

细胞病理自动阅片关键技术

何勇军 著



科学出版社

细胞病理自动阅片关键技术

何勇军 著

科学出版社

北 京

内 容 简 介

本书系统地论述了细胞病理自动阅片的关键技术,内容涵盖作者近年来的最新研究成果。全书共5章,内容包括自动阅片系统结构及关键技术、显微镜聚焦与样本扫描、光照不均匀问题及其补偿、面向阅片环境的细胞图像分割、细胞图像类别划分与识别等。

本书可作为高等院校从事人工智能、显微镜自动聚焦以及细胞图像处理相关研究的硕士、博士研究生的参考书,也可从事计算机信息科学、人工智能和数据挖掘的科技工作人员和工程应用人员提供参考。

图书在版编目(CIP)数据

细胞病理自动阅片关键技术/何勇军著. —北京:科学出版社, 2018.8

ISBN 978-7-03-057542-5

I. ①细… II. ①何… III. ①细胞学-病理学-图象识别
IV. ①R361-39

中国版本图书馆CIP数据核字(2018)第100144号

责任编辑:王 哲 / 责任校对:郭瑞芝

责任印制:张 伟 / 封面设计:迷底书装

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

北京教图印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2018年8月第 一 版 开本:720×1000 B5

2018年8月第一次印刷 印张:10 1/2 插页:1

字数:214 000

定价:69.00 元

(如有印装质量问题,我社负责调换)

前 言

近年来,人工智能技术迅猛发展,在各行各业应用越来越广泛,受到了广泛的关注。国家也从战略层面提出了我国人工智能发展规划。人工智能和医疗结合对医疗信息管理、病理诊断以及医疗数据挖掘等都有着深远影响。自动阅片是人工智能和细胞病理诊断相结合的典型代表。传统的细胞病理诊断完全依靠医生手动操作显微镜平台,用眼睛在镜下视野搜索异常细胞。这不仅给医生带来了繁重的劳动,而且降低了诊断的准确性。自动阅片技术控制显微镜电动平台自主移动和聚焦,并采用工业相机拍摄到清晰的细胞图像,然后进行图像分割和识别,找出上皮细胞,并对细胞的各项参数进行精准测量,最后将异常细胞罗列出来供医生进一步确认。将医生从繁重的细胞搜索中解放出来,仅仅对有异议的细胞进行复核,有效提高了医生的诊断准确率。自动阅片技术的发展,将极大影响细胞病理诊断。

细胞病理自动阅片技术是一个多学科融合的先进技术,是人工智能和医疗相结合的典型范例。首先是清晰图像的获取,这是进行识别和细胞测量的第一步。显微镜下的图像经过高倍放大,微小的误差就会被放大。而整个载物台无法做到绝对平整,在安装时也无法做到与透镜径向方向垂直。此外,被观察的细胞层本身也存在起伏。这就使得不同的视野其焦平面并不在一个与物镜径向垂直的平面上。因此,聚焦是必要的。另一方面,每个位置都进行聚焦也忽略了相邻的若干视野可能在一个平面上的事实。因此还需要识别清晰程度,以决定是否需要聚焦。此外,对于一个标本的扫描,常常需要扫描三百个以上的视野,如何做到高效扫描也是需要解决的问题。

自动阅片系统的另一个关键技术是光照补偿。无论显微镜上是卤素灯或最新的LED灯,其光照不可能做到完全均匀。这对于基于灰度值的光学测量来说,将引起较大的误差,因此需要进行光照补偿。光照补偿在人脸识别等领域已经有比较广泛的研究。然而这些方法着眼的任务是分割或者识别,而不是测量,必然存在较大误差。因此,如何针对光学测量任务,提出有效的光照补偿方法也显得特别重要。

由于病理图像的细胞存在着较多的重叠现象,而某些严重的细胞病变常常表现为重叠细胞存在问题。因此细胞分割是识别与测量的必备前提。细胞分割不同于一般的图像分割任务。一方面,细胞存在细胞核和细胞质,二者需要分离以计算细胞特征。另一方面,细胞与细胞之间存在重叠。此外,图像中存在大量的垃圾杂质以及染色不均匀现象。这些都给细胞分割带来了挑战。系统性能的优劣在很大程度上取决于细胞分割的好坏。

