但关于妇科分泌物光学显微图像的自动成分分析尚未有临床应用,完全是依靠人工分析。随着人们健康意识的不断增强,人工分析的弊端日益突出:(1)人工分析需要较强的专业知识,需要经过专门训练;(2)由于待分析样本数量庞大,使得劳动强度大,人力成本过高;(3)主观性强,检测结果缺乏统一和客观标准。因此,亟需开发一套自动检测相关成分的算法和设备。通过观察妇科阴道分泌物(简称分泌物)的光学显微图像,可以看出其中的成分多且复杂。图1标识出图像中可能出现的其中几种成分(其他如:支原体、衣原体、红细胞等成分不再列出)。从图1看出,相同成分的形态也差异很大,而且有许多重叠、遮挡等现象,对于准确检测造成了极大困难。

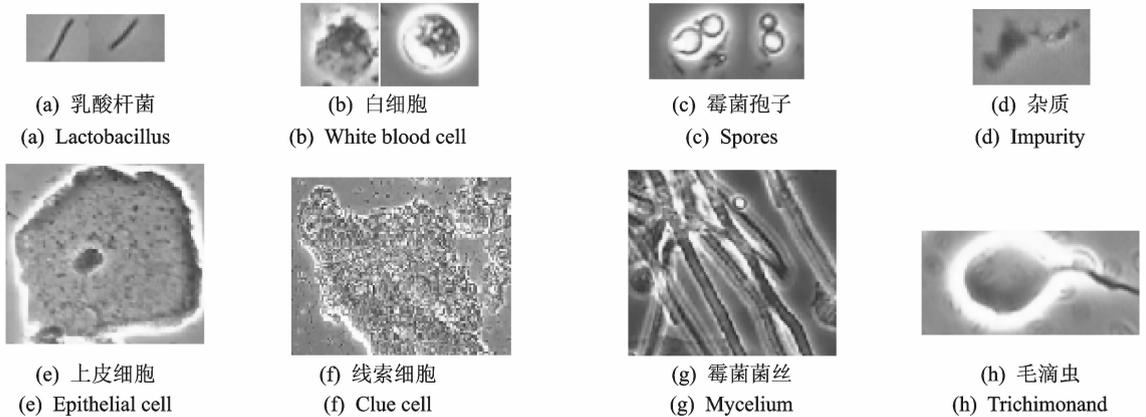


图1 分泌物显微图像中的几种成分

Fig. 1 Kinds of component in microscopic images of secretions

近几年,国内外学者做了很多与光学显微杆菌图像检测相关的工作。在针对结核杆菌检测方面的工作中,由于样本经过染色,使得结核杆菌与背景之间的颜色差异较为明显。研究人员多数从颜色通道入手,提出了很多检测方法。文献[1]基于红色和绿色通道使用全局自适应阈值分割,在RGB颜色空间中用 $R-G$ (即 R 通道减去 G 通道)使得结核杆菌细胞能够在图像背景中更为明显,再选取恰当的阈值进行分割,最后用数学形态学、颜色及面积大小去除杂质。文献[2]在HSI颜色空间中对像素分类,进行连通区域标记、区域面积大小阈值去除噪声、区域分类等方法检测杆菌目标。文献[3]分别在YCbCr与Lab两个颜色空间取出Cr与 a 通道进行阈值分割,进而提取杆菌。文献[4]则使用HSV颜色空间中的 H 与 S 通道,用Otsu自适应阈值算法进行分割提取目标杆菌。文献[5]针对蔬菜中的大肠杆菌快速检测问题,提出采用形态特征参数及染色后菌体区域颜色特征参数统计值对大肠杆菌进行快速识别,提取多个特征参数建立基于主成分的3层BP神经网络模型。文献[6]提出的基于颜色及梯度统计特征的结核杆菌识别算法,首先利用HSV颜色空间预分割,然后在Lab颜色空间进行自适应分割,最后提取多个形态特征用贝叶斯分类器进行分类。文献[7]针对萋尔-尼尔逊方法染色的显微图像,对结核杆菌的提取和识别进行研究,用形态学方法进行分割、提取特征,最后用BP(Back propagation)神经网络进行识别。除此之外,文献[8-9]研究了杆菌目标的检测。

