

引用格式: GAO Junzhao, HUANG Dangfei, ZHANG Lechao, et al. Cancer Tissue Recognition by Muller Matrix Imaging Full Array Curve[J]. Acta Photonica Sinica, 2023, 52(2):0210003

高君朝, 黄丹飞, 张乐超, 等. 基于穆勒矩阵成像全阵元曲线的癌变组织识别[J]. 光子学报, 2023, 52(2):0210003

# 基于穆勒矩阵成像全阵元曲线的癌变组织识别

高君朝<sup>1,2</sup>, 黄丹飞<sup>1,2</sup>, 张乐超<sup>1,2</sup>, 宋东<sup>3</sup>, 洪景辉<sup>3</sup>, 张丽丽<sup>1,2</sup>,  
唐鸿宇<sup>1,2</sup>, 周尧<sup>1,2</sup>

(1 长春理工大学 光电工程学院, 长春 130022)

(2 长春理工大学中山研究院, 中山 528437)

(3 吉林大学第一医院, 长春 130021)

**摘要:**改进原有背向穆勒矩阵成像装置,采集未染色肺癌、乳腺癌切片,基于穆勒矩阵各阵元的物理含义,借鉴光谱数据生成偏振立方体,纵向提取任意像素全阵元曲线,展示曲线特有的穆勒矩阵信息分析方式。采用夹角余弦曲线分类方法,对照病理标注生成混淆矩阵。其中肺癌、乳腺癌精度可达 89.59%、87.82%,高倍率乳腺癌精度为 77.52%。全阵元曲线较单一阵元分类准确率更高。本文方法的可视化分析、像素分类、跨尺度特性表明在偏振检测领域有很大的挖掘空间以及应用潜力。

**关键词:**偏振成像;穆勒矩阵;癌变组织;特征提取;病理切片检测

中图分类号:O436.3

文献标识码:A

doi:10.3788/gzxb20235202.0210003

## 0 引言

根据国际癌症研究机构统计数据,全球新增癌症人数共计 1 929 万人左右,中国新增癌症患者 457 万人,占比 23.7%,成为癌症新增人数最多的国家<sup>[1]</sup>。癌症作为全球第二大常见死亡原因,已成为当前全球共同面对的持续性公共卫生挑战<sup>[2]</sup>。现如今,癌症最成熟的治疗方法为手术切除,判断是否切除干净依靠人工病理切片检测。手术过程需要反复确认病变组织是否切除干净,每次检测都需要制作病理切片人工检测。病理切片制作繁琐耗时,通常将病变组织包埋在石蜡块里,用切片机切成薄片,再用苏木精-伊红(H-E)染色。人工切片检测确认过程时间长效率极低。因此临床急需一种跳过染色过程、精确、快速、高灵敏、非接触、无损伤、高安全性等特点的辅助检测手段。

偏振是独立于光强、波长、相位之外的一个新的信息维度。偏振成像光学检测技术基本步骤为利用入射光照射样品探测出射光,所获取入射光和出射光的差异即为样品的物理信息。偏振成像是一种无创、无标记、非接触的光学检测技术。常由穆勒(Mueller)矩阵来表征探测的偏振信息。已经被证明穆勒矩阵成像包含丰富的生物微观结构信息,应用于各种病理组织<sup>[3]</sup>。穆勒矩阵信息量大,不仅包含非偏振的光强变化信息,还能提供额外 15 个阵元的信息。穆勒矩阵成像跨尺度的特性,即使在低分辨率和宽场条件下也能获取微观结构编码丰富的信息<sup>[4]</sup>。成像结果既有较大的视野范围,同时还能反映微观结构信息的分布情况。

通过斯托克斯矢量获得穆勒矩阵的 16 个元素,必须在不同偏振调制下测量至少 16 个独立的强度图像<sup>[5]</sup>。穆勒矩阵求解采用基于双波片旋转法的穆勒矩阵 30 次测量装置<sup>[6]</sup>,能够达到的精度在  $10^{-2}$  量级。尽管偏振检测技术具有信息大、兼容强、尺度广等优势,但穆勒矩阵的各个阵元值同样品的微观结构没有直接的联系,混合了大量影响因素(散射、传播、样品角度、厚度等等),不易表征。要得到样品微观结构特征的特异性指标,目前研究的较多信息提取主要在两个方面,一方面从物理学的基本原理出发,建立起严谨的理

基金项目:国家自然科学基金(No.61893096014)

第一作者:高君朝,642283821@qq.com

通讯作者:黄丹飞,huangd\_f@163.com

收稿日期:2022-09-23;录用日期:2022-11-28

<http://www.photon.ac.cn>

论,提出分离<sup>[7]</sup>或者变换的方法,此方法对数据成像质量要求很高,分解变换出有用信息但也同样保留了干扰信息,并且信息展示多局限于二维空间差异信息提取。另一方面,直接跳过解释过程,制备大量数据,借助机器学习端到端寻找特异性指标<sup>[8]</sup>,但机器学习对特征提取过程解释性差,训练繁琐耗时过长,并且不适用原有穆勒矩阵的阵元分析方式。

通过研究的不断深入,穆勒矩阵阵元存储的偏振物理学信息及其影响逐渐被解读。本文将从改造原有背向穆勒矩阵成像系统光源以及入射方向入手,提高成像质量。并继续采用双波片旋转30次的测量方式生成穆勒矩阵,但区别于以往数据分析和处理方式。借助整理归类的穆勒矩阵物理信息,链接全新的排列次序,类比遥感领域光谱数据处理方法生成穆勒矩阵数据立方体。纵向以像素为单位提取全阵元曲线。全阵元曲线调用整体偏振特征,同时也方便开展单独偏振性质分析。借鉴光谱角制图用于抵消光路夹角带来系统误差的方法,遍历整个二维空间实现目标分类,此方法耗时短易于实现。在此基础之上,样品选用未染色肺癌以及乳腺癌病理切片,依照上述方法,开展对信息可视化分析,特征提取分类,跨尺度实验。结果证明,区别于常用的二维图像观测方式,像素级全阵元信息观测目标选取更灵活,观测形式更多变,抗干扰能力更强,对癌变非癌变及背景区分度好精度较高。跨尺度特性实现低倍率成像依旧能完成特征提取。跳过染色过程以及辅助人工检测,对临床缩短病理检测时间有十分重要的意义。