病理图像中的物质种类繁多。既有各种细胞，包括上皮、中性粒、淋巴等不同细胞，还有这些细胞以各种方式和组合形成的团状细胞，以及各种垃圾杂质等。对这些物质要分别予以处理。正确识别各类物质是分别处理的前提，因此需要进行有效的识别。识别做得越精细，对自动阅片系统的支撑越充分。

本书重点论述上述关键技术，其内容自成体系。全书共 5 章。第 1 章是自动阅片系统结构及关键技术，介绍自动阅片系统的系统结构，并总览了该系统所需要的关键技术。第 2 章是细胞图像采集，主要论述清晰的细胞图像采集方法。其中包括全自动显微镜下的聚焦和样本扫描方法。第 3 章是光照补偿，论述显微镜下图像存在的光照问题以及补偿方法。第 4 章是细胞图像分割，以自动阅片为背景，论述细胞图像的分割方法，给出一个统一的图像分割框架。第 5 章是细胞图像识别，论述细胞图像的识别方法。

本书的撰写是作者及其团队集体共同努力完成的，需要特别感谢的是研究生梁隆恺、赵晶、卢玉、余莲、卢祎、张雪媛、郭云雪等。感谢谢怡宁老师和黄金杰老师予以的大力支持。由于作者水平有限，本书难免存在不妥之处，希望广大读者批评指正。

目 录

前言

第 1 章 自动阅片系统结构及关键技术	1
1.1 细胞图像采集	3
1.2 光照补偿	4
1.3 细胞图像分割	5
1.4 细胞图像识别	6
1.5 本章小结	7
参考文献	7
第 2 章 细胞图像采集	8
2.1 引言	8
2.2 相关技术概述	11
2.2.1 光学成像系统原理	11
2.2.2 自动聚焦方法	13
2.3 一种新的显微镜快速聚焦方法	15
2.3.1 现有的清晰度评价函数	16
2.3.2 新提出的清晰度评价函数	18
2.3.3 变步长爬山法	21
2.3.4 实验与分析	22
2.4 基于图像清晰度识别的快速扫描方法	24
2.4.1 引言	25
2.4.2 基于清晰度识别的快速聚焦扫描方法	26
2.4.3 实验与分析	31
2.5 基于多点聚焦的快速扫片方法	40
2.5.1 引言	40
2.5.2 三点聚焦快速扫片方法	41
2.5.3 改进的多点聚焦快速扫片方法	42
2.5.4 实验与分析	44
2.6 本章小结	45

参考文献	46
第3章 光照补偿	50
3.1 引言	50
3.2 现有的光照补偿方法	50
3.2.1 灰度变换法	50
3.2.2 同态滤波方法	51
3.2.3 Retinex 算法	51
3.2.4 基于梯度域的增强方法	52
3.2.5 基于反锐化掩模的增强方法	53
3.3 DNA 倍体分析技术与朗伯比尔定律	53
3.4 技术路线	56
3.4.1 方法流程	56
3.4.2 图像采集	57
3.4.3 分块阈值分割	57
3.4.4 背景填充	57
3.4.5 平滑	58
3.4.6 偏差估计与补偿	58
3.5 实验方案结果对比	60
3.5.1 实验设置	61
3.5.2 实验效果	61
3.5.3 实验与分析	63
3.6 本章小结	65
参考文献	65
第4章 细胞图像分割	67
4.1 引言	67
4.2 相关技术概述	70
4.2.1 DNA 倍体分析技术原理	70
4.2.2 图像分割技术	72
4.2.3 GMM 模型	74
4.2.4 图像修复方法	75
4.3 复杂背景下的宫颈细胞核分割方法	76
4.3.1 基于参数自适应的局部阈值法	76