上述研究对杆菌的分割识别都取得了一定进展,但其问题背景相对简单,图像中的成分较为单一,均是针对染色图像做处理,而染色过程中,染色剂的浓度、剂量都会对后续的分割识别产生影响;对于图像中含有大量其他细胞成分的情况没有给出具体的处理方法。由于本课题所针对的图像为妇科分泌物光学显微图像,其中除目标乳酸杆菌外还有很多其他成分,如:上皮细胞、白细胞、霉菌孢子、线索细胞和

杂质等。而且在妇科临床医学检验中上述的各种成分均需要在一幅图像中同时检测,故不能只针对乳酸杆菌进行染色。因此,无法根据颜色信息进行乳酸杆菌的检测,所以上述文献中的方案不能有效地解决本课题中的问题。经过调研国内外有关图像分割^[10-17]方法,针对目前的临床需求,本文提出了一种能够自动快速检测并统计分泌物显微图像中乳酸杆菌数量的有效算法。

1 窗口 Laws 纹理能量背景提取

对于背景的提取,由于乳酸杆菌目标较小,而其他成分面积较大,因此分布在乳酸杆菌上的 Laws 能量值相对其他成分较低,用此特征进行分割会取得较好的效果。

1.1 窗口 Laws 能量

窗口 Laws 能量可以有效描述目标的纹理特征,可表示为

$$\mu_w = \frac{1}{n^2} \sum_{i=-k}^k \sum_{j=-k}^k I(x+i, y+j) \quad (1)$$

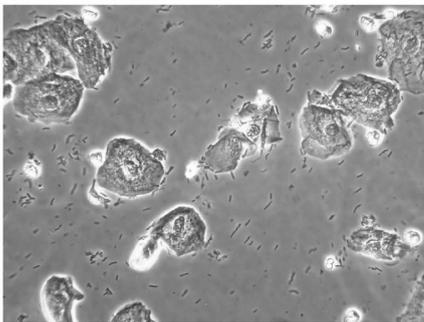
式中: μ_w 为图像中窗口 W 内的均值; n 为窗口 W 的大小 $n=2k+1$; $I(x, y)$ 为图像 I 在 (x, y) 位置的灰度值。

$$E(x, y) = \frac{1}{n} \sum_{i=-k}^k \sum_{j=-k}^k |I(x+i, y+j) - \mu_w| \quad (2)$$

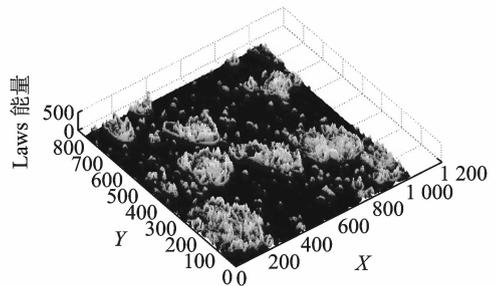
式中: $E(x, y)$ 为图像中以 (x, y) 为中心的窗口 W 中像素相对平均值的平均误差,也就是本文中所定义的在图像 (x, y) 位置上的 Laws 能量值。

1.2 数据分析与背景提取

由于处理过程中颜色信息无法利用,故先对图像进行灰度化,再计算图像的 Laws 能量分布。如图 2 (b)所示,图像中上皮细胞部分的 Laws 能量值比背景部分的能量值高出很多,同样乳酸杆菌附近的能量比背景区域也略微高一些,但明显低于上皮细胞所在的区域。