## 1 材料和方法

### 1.1 装置及原理

实验室搭建于光学平台的双波片偏振成像穆勒矩阵检测系统,配合波片旋转采集30组光强图像生成穆勒矩阵。该装置主要开展病理切片癌变检测原理验证,已实现偏振信息提取,穆勒矩阵求解分解等功能<sup>[9]</sup>。保留原有机电转动调制的方法,与相位延迟调制相比,不易受温度影响。但是原有系统光源采用裸漏LED灯珠焊接,没有灯罩和聚光镜片导致出射散光漏光干扰严重。原有光路斜入射的图像采集方式成像重影和干扰很大。综上,经改进之后背向穆勒矩阵测量实验装置及原理图如图1所示。

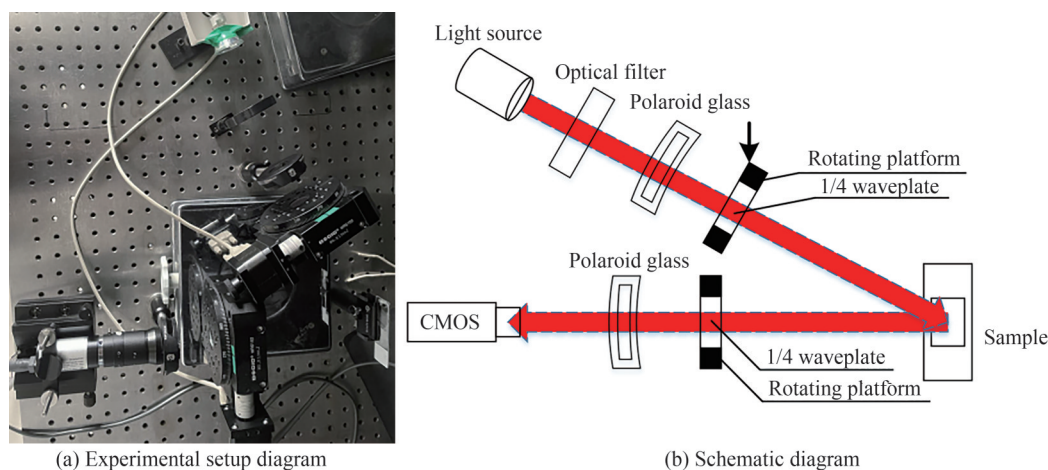


图1 背向穆勒矩阵测量实验装置及原理

Fig. 1 Experimental device and schematic diagram for measuring backward Mueller matrix

系统光源更换为带有反射灯罩和聚光镜面的卤素灯,波长633 nm的窄带滤光片。线偏振片工作波长400~700 nm,消光比大于1 000:1,直径25.4 mm,透过率大于80%。二相混合式驱动器(SJ2H504DM,常州双杰)驱动两个电控旋转台(北光世纪公司的MAS102系列精密电控旋转台,其传动比为180:1,最大速度25°/s,刻度盘分度为1°,分辨率0.000 2°,重复定位精度小于0.005°),夹持1/4波片(武汉优光科技的胶合真零级波片,型号为WPF4225-633,工作波长为633 nm,尺寸 $\Phi=25.4$  mm)。相机采用QHYYCCD所生产的QHL-II-M系列相机。相机感光面像素大小5.2  $\mu\text{m}$ ,分辨率1 280 $\times$ 1 024,最大帧率为30帧/s,有效像素数130万,接口为1.25英寸,量子效应在633 nm波长时可以达到60%。相机垂直样品采集,避免捕获镜面反射光。受成像工作距离和旋转台尺寸限制,入射光和出射光最小角度约30°。

光波任意的偏振状态(即光波电矢量的振动状态)都可以用四个可测量的光强量组成的一个四维的数学矢量(即Stokes向量)来表达。用Stokes向量来描述入射光和出射光,两个光的Stokes向量可以用一个描述介质的 $4 \times 4$ 的矩阵联系起来。入射偏振光与介质相互作用后,出射的偏振光的特征由一个新的Stokes向量 $S'$ 描述,则入射光的 $S$ 与 $S'$ 之间可以由一个矩阵联系起来。

$$\begin{bmatrix} I' \\ Q' \\ U' \\ V' \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{31} & M_{32} & M_{33} & M_{34} \\ M_{41} & M_{42} & M_{43} & M_{44} \end{bmatrix} \begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} \quad (1)$$

穆勒矩阵包含了被测样品的全部偏振光学特征信息。获取偏振信息即是寻找特异性指标的过程,而特异性指标蕴藏于穆勒矩阵各元素之中。

穆勒矩阵包含样品的全部偏振性质。而基本的偏振性质包含二向色性、起偏、相位延迟、退偏。二向色性描述透射(或反射)光强随入射偏振态的变化,偏振态的方向性为Stokes向量首项为1下最大差异程度。因此对应穆勒矩阵二向色性的 $0^\circ$ 、 $45^\circ$ 和圆二向色性三个分量存储在第一行阵元中。起偏是指光与样品相互作用后偏振度增加的现象,出射光的Stokes向量由样品穆勒矩阵的第一列决定,因此对应穆勒矩阵起偏对应穆勒矩阵首列。相位延迟即存在两个速度最大差异的本征偏振态,电场振动产生的时间延迟。退偏表明偏振度下降的情况,退偏器的穆勒矩阵可简化为对角形式。

样品的方位角取向不确定是偏振检测中现实存在的问题<sup>[10]</sup>。相对偏振方向偏离,二向色性的值基本不变,但相位延迟和散射退偏值会增加。保证镜片垂直和运动位置准确,可最大程度减小影响。

入射光强相对分布不均匀是影响穆勒矩阵的不可忽视的问题。相对光强强度偏差越大,二向色性和相位延迟值增加,散射退偏值则减小。因此尽量缩减入射光的角度就能保证相对光强偏差减小。但机械转动调制方法的入射角度不能消除,因此要从算法上选择减小入射角度影响的分类方法。

穆勒矩阵成像装置,在原有二维平面强度信息拓展了偏振的信息维度。而在遥感领域的光谱数据,在原有二维平面强度信息拓展了波段的信息维度。两者都是对光进行了调制拓展纵向维度信息,区别只是所代表的物理含义不同。