4.3.2	分水岭算法的优化	77
4.3.3	粗分割算法	78
4.3.4	实验与分析	79
4.4	基于识别的细胞精细分割方法	84
4.4.1	分类器训练	85
4.4.2	精细分割	87
4.4.3	细胞核分割算法	88
4.4.4	实验与分析	89
4.5	重叠细胞分割中异常区域的重构	95
4.5.1	引言	95
4.5.2	模型训练	96
4.5.3	算法实现	98
4.5.4	实验与分析	101
4.6	本章小结	105
	参考文献	106
第 5 章	细胞图像识别	110
5.1	引言	110
5.2	宫颈细胞图像的细胞学基础	112
5.2.1	获取宫颈细胞	113
5.2.2	宫颈细胞分类	114
5.2.3	宫颈细胞图像识别方法框架	115
5.3	宫颈细胞图像预处理及分割方法	117
5.3.1	图像增强方法	117
5.3.2	宫颈细胞图像滤波方法	121
5.3.3	宫颈细胞图像分割方法	123
5.3.4	实验与分析	126
5.4	图像特征提取与选择	129
5.4.1	宫颈细胞图像特征提取	129
5.4.2	NF 特征集合	129
5.4.3	PF 特征集合	132
5.4.4	ReliefF 算法	133
5.4.5	实验与分析	135
5.5	两阶段宫颈细胞图像分类器	137

5.5.1 两级分类器融合.....138

5.5.2 实验与分析140

5.6 基于随机性的成团细胞核图像合成方法.....144

5.6.1 细胞类别划分.....144

5.6.2 合成方案144

5.6.3 实验与分析150

5.7 本章小结154

参考文献.....155

彩图

第1章 自动阅片系统结构及关键技术

随着计算机技术的飞速发展,人工智能技术已经广泛应用在人们生活的各个方面。在生物医学领域,日益成熟的图像处理技术被广泛应用,计算机辅助诊断技术推动了生物医学的发展。细胞核图像的DNA定量分析和病变细胞识别是医学图像处理技术的研究重点。它采用图像处理和机器学习技术定量分析人体特定部位的脱落细胞或组织切片。采用特征提取技术提取细胞核图像的各种特征参数,然后通过训练分类器来判断细胞是否发生癌变,并对病变细胞进行识别和分级。

中国人口众多,使得国人的肿瘤防控数据对全球癌症防控意义重大,全球新发癌症病例约有22%出现在中国,其中27%的癌症死亡病例也在中国。在全球范围内,宫颈癌是排名第4位的恶性肿瘤,也是威胁女性健康的主要问题。2012年全球宫颈发病数为52.8万例,年死亡数为26.6万例。其中85%的宫颈癌病例在发展中国家发生,而这些地区因癌症死亡的主要原因也是宫颈癌。数据显示,2013年发展中国家子宫颈癌的发病率和死亡率分别是发达国家的1.64倍和2.10倍。中国一直在努力降低子宫颈癌的发病率和死亡率,至今中国的宫颈癌发病率和死亡率都得到了大幅度的下降^[1-3]。

宫颈癌的发病时间较长,宫颈细胞从早期的病变到晚期的恶化需8~10年。宫颈癌变一般有四个阶段:宫颈不正常增长、宫颈原位癌、宫颈早期浸润癌、宫颈浸润癌。癌变的前三个阶段,病变较缓慢而且症状并不明显,但是从第三到第四阶段,病变明显、时间较短而且有严重的病痛。宫颈癌是容易预防的,这与其他癌症不同。因为宫颈在早期病变时的治疗效果很好,比晚期的治疗效果高很多。所以,痊愈宫颈癌的关键就是定期检查,并在发现细胞病变时对症治疗。

宫颈癌有很多筛查方法,宫颈脱落细胞涂片检查是最普遍、最常见的宫颈癌前病变和早期宫颈癌检查的方法,这种宫颈癌早期筛查的细胞学检验手段是肿瘤防治史上最重要的成就之一。该方法需要经验丰富的病理医生通过镜下观察发现病变细胞后做出诊断。然而,随着癌症的频发,该技术已无法适应现实需求。一方面,该诊断需要医生经过长期的专门训练,对医生要求较高;另一方面,医生在镜下观察完全根据经验,具有主观性,容易因视觉疲劳而导致误诊率上升。近年来发展起来的自动阅片技术能有效解决这些问题。该技术只对细胞染色,显微镜在软件的控制下自动聚焦扫描,拍摄镜下图像,并在识别的基础上准确测量细胞的各项参数,最后将异常的细胞核罗列出来,辅助医生诊断。通过该技术,医生只需接受短时间的培训,仅需复核系统挑选出的病变细胞核即可做出诊断,诊断准确率更高。