(a) 分泌物图像
(a) Secretion image



(b) Laws能量分布图
(b) Laws energy distribution map

图2 分泌物图像及其 Laws 能量分布图

Fig. 2 Secretion image and its Laws energy distribution map

为减少图像中细节损失,在提取背景时先用 Laws 能量对原图进行增强,再使用 Otsu 自适应阈值法对增强后的图像进行分割,高于阈值的像素为所要去除的成分,低于阈值的部分为背景。由于有些细胞成分的部分区域 Laws 能量值也会比较低,使得分割时会出现孔洞、断裂等情况,因此需要对分割后的图像进行填充孔洞、滤波等操作。再对区域进行标记,将区域面积小于 3~5 倍的杆菌细胞面积的区域去掉,最终获得提取背景所需的掩模图像,如图 3 所示。

2 多信息融的乳酸杆菌区域检测

通过大量的观察,乳酸杆菌的灰度值绝大多数情况是低于背景的灰度值(除非图像受到较为严重的离焦模糊的影响,这种情况较少出现,暂不考虑),本文采用 Laws 能量与统计直方图两种方法进行融合,最终得到较好的提取结果。

2.1 基于 Laws 能量的预分割

在 1.2 节中进行区域标记后,选取杆菌面积相近的区间保留,其他的丢弃。虽然这可将大部分的乳酸杆菌区域保留,但同时也无法将部分与乳酸杆菌形态相同的杂质进行去除。如图 4 所示,白色点状区域为疑似乳酸杆菌区域。通过观察可以看出虽然只能将部分乳酸杆菌区域分割出来,但同时也去除了背景中的大部分杂质。



图 3 提取背景的掩模图像

Fig. 3 Mask image for extracting background



图 4 Laws 能量预分割结果

Fig. 4 Pre-segment result by Laws energy

2.2 加权二类阈值联合分割

根据乳酸杆菌的灰度值低于背景的灰度值的特征,可以选取一个合适的阈值将其在背景中分割出来。为了能够得到合适的阈值,本文采用加权二类阈值联合分割方法进行精确分割,其中二类阈值是指欲分割目标的灰度均值与背景灰度均值,具体方法如下:首先,计算背景的平均灰度值 v_1 和 Laws 能量分割出来的图像均值 v_2 ,如图 4 所示。然后计算分割阈值

$$v = (1 - \frac{1}{\alpha})v_1 + \frac{1}{\alpha}v_2 \quad (3)$$

式中: α 为常数。经大量实验数据统计分析可知,当 α 取 $2/5$ 时,可以取得最佳的分割效果,故本文中 α 取值 $2/5$ 。为了更准确地计算出背景的平均灰度值 v_1 ,可将图 3,4 作“或”操作得到更精确的背景掩模图像。图 5 为用 v 分割的结果图。

从图 5 中可看出,由于光照不均,导致图像的左下方的杂质及背景也分割出来,通过后续的区域面积滤波(即利用杂质与背景部分的连通区域面积较大这一特征将其去除)可以消除这种影响。

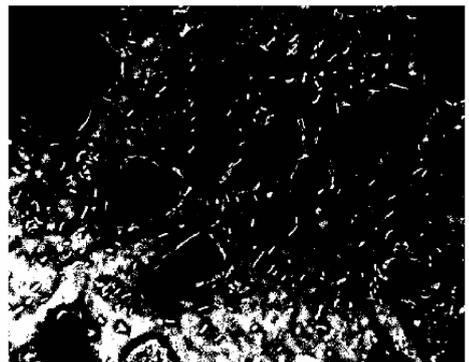


图 5 使用阈值 v 的分割结果

Fig. 5 Segment result by threshold v

3 乳酸杆菌的检测

按照上述得到的分割结果,可发现内仍有许多杂质,且杂质灰度与乳酸杆菌细胞的灰度相似,但形态上有所差别,如图 6 所示。根据乳酸杆菌与杂质形态上特征的差异,可提取面积大小、长宽比和占空比等特征去除杂质,为提高精度,还可以再加上区域内的灰度均值与背景灰度均值的差值(称为灰度差值)这个特征。