图像分类方法中,光谱角度填图(Spectral Angle Mapping, SAM)即夹角余弦方法,是成像光谱图像处理技术中谱形识别的主要方法之一。原理是将立体数据视为多维空间的矢量,利用解析方法计算未知谱线与目标谱线之间矢量的夹角,根据夹角的大小来确定谱线间的相似程度,以达到识别分类的目的。角度求解公式为

$$\theta(x^*, x_i) = \arccos \left[ \frac{x_i^T x^*}{(\|x_i\|)(\|x^*\|)} \right] = \arccos \left[ \frac{x_i^T x^*}{\sqrt{(x^*)^T x^*} \sqrt{(x_i)^T x_i}} \right] \quad (2)$$

式中, $\theta$ 为未知向量 $x^*$ 和已知向量 $x_i$ 的夹角。

此分类法适合于谱形总体相似性分析的图像分类。可以改善光强和入射角度引起的谱线变形,其优越性显而易见。

夹角余弦分类方法依据是谱线的矢量方向而非矢量的长度。由于矢量的长度关联到影像亮度而角分类法对亮度值的影响并不敏感。正是由于此方法仅仅需要谱线矢量的方向而不关注影像本身的亮度。因此将其应用于偏振数据的分析,对偏振方向上的光强分量并不敏感,直接缓解系统成像夹角对系统穆勒矩阵的影响。

光谱角填图法,是比较端元光谱向量和每一个像素的向量在 $n$ 维空间中的角度。角度越小代表与参考光谱越匹配,当像素比特定的最大角度的弧度阈值更大的时候就不会被分类。应用于偏振数据,使用光谱角度描述并不准确,但阈值依据为所设定余弦夹角弧度,因此后文将其统一叙述为夹角余弦方法。

## 1.2 穆勒矩阵阵元分析

穆勒矩阵反映的是样品的偏振信息。面光源穆勒矩阵成像适用于均匀和非均匀散射介质,研究表明各个穆勒矩阵元也反映了样品的大量信息<sup>[11]</sup>。测量样品的穆勒矩阵并不困难,均匀散射体样品(如球状散射体聚苯乙烯小球)结构单一尚能依靠理论解释,但针对复杂结构样品(如病理组织)根据穆勒矩阵获取有用



的信息十分困难。

穆勒矩阵拥有 16 个阵元,整体的物理意义可以分为四个部分,光强信息、二向色性、起偏能力、散射退偏和相位延迟<sup>[12]</sup>,如图 2 穆勒矩阵阵元信息示意图所示, $M_{11}$  反应样品对入射光束强度的影响; $M_{12}$ 、 $M_{13}$ 、 $M_{14}$  反应样品对入射光波二向色性的影响; $M_{21}$ 、 $M_{31}$ 、 $M_{41}$  反应样品对入射光波中非偏振光的影响;其余分量反应样品对入射光波退偏振和相位延迟特性的影响。

随着研究的不断深入,穆勒矩阵的部分阵元其含义逐渐被认识<sup>[12]</sup>。二向色性信息中  $M_{14}$  的值与圆二向色性相关。 $M_{21}$ 、 $M_{31}$  阵元与非偏振光起偏性质成反比。穆勒矩阵的  $M_{24}$ 、 $M_{34}$  阵元表示各向异性纤维结构信息。对于线双折射样品, $M_{24}$ 、 $M_{34}$  阵元(以及  $M_{42}$  和  $M_{43}$  阵元)与  $\sin\delta$  成正比,而  $M_{23}$  和  $M_{32}$  以及  $M_{22}$ 、 $M_{33}$  和  $M_{44}$  阵元都与  $\cos\delta$  成正比, $\delta$  为相位延迟大小。而对角阵元( $M_{22}$ 、 $M_{33}$ 、 $M_{44}$ )都近似为 1,说明退偏较弱。一个对角的穆勒矩阵可以代表为一个纯粹的退偏器,即  $\text{diag}(1, a, b, c)$ ,其中,  $|a|, |b|, |c| \leq 1$ 。

均匀介质内部结构单一,探测物理信息一般在穆勒矩阵单独阵元中体现。例如,油雾粒子是均匀同性球形散射粒子,具有对称性,只影响穆勒矩阵的散射退偏项,因此只对低、中、高浓度下的  $M_{22}$ 、 $M_{33}$  和  $M_{44}$  一维强度分布进行分析<sup>[13]</sup>;在穆勒矩阵对光学器件表面缺陷的探测中可选择  $M_{33}$  阵元开展分析。本文验证得出不同观测角度  $M_{33}$  的正负可鉴别光学器件缺陷<sup>[14]</sup>。

生物组织微观结构十分复杂,一般为高散射的介质。生物组织中不但包含各向同性为主的散射体,例如血细胞,还包含了如骨骼肌、胶原纤维、神经元、心肌、癌组织等有序的各向异性纤维状结构,更复杂的情况,如细胞中的微丝微管等柱状结构本身具有双折射,而组织散射体间隙也存在着糖类手性分子等旋光和双折射的效应。偏振光通过某些介质后透射出来的光仍然是偏振光,但是偏振光的振动平面向左或向右旋转了一定的角度,这叫做旋光现象(圆双折射)。

在生物学样品偏振检测中,所测得的穆勒矩阵非常复杂( $M_{14}$ 、 $M_{41}$  不为零,矩阵转置不对称等等)。目前,依靠理想化实验仿体对内部结构进行深入研究。仿照真实生物组织样本制备替代物质混合的仿体作为模型,探测微观结构信息<sup>[15]</sup>。

真实生物样品的穆勒矩阵信息提取工作,关注点只停留在对个别有较大差异的阵元,选择性忽略其它阵元的差异信息。例如,未染色的脱蜡的癌变切片的偏振性质主要集中在  $M_{24}$  和  $M_{34}$  阵元中,直接丢弃了其它 14 个阵元信息。但由于实际情况更为复杂,不同癌变组织内部结构不尽相同,相对明显的信息可能在其它阵元当中体现,唯独两个阵元并不足以得出结论。因此临床判别依据信息应该是越多越好,但也同样对差异信息提取提出了很高的要求。本文研究穆勒矩阵全阵元曲线希望通过提取整体特征的方式全面呈现穆勒矩阵内部差异信息。

### 1.3 样本选择

双旋转波片穆勒矩阵检测系统依靠控制电控旋转平台旋转波片,避光采集供穆勒矩阵求解的 30 张样品图像数据。实验室前期验证过偏振对未染色病理切片检测的可行性<sup>[9]</sup>。本文研究主要使用如图 3 所示肺癌和乳腺癌病理切片。经由医院病理科室标注出对应癌变区域。未染色切片样本为吉林大学第一医院提供,这项工作得到了吉林大学第一医院伦理委员会的批准。