经过三年的努力，我们自主研发了一台细胞病理自动阅片系统系统。整个系统的组成如图 1-1 所示。计算机通过两个端口分别控制工业摄像机和控制盒，控制盒可以控制平台的移动和聚焦。

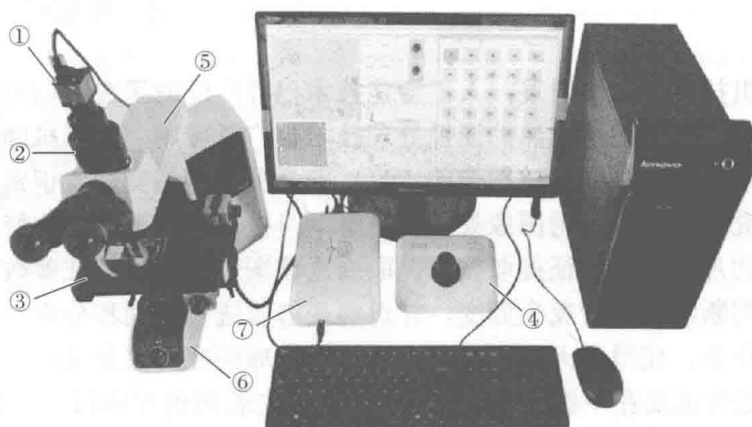


图 1-1 系统组成示意图

- ① 高清数码摄像机；② 数码摄像机适配器；③ 电动载物台；④ 控制遥杆；
⑤ 显微镜；⑥ 显微镜光源盒；⑦ 控制盒

计算机运行环境：CPU:i5-4460s，内存：8G，操作系统：Windows 7，编程语言：VC++。其中，电动载物台为自主开发，通过两个步进电机控制平台，使平台可以在水平方向进行前后左右四个方向的移动，如图 1-2 所示。

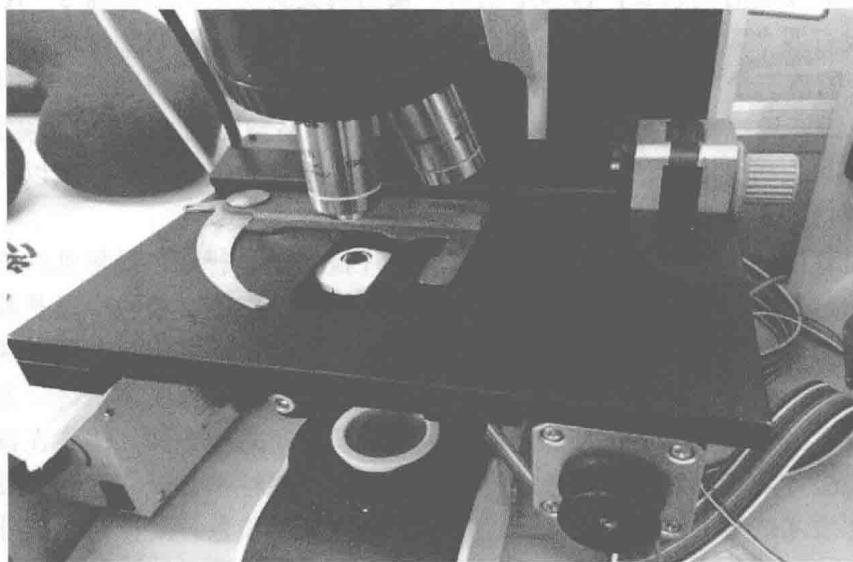


图 1-2 电动载物台

细胞 DNA 显微分光图像自动分析仪结构如图 1-3 所示。系统工作原理：首先光
此为试读, 需要完整PDF请访问: www.ertongbook.com

源通过滤光片照射置于载物平台上的玻片, 图像通过显微镜由数码摄像机采集, 然后将图像通过 USB 上传至计算机, 计算机计算图像的清晰度后通过移动 Z 轴改变平台到显微镜物镜的距离, 最终移动到一个清晰图像的位置。