图 6 乳酸杆菌杂质

Fig. 6 Lactobacillus and impurities

图 5 的分割结果中各个区域的面积大小、占空比、长宽比和灰度差值的统计结果,如图 7 所示。图 5 中分割出区域数量为 655 个,为了检验各个特征去除杂质的效果,本文通过人工标定出来了其中所含有的乳酸杆菌数量为 70 个。

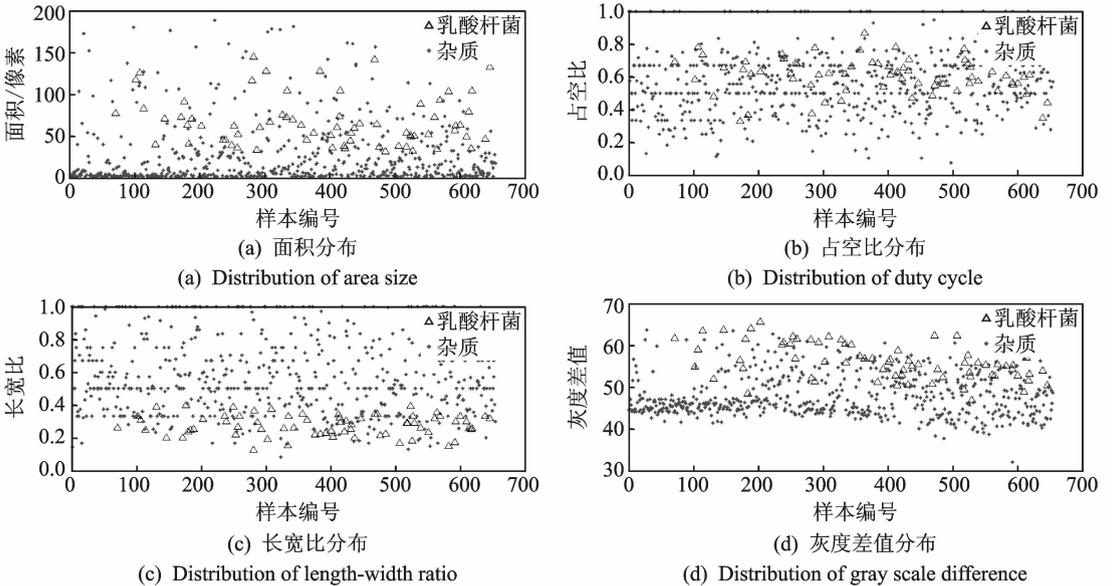


图 7 各个特征的统计结果

Fig. 7 Statistical results of characteristics

通过观察图 7 可以看出,面积、长宽比和灰度差值这 3 个特征均可以起到较好的去杂质结果,而乳酸杆菌与杂质在占空比这个特征中的分布没有较好的聚集性,分类效果很差,原因是由于杂质形状多样且形态复杂,仅仅利用占空比这一特征无法取得好的分类效果。因此,本文选用待识别区域的面积、长宽比和灰度差值这 3 个特征进行最终的识别特征因子。经大量的实验数据分析与验证,本文最终选用这 3 个特征因子参数进行乳酸杆菌的识别,并且取得较好的识别效果。各参数取值如表 1 所示。

表 1 各个特征的特征取值

Table 1 Parameter selection of features

面积大小/像素	长宽比	灰度差值
35~150	≤ 0.4	≥ 50

4 实验结果及分析

在 Intel 2.5 GHz 处理器,4 GB 内存,Windows 7 操作系统,Microsoft Visual Studio 2010 编译环境下实现上述算法。在 Matlab R2013b 下进行实验数据分析。为了检验本文算法的有效性,对大量图像进行了测试。其中 25 个样本的检测结果如表 2 所示。实际测试的样本数为 1 500 幅图像,其中有 1 413 幅能够达到医学临床标准。剩余的 37 幅图像中,有 14 幅由于相机离焦造成图像模糊,识别误差较大