肺癌组织与正常组织之间结构差异明显,成像效果较好,是辅助分析方案验证的最佳选择。选用有明显背景干扰并无法从依靠单一信息提取差异的成像数据来提出整体特征分析的解决方案。乳腺组织较肺组织内部功能结构更复杂,浸润类型尤其复杂,因此选用

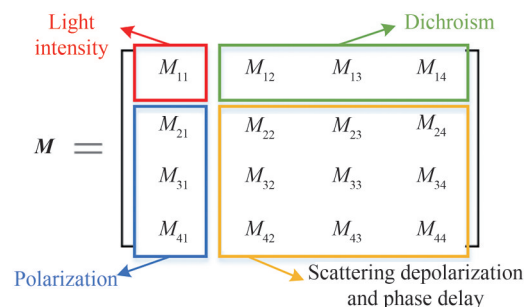


图 2 穆勒矩阵阵元信息示意图

Fig. 2 Schematic diagram of Mueller matrix element information

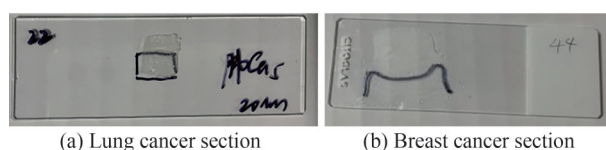


图 3 未染色病理切片样品

Fig. 3 Unstained pathological section sample

未染色浸润型乳腺癌病理切片作为样品,开展特征提取以及跨尺度验证实验。

穆勒矩阵每个阵元信息容易获得,但各阵元信息难以解读,内部关系十分复杂,阵元之间更没有具体表达公式。穆勒矩阵同偏振性质内部结构特征联系最为紧密。因此依据目前已探寻清楚阵元的分类状况,链接所有阵元用于表征生物样本的整体结构特征,具体链接情况如图4所示。

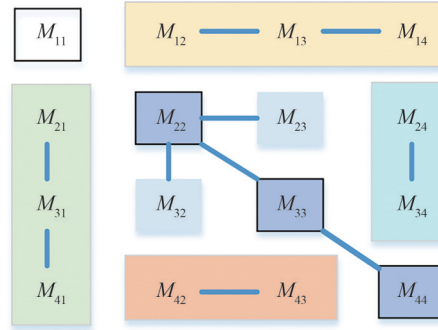


图4 穆勒矩阵阵元信息链接图

Fig. 4 Mueller matrix element information link diagram

归一化穆勒矩阵依照强度信息生成,因此代表强度的阵元 $M_{11}$ 放首位,其余依次按照分类的二向色性、起偏能力、相位延迟和退偏特性展开。在退偏和相位延迟特性中,被探明的相位延迟特性分类中 $M_{24}$ 、 $M_{34}$ 、 $M_{42}$ 、 $M_{43}$ 与 $\sin\delta$ 成正比,而 $M_{23}$ 和 $M_{32}$ 以及 $M_{22}$ 、 $M_{33}$ 和 $M_{44}$ 阵元都与 $\cos\delta$ 成正比。对角阵元 $M_{22}$ 、 $M_{33}$ 、 $M_{44}$ 表明退偏特性。最终所链接阵元顺序依次为 $M_{11}$ 、 $M_{12}$ 、 $M_{13}$ 、 $M_{14}$ 、 $M_{21}$ 、 $M_{31}$ 、 $M_{41}$ 、 $M_{24}$ 、 $M_{34}$ 、 $M_{42}$ 、 $M_{43}$ 、 $M_{23}$ 、 $M_{32}$ 、 $M_{22}$ 、 $M_{33}$ 、 $M_{44}$ 。

将内部复杂无法解释的关系简化为链接的曲线信息,以可视化曲线的形式直观的展现出来。样品的结构差异信息贮存在穆勒矩阵中,经由链接处理,样品的结构信息转移于可视化的曲线之中。因此问题就由复杂的平面空间特征提取,转换为曲线特征提取。曲线表现形式多样,特征提取方法众多的优势也同样被继承。

## 2 实验结果

### 2.1 数据制备及分析

采集肺癌切片强度信息生成穆勒矩阵,再进行归一化处理。归一化展示的是趋势信息。归一化方便后续数据处理,其次保证运行加快收敛。如图5所示为肺癌切片22号归一化穆勒矩阵。

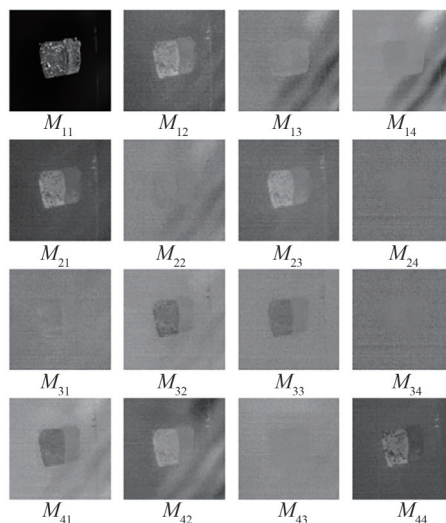


图5 肺癌切片22号归一化穆勒矩阵

Fig. 5 Lung cancer slice No. 22 normalized Mueller matrix

$M_{11}$ 为光强信息,切片组织成像清晰。前文理论分析提及未染色的脱蜡的癌变切片的偏振性质主要集中在 $M_{24}$ 和 $M_{34}$ 阵元中<sup>[16]</sup>。图5所示的 $M_{24}$ 、 $M_{34}$ 阵元背景与组织信息混叠难以区分,因此由于生物组织内部结构复杂性,原始选取目标阵元观测理论并不适用。其它阵元出现的背景条带干扰来自遮光布褶皱条纹。穆勒矩阵包含庞大的组织结构信息,受多种因素影响,偏振特征信息始终存在,但表现的强度可能有所衰减。全阵元差异信息相比于个别阵元无疑更加可靠,全阵元更是包含个别阵元,因此有更强的稳定性和兼容性。

保留原有二维强度信息,生成全新的纵向包含偏振物理意义的穆勒矩阵立方体数据集,如图6所示。将平面二维扩充至三维,增添了纵向的维度。扩充维度的方式同遥感领域光谱数据相同,区别在于表示物理含义不同。在遥感领域拥有一套成熟的数据生成和处理方案,因此借助光谱数据和曲线的生成方法,使用遥感领域ENVI5.3软件,开展偏振数据立方体合成以及验证工作。