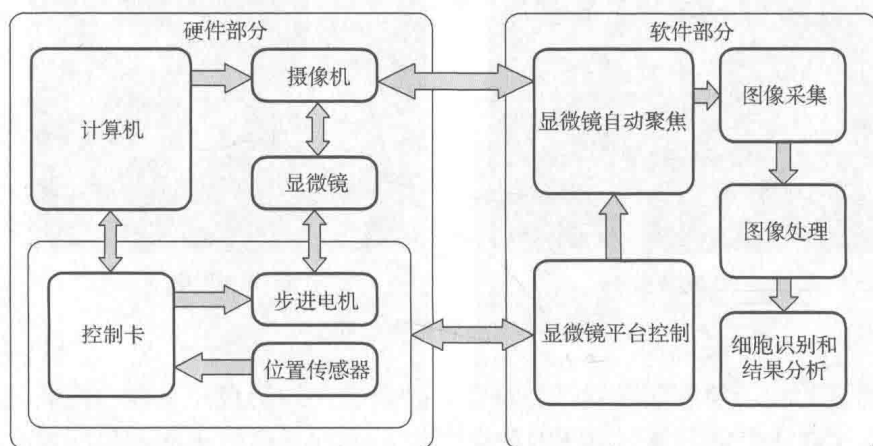


图 1-3 系统结构图

在显微镜对细胞标本进行扫描的时候, 由于摄像头可视范围有限的原因, 不能一次将整个细胞标本的图像采集下来, 所以需要进行多次采集才能将获取整个细胞标本的图像, 然后进行图像分析。扫描时的移动是通过移动电动载物台来实现摄像头的相对移动, 移动使用的是 X 轴电机和 Y 轴电机。

该系统需要的关键技术包括细胞图像采集、光照补偿、细胞图像分割和细胞图像识别, 下面分别简要介绍。

1.1 细胞图像采集

图像采集的任务在于获取清晰的镜下细胞图像, 供分析软件进行分析。这项技术包含两个重要的步骤: 扫描和聚焦。

标本玻片被放置在载物台上, 电动平台在软件的控制下, 带动玻片移动, 使得相机可以拍摄到标本上所有的细胞。由于载物台与显微镜观察方向不是绝对垂直的、载物台制作存在误差以及玻片制作存在不平整等问题, 所以当平台移动时被观察物体并不总是在焦平面上。而且显微镜下的图像是经过高倍放大的, 所以细微距离的变化都会导致镜下图像变模糊。如图 1-4(a) 所示, 如果被观察视野不在焦平面上, 即处于离焦状态, 图像是模糊的, 无法进行正常分析。通过聚焦后图像变得清晰, 如图 1-4(b) 所示。因此, 平台每次移动后, 都需要通过聚焦调整玻片到物镜的距离, 使采集的图像是清晰的。

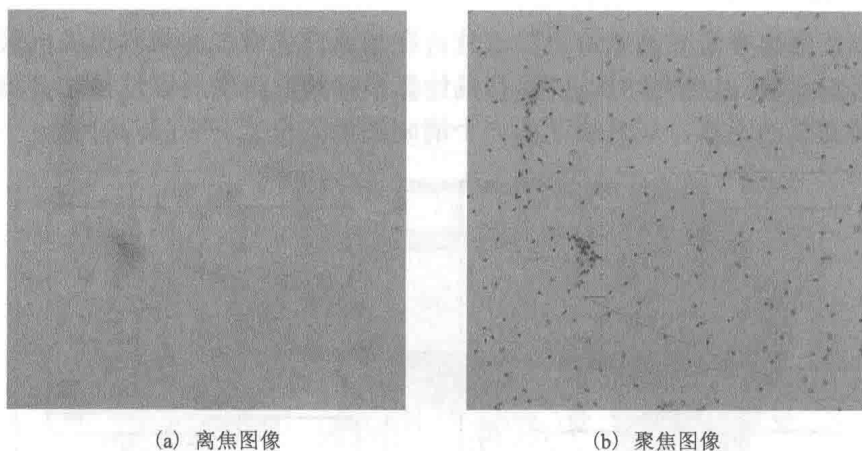


图 1-4 离焦和聚焦的图像

扫描的目的在于控制显微镜平台移动，使得镜头遍历玻片上被扫描区域的各个位置，并通过自动聚焦采集到清晰的图像^[4-6]。标本在显微镜下的图像经过高倍放大之后，在电子摄像头中仅能看到很有限的一部分区域，然而医学的诊断需要大量的图像才能保证诊断结果的可靠性。因此，需要扫描整个样本玻片，通过不断移动平台的位置，采集大量的样本图像。如图 1-5 所示，一个样本需要扫描的图像为 332 幅，如何利用相邻视野之间的关系减少聚焦移动的次数以提高扫描效率是需要解决的关键问题之一。