表 2 乳酸杆菌的检测结果

Table 2 Detection results of lactobacillus

样本编号	耗时/ ms	检测出乳酸 杆菌数量/个	实际乳酸 杆菌数量/个	误检乳酸 杆菌数量/个	漏检乳酸 杆菌数量/个	临床判读 正误
1	209.005	21	23	2	4	正确
2	205.491	58	63	3	8	正确
3	207.185	32	30	2	0	正确
4	220.855	9	7	2	0	正确
5	218.935	22	24	1	3	正确
6	221.386	40	65	0	25	错误
7	208.398	4	4	0	0	正确
8	223.378	44	40	6	2	正确
9	206.036	43	46	2	5	正确
10	208.504	17	13	4	0	正确
11	199.101	3	17	1	15	错误
12	205.654	15	12	3	0	正确
13	210.668	69	73	3	7	正确
14	221.484	38	43	0	5	正确
15	201.867	27	34	1	8	正确
16	210.436	65	67	1	3	正确
17	204.538	5	3	2	0	正确
18	215.843	18	21	2	5	正确
19	224.657	86	81	7	2	正确
20	214.516	32	34	1	3	正确
21	207.428	19	18	1	0	正确
22	198.853	26	23	4	1	正确
23	204.731	21	25	1	5	正确
24	214.275	34	34	3	3	正确
25	209.362	31	36	1	6	正确

(如第 6 个样本);17 幅由于大量乳酸杆菌附着在上皮细胞上无法分割导致检测误差太大(如第 11 个样本);6 幅由于图像光照不均导致分割不足造成检测误差过大(由于此研究没有其他工作可作数据对比,因此表 2 中只做了和人工检测结果的对比)。实验结果表明,本文算法能够较为准确地检测出图像中的乳酸杆菌数量,能够达到医学临床的要求。

5 结束语

针对妇科分泌物光学显微图像中的乳酸杆菌的检测,采用 Laws 能量提取背景,并通过采用加权二类阈值联合分割的方法进行乳酸杆菌的分割,最后根据乳酸杆菌的面积大小、长宽比、灰度差值这 3 个特征值的范围检测出乳酸杆菌,并统计乳酸杆菌数量。实验结果表明,本文方法能有效地检测并统计出图像中的乳酸杆菌的数量,并且能够应用于自动化检测,减轻医师的负担。

参考文献:

- [1] Costa M G F, Costa F, Cicero F F, et al. Automatic identification of mycobacterium tuberculosis with conventional light microscopy [C]// Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vancouver, BC:IEEE, 2008:382-385.
- [2] Nayak R, Shenoy V P, Galigekere R R. A new algorithm for automatic assessment of the degree of TB-infection using images of ZN-stained sputum smear [C]// International Conference on Systems in Medicine and Biology. Kharagpur: IEEE, 2010:294-299.
- [3] Sotaquira M, Rueda L, Narvaez R. Detection and quantification of bacilli and clusters present in sputum smear samples: A novel algorithm for pulmonary tuberculosis diagnosis [C]// International Conference on Digital Image Processing. Bangkok: IEEE, 2009:117-121.
- [4] Makkapati V, Agrawal R, Acharya R. Segmentation and classification of tuberculosis bacilli from ZN-stained sputum smear images [C]// Automation Science and Engineering, CASE 2009, IEEE International Conference on. Bangalore: IEEE, 2009: 217-220.
- [5] 丁筠, 殷涌光, 王旻. 蔬菜中大肠杆菌的机器视觉快速检测[J]. 农业机械学报, 2012, 43(2): 134-139.
Ding Yun, Yin Yongguang, Wang Min. Rapid detection based on machine vision for escherichia coli in vegetables[J]. Transactions of the Chinese Society for Agricultural Machinery, 2012, 43(2): 134-139.
- [6] 翟永平, 周东翔, 刘云辉. 基于颜色及梯度统计特征的结核杆菌目标识别[J]. 国防科技大学学报, 2012, 34(5): 146-152.
Zhai Yongping, Zhou Dongxiang, Liu Yunhui. Recognition of mycobacterium tuberculosis in microscopic images based on color and gradient feature[J]. Journal of National University of Defense Technology, 2012, 34(5): 146-152.
- [7] 孙涵璇, 谢凤英, 姜志国, 等. 基于 BP 神经网络的结核杆菌目标识别[J]. 中国体视学与图像分析, 2010, 15(1): 13-17.
Sun Hanxuan, Xie Fengying, Jiang Zhiguo, et al. Identification of mycobacterium tuberculosis based on BP neural network [J]. Chinese Journal of Stereology and Image Analysis, 2010, 15(1): 13-17.
- [8] 徐超, 周东翔, 刘云辉, 等. 光学显微结核杆菌图像的区域分割算法 [J]. 国防科技大学学报, 2014, 36(5): 79-86.
Xu Chao, Zhou Dongxiang, Liu Yunhui, et al. A segmentation algorithm for optical microscopic mycobacterium tuberculosis based on region segmentation[J]. Journal of National University of Defense Technology, 2014, 36(5): 79-86.
- [9] 王慧莹. 基于计算机视觉技术快速检测食品中蜡样芽孢杆菌的研究[D]. 长春: 吉林大学, 2011.
Wang Huiying. Research on rapid detection of bacillus cereus in food by computer vision[D]. Changchun: Jilin University, 2011.
- [10] 刘松涛, 殷福亮. 基于图割的图像分割方法及其新进展[J]. 自动化学报, 2012, 38(6): 911-922.
Liu Songtao, Yin Fuliang. The basic principle and its new advances of image segmentation methods based on graph cuts[J]. Acta Automatica Sinica, 2012, 38(6): 911-922.
- [11] Laws K I. Texture images segmentation [D]. California: University of Southern California, 1980.

- [12] Malik J, Belongie S, Leung T, et al. Contour and texture analysis for image segmentation[J]. International Journal of Computer Vision, 2001, 43(1):7-27.
- [13] Chen X J, Udupa J K, Alavi A, et al. GC-ASM: Synergistic integration of graph-cut and active shape model strategies for medical image segmentation [J]. Computer Vision and Image Understanding, 2013,117(5): 513-524.
- [14] Zhang Hui, Zhu Quanyin, Guan Xiangfeng. Probe into image segmentation based on sobel operator and maximum entropy algorithm [C]//Computer Science and Service System (CSSS), 2012 International Conference on. Nanjing: IEEE, 2012:238-241.
- [15] Preetha M M S J, Suresh L P, Bosco M J. Image segmentation using seeded region growing [C]//Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on. Kumaracoil: IEEE, 2012: 576-583.
- [16] 王斌,李洁,高新波. 一种基于边缘与区域信息的先验水平集图像分割方法[J]. 计算机学报, 2012,35(5):1067-1072.
Wang Bin, Li Jie, Gao Xinbo. An edge and region based level set method with shape priors for image segmentation[J]. Chinese Journal of Computers, 2012, 35(5):1067-1072.
- [17] 唐利明,黄大荣,李可人. 基于变分水平集的图像分割模型[J]. 数据采集与处理, 2014,29(5):704-712.
Tang Liming, Huang Darong, Li Keren. New model based on variational level set for image segmentation[J]. Journal of Data Acquisition and Processing, 2014,29(5):704-712.

作者简介:



郭玉坤(1989-),男,硕士研究生,研究方向:数字图像处理和机器视觉等, E-mail: guoyk2009@outlook.com。



马丽文(1988-),女,硕士研究生,研究方向:数字图像处理和机器视觉等。



李金屏(1968-),男,博士,教授,研究方向:数字图像处理、机器视觉、模式识别和优化算法等。