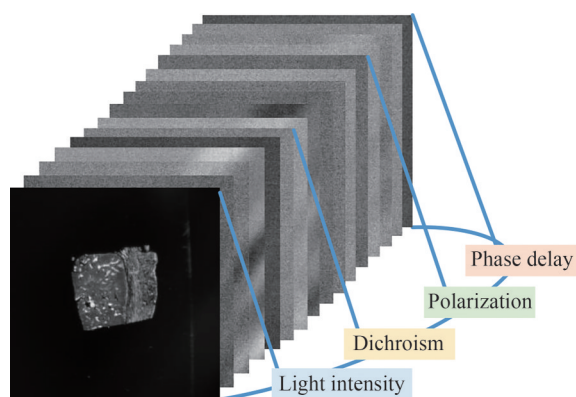


图6 穆勒矩阵数据立方体  
Fig. 6 Mueller matrix data cube

ENVI软件自带感兴趣区(Region Of Interest, ROI)绘制工具,数据集任意像素可以提取出对应的特征曲线。依据切片标注区域选取癌变组织、正常组织、背景三个感兴趣区绘制特征曲线,如图7所示。红色方形A代表癌变组织,绿色圆形B代表正常组织,蓝色三角形C代表背景,7(a)绘制感兴趣区域与7(b)生成曲线,颜色符号相互对应。7(b)纵坐标表示所在阵元偏振的强度信息,横坐标正是立方体数据顺序依次为 $M_{11}$ 、 $M_{12}$ 、 $M_{13}$ 、 $M_{14}$ 、 $M_{21}$ 、 $M_{31}$ 、 $M_{41}$ 、 $M_{24}$ 、 $M_{34}$ 、 $M_{42}$ 、 $M_{43}$ 、 $M_{23}$ 、 $M_{32}$ 、 $M_{22}$ 、 $M_{33}$ 、 $M_{44}$ 。全文分类标注及曲线坐标含义相同且对应,后文不做赘述。不同目标所生成的特征曲线的差异,正是内部结构上的差异,曲线的每个点位保留其物理含义。特征曲线表征的是目标的整体特征,强度大小则是程度信息。

观测图7(b)特征曲线,单从1点位二维光强信息无法区分正常组织(91.6)以及癌变组织(93.8)。特征曲线特异性指标不明显,点位邻近(同为邻近点3、4、6、8、9、12、15、16)。其中前文提到的纤维特性(点8、9),多种因素和干扰导致折线拐点不能提取特征信息。存在区分特性明显的点位(2、5、7、10、11、13、14),同样

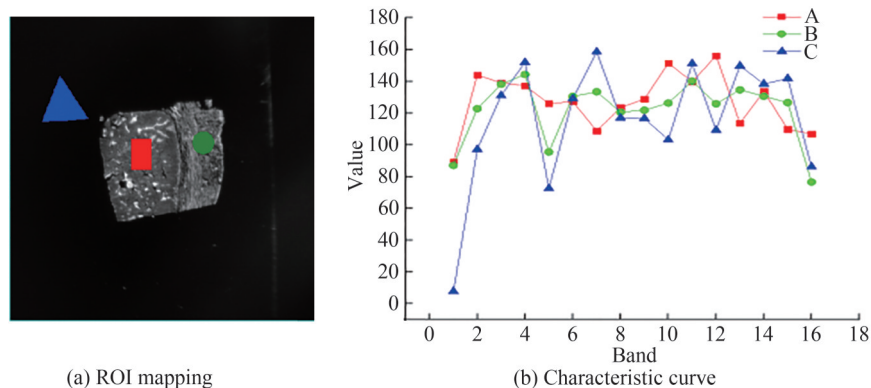


图7 肺癌22号感兴趣区选取及特征曲线图  
Fig. 7 Selection and characteristic curve of 22 ROI of lung cancer



具有对应的物理含义,2点是二向色性,5点位代表偏振特性信息。10、11与 $\sin\delta$ 的相位延迟有关。13、14和退偏有关。特征曲线不仅满足穆勒矩阵常用的二维平面的解读方式,同时支持多个点位联合分析。例如4组联合点(8、9、10、11)与相位延迟成正比,观测正常组织和癌变组织特征曲线保持在癌变曲线之上,得出正常组织的相位延迟特性更强。

穆勒矩阵全阵元数据分析区别于二维空间图像的描述形式,曲线的表现形式灵活多变。归一化穆勒矩阵是偏振程度化信息。雷达图对单独或者归类的趋势信息展示十分直观,尤其适合所生成的偏振数据。如图8所示肺癌22号特征曲线雷达图。

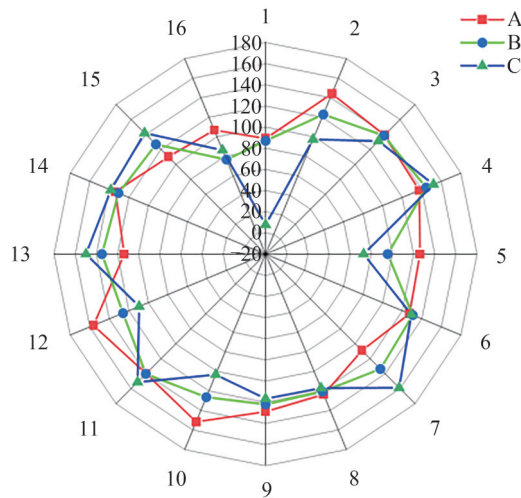


图8 肺癌22号特征曲线雷达图

Fig. 8 Radar figure of lung cancer characteristic curve 22

距圆心距离反映对应信息的强弱程度。单独阵元内部结构差异在雷达图单线上很容易观测。例如:前文提到未染色切片判别依据 $M_{24}$ 、 $M_{34}$ 阵元(对应图8中8、9)背景与组织信息混叠,雷达图8、9线上,信息点在雷达图上几乎重合。雷达图更容易展示同一类别所在偏振特性的趋势。例如:8、9整体代表纤维特征,观测单独阵元尽管信息混叠,但联合观测8和9的连线,可以看出癌变组织更趋向雷达外,正常组织趋向中心,而背景则是夹在中间。正是解释癌变组织相比于正常组织确实存在纤维特性差异,但由于生物样本复杂性导致此特性并不明显。当然观测同一类型的整体所包络雷达图形,正是整体偏振信息的体现。