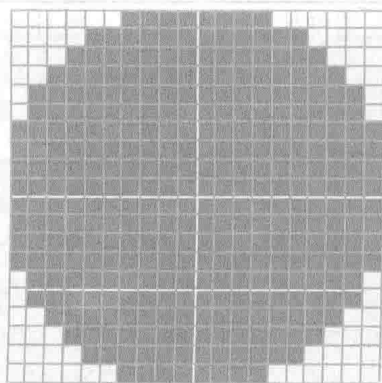
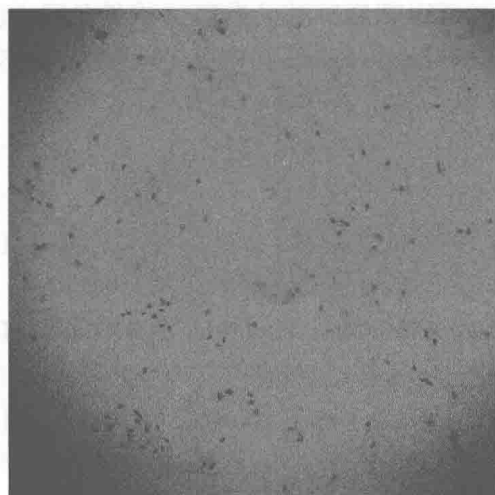


图 1-5 扫描区域图

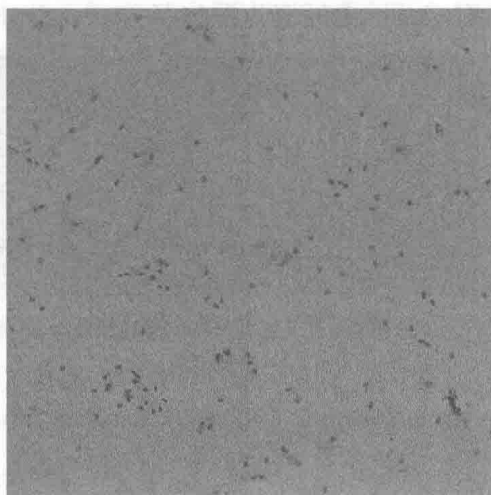
1.2 光照补偿

对细胞 DNA 相对含量的准确测量是倍体分析技术实现准确诊断的先决条件。然而，显微镜无论采用卤素灯还是 LED 灯，都不可避免地存在光照不均的问题。如

图 1-6(a) 为光照不均匀图像。可以看出, 图像的四周偏暗, 中间偏亮。这使得依赖于灰度值的光学测量在不同位置测量相同细胞的 DNA 含量, 其结果也不相同, 甚至相差甚远。这将直接增加系统的测量误差, 降低倍体技术测量结果的可信程度。因此, 必须通过有效的光照补偿减弱或消除光照不均带来的影响, 以获得图 1-6(b) 所示的均匀图像。



(a) 光照不均图像



(b) 光照均匀图像

图 1-6 光照对图像的影响

1.3 细胞图像分割

自动阅片技术的实质就是计算机辅助阅片的 DNA 倍体分析技术, 该技术首先通过照相机采集显微镜下的细胞核图像。然后利用图像分割技术将图像中的细胞核分割出来, 并利用特征提取技术提取细胞核的形状、颜色、积分光密度等特征参数。最后根据细胞的生物病理学知识和模式识别方法来定量分析和识别病变细胞。而在整个图像处理和识别的过程中, 细胞图像分割技术是定量分析和识别的基础, 可以为其提供大量的数据支撑。每个步骤之间是相辅相成的, 且是单向连接的, 而细胞分割是后面步骤的输入。所以细胞分割的结果越精确, 后续操作的准确率就越大。因此, 在自动阅片技术中, 细胞分割环节占据着不可或缺的重要地位^[7-9]。

然而, 计算机辅助阅片有复杂的应用环境, 如图 1-7 所示。首先, 复杂的背景条件, 即光照不均、背景阴影、染色深浅不一致等问题。其次, 脱落宫颈细胞采集不可避免地存在微生物、细胞碎片、血液、污染物等杂质, 使标本中存在大量的碎

片、黑斑、丝状絮状垃圾和一些聚集成团的腺细胞核等垃圾杂质。最后，细胞核在沉降时会出现层叠，导致图像中存在大量的重叠细胞核。重叠细胞核分割问题对细胞分割方法提出了新的挑战。如果我们可以在图像分割阶段就将这个复杂条件排除，并尽可能地分割重叠的细胞，那么就会得到更好的训练数据，并得到更高的识别率。而且在测量细胞 DNA 含量时可以更准确，最重要地是提高了自动阅片系统的诊断准确率，可以更好地为医生服务。

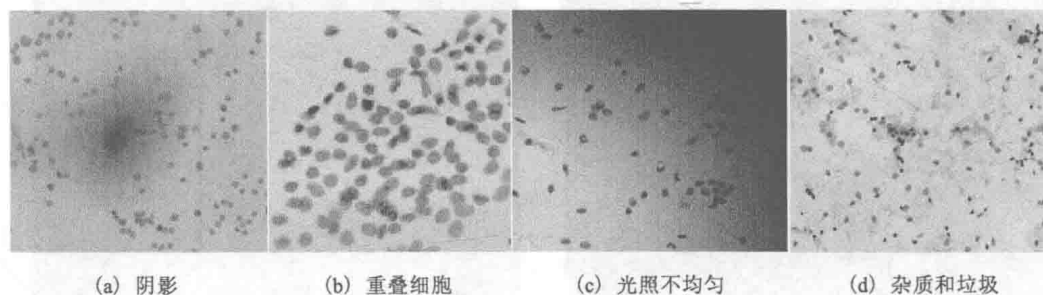


图 1-7 细胞核图像的复杂条件

1.4 细胞图像识别

自动阅片系统需要将细胞核图像进行分类，在类别划分要考虑两个方面的问题。一方面，不同细胞类别图像处理方式不同，因此需要对细胞图像分类精细。另一方面，要确保所有相似的样本被分为一类，以减少分类器正确识别细胞的难度。鉴于此，本书将细胞分为 8 类(如图 1-8 所示)，分别是：单个典型上皮细胞、单个非典型上皮细胞、2 个上皮细胞、3 个上皮细胞、4 个及以上上皮细胞、淋巴细胞和固缩核、单个中心粒细胞、2 个及以上中心粒细胞^[7,10]。

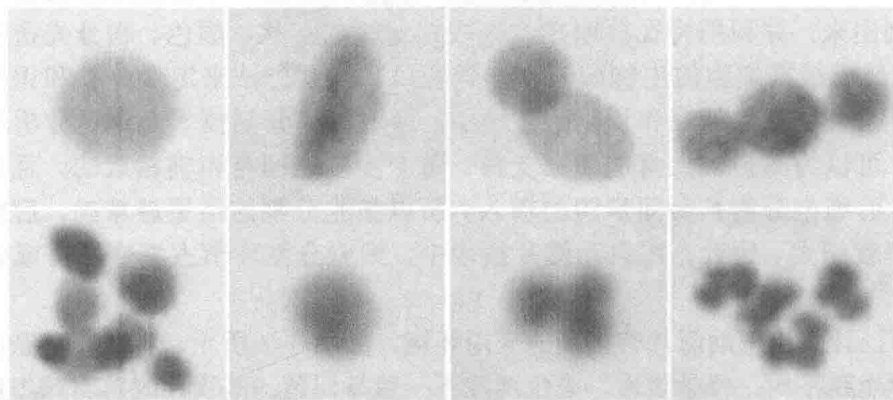


图 1-8 各类细胞图像

1.5 本章小结

本章介绍了自动阅片系统的体系结构。该系统是一个综合光学、机械、电控和软件以及病理诊断等技术的综合系统。其关键技术包括清晰细胞图像的获取、不均匀光照的补偿,细胞分割和细胞识别等。本书将集中在这些关键问题的解决上,为实现鲁棒准确的病理自动阅片技术奠定基础。