穆勒矩阵全阵元曲线通过雷达图的形式,直观展示了对单独阵元、同偏振类型联合阵元、整体偏振信息的解析能力。

## 2.2 穆勒矩阵全阵元特征提取

若要采用特征曲线进行特征提取,必须要验证所制备全阵元曲线样本同类型曲线所包含信息的相似性。以及不同类型曲线特征的差异性。对肺癌切片病理标注区域分别开展随机采样10组平均特征曲线,同坐标系展示同类组织曲线相类似性,不同组织特征曲线差异性。如图9所示肺癌切片癌变和正常随机样本特征曲线内部特征对比情况。以像素为单位开展特征提取,针对切片标注区域,对癌变和正常组织开展随机采集各10个随机区域如图9(a)所示。并分别绘制曲线图,如图9(b)癌变组织采样特征曲线,内部曲线几乎重叠,具有空间相似性。使用如图9(d)和图9(e)瀑布图,更加清晰的展示特征曲线的相似性。同理对比图9(c)内部代表的正常组织的随机10个全阵元曲线,也有同样的结论。10组随机样本曲线趋势几乎重叠,但癌变组织和正常组织之间曲线又存在明显的差异。

证明相同组织结构的特征曲线相似,不同组织结构特征曲线存在差异,因此可以匹配全部样本差异信息获取分类结果,肺癌切片特征曲线采样分类过程如图10所示。

光谱角填图法分类是依据设置阈值和标准谱线库,对全数据匹配遍历得到分类结果。标准谱线库可以从本数据获取,也可以通过其它数据获取。选择依据自身数据将生成的平均曲线如图10(b)所示作为标准谱线库。光谱分类中涉及到光谱角填图法,遥感领域经验阈值通常设置是0.15,阈值对应角度的弧度值,阈

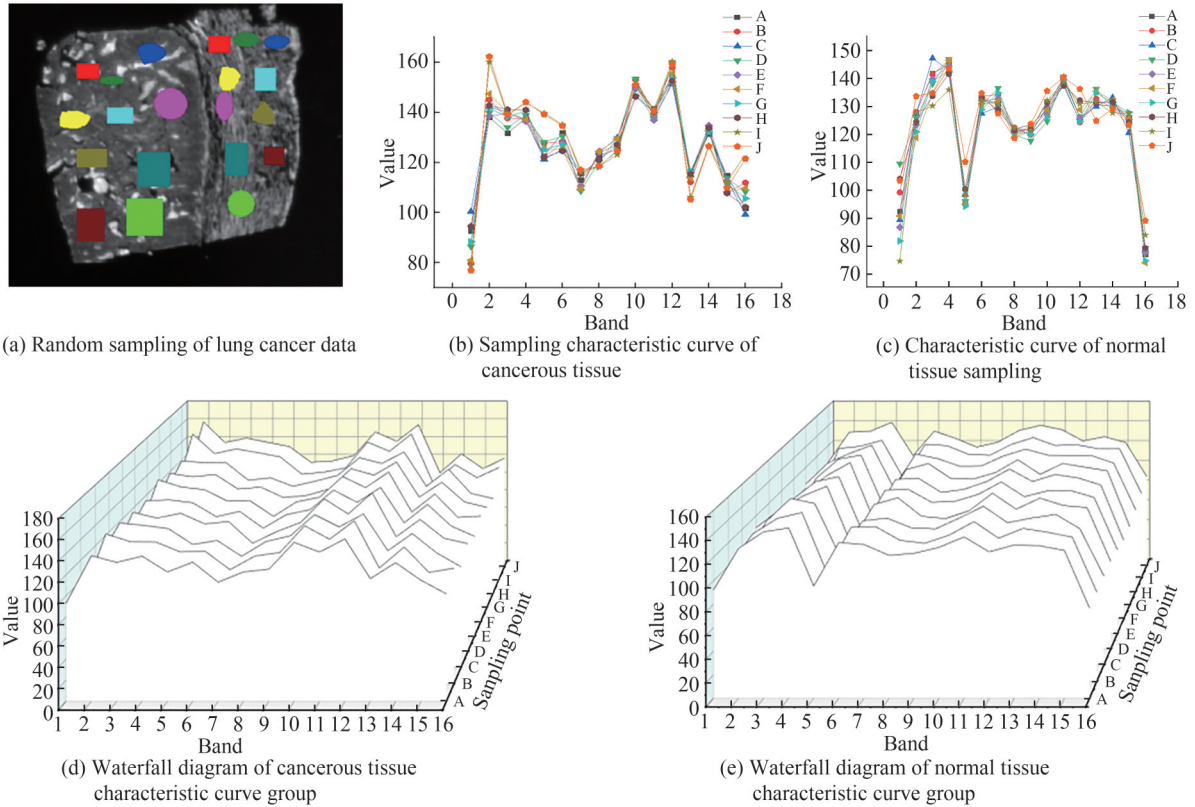


图9 肺癌切片癌变和正常随机样本特征曲线内部特征比对

Fig. 9 Comparison of internal characteristics of characteristic curves of cancerous and normal random samples from lung cancer sections

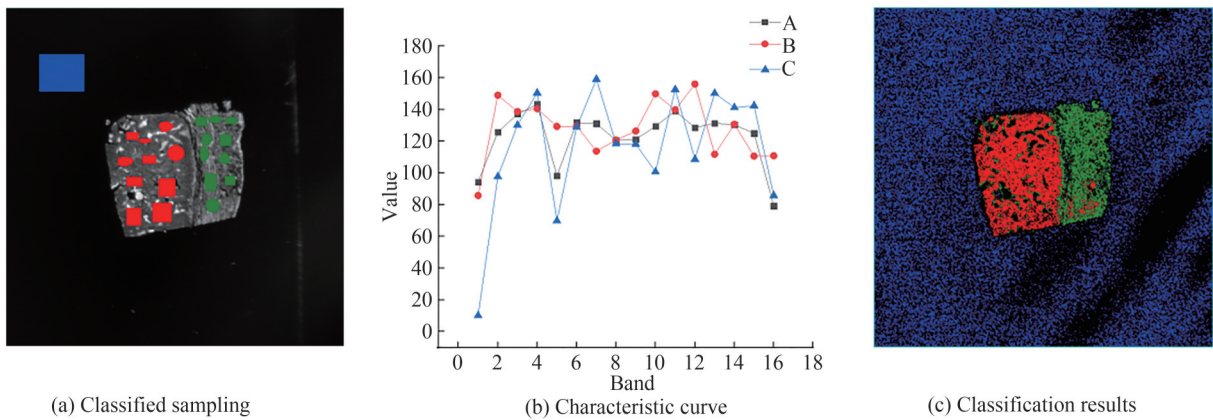


图10 肺癌切片特征曲线采样分类过程

Fig. 10 Classification process of lung cancer slice characteristic curve sampling