参考文献

- [1] Chen W, Zheng R, Baade P D, et al. Cancer statistics in China, 2015. *Ca: A Cancer Journal for Clinicians*, 2016, 66(2):115-132.
- [2] 周晖, 刘昀昀, 林仲秋. 《2017 NCCN 宫颈癌临床实践指南》解读. *中国实用妇科与产科杂志*, 2017, 33(1):100-107.
- [3] Fitzmaurice C, Dicker D, Pain A, et al. The global burden of cancer 2013. *Jama Oncology*, 2015, 1(4):505.
- [4] 梁隆恺. 基于环境感知的显微镜自动聚焦方法. 哈尔滨: 哈尔滨理工大学, 2017.
- [5] 梁隆恺, 赵晶, 何勇军. 一种显微镜自动聚焦算法. *哈尔滨理工大学学报*, 2018, 2: 46-52.
- [6] 卢祎, 张雪媛, 何勇军. 基于焦平面估计的快速扫片方法. *哈尔滨理工大学学报*, 2018, 7(2):120-128.
- [7] Xie Y N, Yu L, He Y J. An overlapping cell image synthesis method for imbalance data. *Analytical Cellular Pathology*, 2018, 6(3):156-163.
- [8] 赵晶, 何勇军, 梁隆恺, 等. 复杂背景下的宫颈细胞核分割方法. *哈尔滨理工大学学报*, 2018, 6(1): 100-110.
- [9] 赵晶. 面向自动阅片系统的细胞分割方法. 哈尔滨: 哈尔滨理工大学, 2017, 4(2): 130-134.
- [10] Lu Y, He Y J. Illumination compensation for microscope images based on illumination difference estimation. *Neurocomputing*, 2018, 8(4): 1025-1033.

第2章 细胞图像采集

2.1 引言

自动阅片的速度决定了医生诊断所花费的时间，而在一般的阅片系统中，采集图像的部分几乎占据了 80% 的时间。因此在阅片系统中，图像采集的速度是非常重要的。现有的阅片系统都是采用先移动平台，再自动聚焦，最后采集图像的方法。这种方法主要时间花费在自动聚焦上，导致整体效果比较慢。提升自动阅片的速度，可以节约医生等待机器的时间，提升医生的诊断效率。快速的扫片方法也能对显微镜下的图像识别提供有力的支撑。因此，对现有的扫片方法进行提升，是非常重要的且有意义的。

目前自动聚焦技术的方法主要可以分为两大类：第一类主动聚焦方法通过测量镜头与被拍摄物体之间的距离，然后将镜头移动到焦点位置达到聚焦的目的。这类方法通常要依赖于测距方法，复杂度高、成本大且难以实现。第二类被动方法以图像信号为反馈，通过比较不同位置图像清晰度的变化趋势实现自动聚焦。随着图像处理技术的进步，被动聚焦技术被广泛应用于自动显微镜中。常见的清晰度评价函数^[1,2]有绝对方差函数^[3]、灰度差分绝对值之和算子^[4,5]、拉普拉斯算子^[6]、能量谱方法^[7]等。国外的自动聚焦起步较早，研究成果较为丰富。其中主动聚焦的方法以测距为基本技术，美国 Honeywell 公司研制了 VAF 系统，该方法采用两个测距窗口，通过后面的三棱镜比较反射出来的内容完成自动聚焦。宝丽莱公司在 SX-70 摄像机上首次采用了超声波自动聚焦系统，通过计算反射时间来确定焦点位置。日本的佳能照相机最先推出了基于红外线测距的 AF35M 照相机，其原理类似 SX-70，用红外线替代了超声波测距。这类方法首先要有相关的硬件支持，然后还要精确安装，所以这种方法的造价极高，难以普及。

被动聚焦技术主要是基于图像处理的聚焦方法，这种方法相对于主动聚焦，不需要额外的硬件设备，因此受到了广泛的应用，并有大量研究者做出了贡献，提出了各种各样的聚焦函数。

基于梯度的聚焦函数，通过计算图像在某种方式下的梯度幅值的统计量获取图像聚焦信息，Tenenbaum 等^[8]在基于图像灰度梯度的算法上加入阈值，利用图像边缘能量进行图像清晰度评估。Jarvis 等^[9]在原有的基础上提出了改进，同时计算图像水平梯度和竖直梯度方向灰度差的绝对值之和作为评价函数。2016 年，Nayak 等