值越小分类的要求越苛刻,相比默认0.1的阈值适当放宽分类条件。并未查阅到生物组织相关此方法最适阈值的研究,因此本文也采用最大弧度角0.15的匹配精度,采用夹角余弦方法遍历整个穆勒矩阵数据集。最终分类结果如图10(c)所示。

特征曲线整体的差异信息作为标准,跳过解释过程匹配得到分类结果。从图10(c)分类结果看穆勒矩阵曲线完全能区分癌变、非癌变组织以及背景区域,边界清晰细节明显。值得注意的是受外界干扰的斜条纹黑色没有被归类,并通过穆勒矩阵分析显现出来,也能反应特征曲线探测的精准性。

评价图像分类精度依靠混淆矩阵展示。图10(c)中黑色区域为超出阈值无法匹配未定义曲线,根据实际应用场景,未判别类型应归类无用信息属于背景。因此将未匹配区域全部转为背景,得到如图11(a)所示分类结果被测图。依据医院病理科室医生标注癌变区域,手绘标准模板如图11(b)所示,还原真实临床标注



组织分布情况。

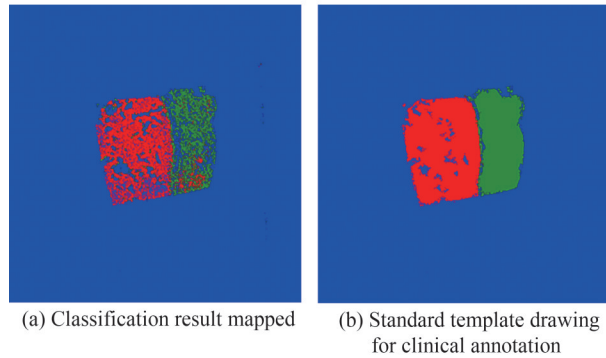


图 11 肺癌切片精度判别被测图和模板图

Fig. 11 Accuracy discrimination of lung cancer slice measured map and template map

输入图 11 被测图以及标准图生成混淆矩阵如表 1 所示。

表 1 肺癌切片分类结果混淆矩阵

Table 1 Confusion matrix of lung cancer slice classification results

Actual/forecast	Real cancerous	Real normal	Real background	Total
Cancerous tissue	46 070	3 539	9	49 618
Normal tissue	5 129	29 803	139	35 071
Background	14 979	10 301	811 631	836 911
Total	66 178	43 643	811 779	921 600

通过表 1 混淆矩阵可知,全图 921 600 像素中,背景所占像素数最多,说明背景准确性对总体准确性影响最大。纵向观测错分情况中,组织错分为背景相比于组织内部错分更严重,说明背景对组织分类准确性影响最大。标准背景错分为组织较少,说明背景同组织差异明显。组织内部正常和癌变存在互相错分,但比例不大,保证了较高的识别准确率,这是因为肺癌组织结构同正常组织结构有明显差异。这说明穆勒矩阵全阵元曲线探测目标结构特征的可行性。观测准确率生成如表 2 肺癌切片整体分类精度统计。

表 2 肺癌切片整体分类精度统计

Table 2 Overall classification accuracy statistics of lung cancer sections

Category	(Classification pixel) accuracy/%
Cancerous tissue	(46 070/49 618) 92.85%
Normal tissue	(29 803/35 071) 84.98%
Background	(811 631/811 779) 96.98%
Overall accuracy	(887 504/921 600) 96.303%
Kappa	0.809 7

表 2 中,总体分类精度可以达到 96.3%,Kappa 值能够达到 0.809 7。通过混淆矩阵可知癌变组织分类精度 92.85%,正常组织分类精度 84.98%。背景与组织曲线区别大,因此精度也最高可达 96.98%。在辅助诊断中最关心正是正常组织和癌变组织分类状况,由于背景所占像素数最多,影响观测组织真实精度状况。因此去除背景,生成正常组织和癌变组织的精度统计如表 3 所示。肺癌切片组织分类精度为 89.59%。

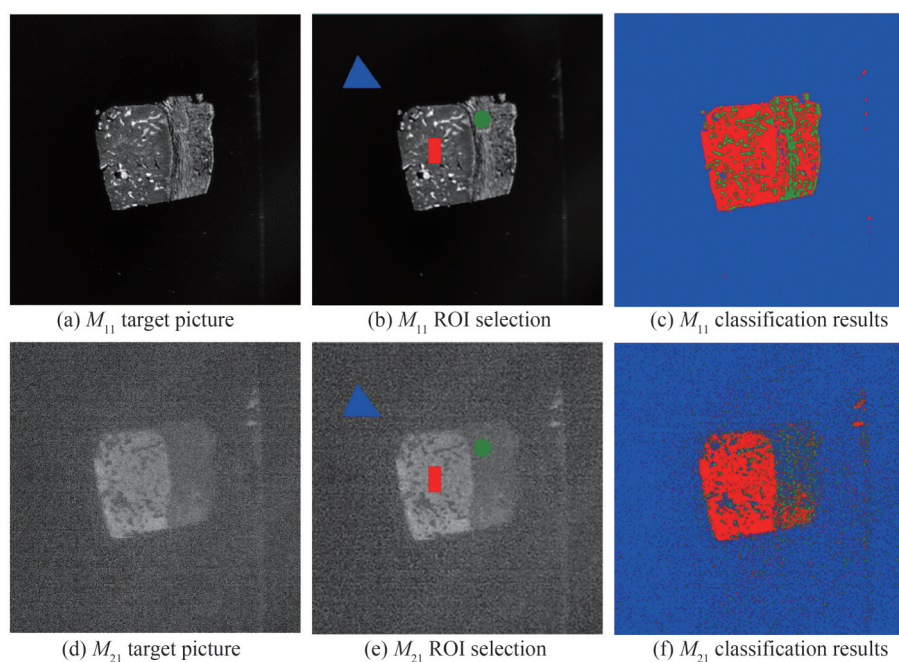
上述内容展示了穆勒矩阵整体分类结果及精度,为了更全面的展示整体分类相较于个别阵元分类的优势。设置单一阵元的分类对照实验。选取阵元中成像清晰的  $M_{11}$  和  $M_{21}$  阵元,借助 ENVI 自带平面监督分类支持向量机(Support Vector Machines, SVM)方法使用默认参数开展实验。为了增强可比性,两阵元使用同

表3 肺癌切片正常组织和癌变组织的精度统计

Table 3 Accuracy statistics of normal and cancerous tissues in lung cancer sections

Category	(Classification pixel)accuracy/%
Cancerous tissue	(46 070/49 618)92.85%
Normal tissue	(29 803/35 071)84.98%
Overall accuracy	(75 873/84 689)89.590 1%

一感兴趣区域,以及使用上文分类同一标准模板图11(b)。整个过程如图12所示。

图12 肺癌切片 $M_{11}$ 和 $M_{21}$ 单阵元SVM分类过程Fig. 12 Single element SVM classification process for  $M_{11}$  and  $M_{21}$  lung cancer slices

最终,生成混淆矩阵,其中 $M_{11}$ 癌变组织分类精度61.9%(53 843/86 982),正常组织分类精度53.85%(18 106/33 621)。 $M_{21}$ 癌变组织分类精度59.34%(54 634/92 074),正常组织分类精度14.30%(17 847/124 771)。结果展示以及精度状况与整体分类结果差异明显。

### 2.3 跨尺度特性

相比于肺癌组织,浸润型乳腺组织结构功能复杂,区分难度较大。浸润型乳腺癌为癌细胞侵袭入正常组织内部,切片上记号笔所圈区域为癌变区域。实验设备更换高倍率镜头,选取切片交界红框区域。生成数据立方体选取特征曲线并分类得出结果的完整过程,如图13所示。

经由全阵元曲线开展分类实验,总体分类精度可以达到77.5%。说明高倍率成像穆勒矩阵数据集,可以实现特征提取,依靠整体特征信息完成分类。但高倍率成像视野狭小,完整采集单张切片需要多次平移采样,导致特征提取效率极低。混淆矩阵分析可知,癌变组织准确率只有71%,大部分癌变组织错分于正常组织内部。由于正常组织较薄,选取ROI时包含了较多的背景信息,造成癌变组织和背景混叠导致分类错误,癌变以及正常组织的混叠也受此影响。

利用偏振成像的跨尺度特性,降低倍率扩大成像视野使得单次采样包含整个组织,达到提高效率节省时间的目的。乳腺癌44号低倍率成像穆勒矩阵特征曲线分类如图14所示。

低倍率成像尽管损失内部细节,但是切片的目标区域同样被区分出来。切片上的喷塑字体以及切片夹持器没有误分类。可见低倍率下穆勒矩阵成像,依旧能探测出内部结构特征。同图13成像情况一样,存在癌变组织较薄,导致选取ROI包含背景信息,背景被错分为癌变组织。总体的准确率在87.82%,相比图14精度更高,猜测是由于相同像素下大视野导致背景信息弱化。无论高低倍率都能对内部结构特征进行探测提取,也证明了偏振成像的跨尺度特性,对解决病理检测当中因显微镜观测视野小效率低的问题具有十分重要的意义。

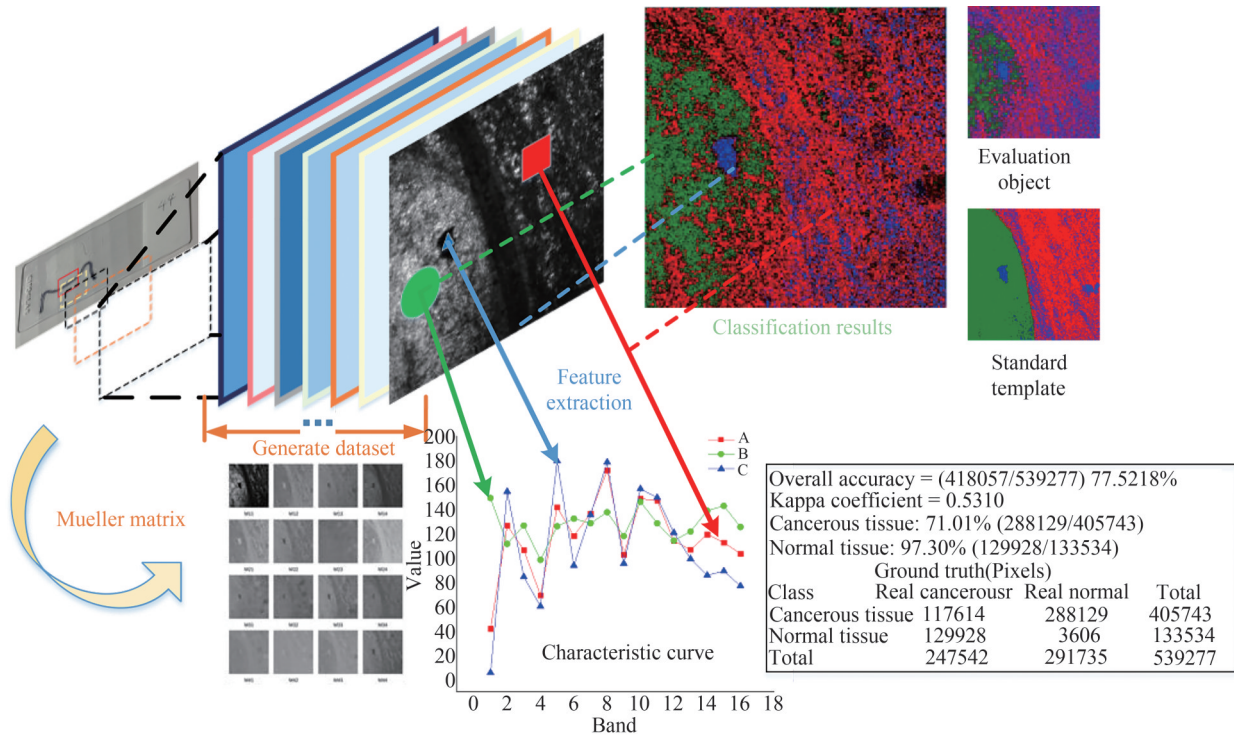


图 13 乳腺癌 44 号高倍率成像穆勒矩阵特征曲线分类

Fig. 13 Mueller matrix characteristic curve classification of breast cancer No.44 high magnification imaging

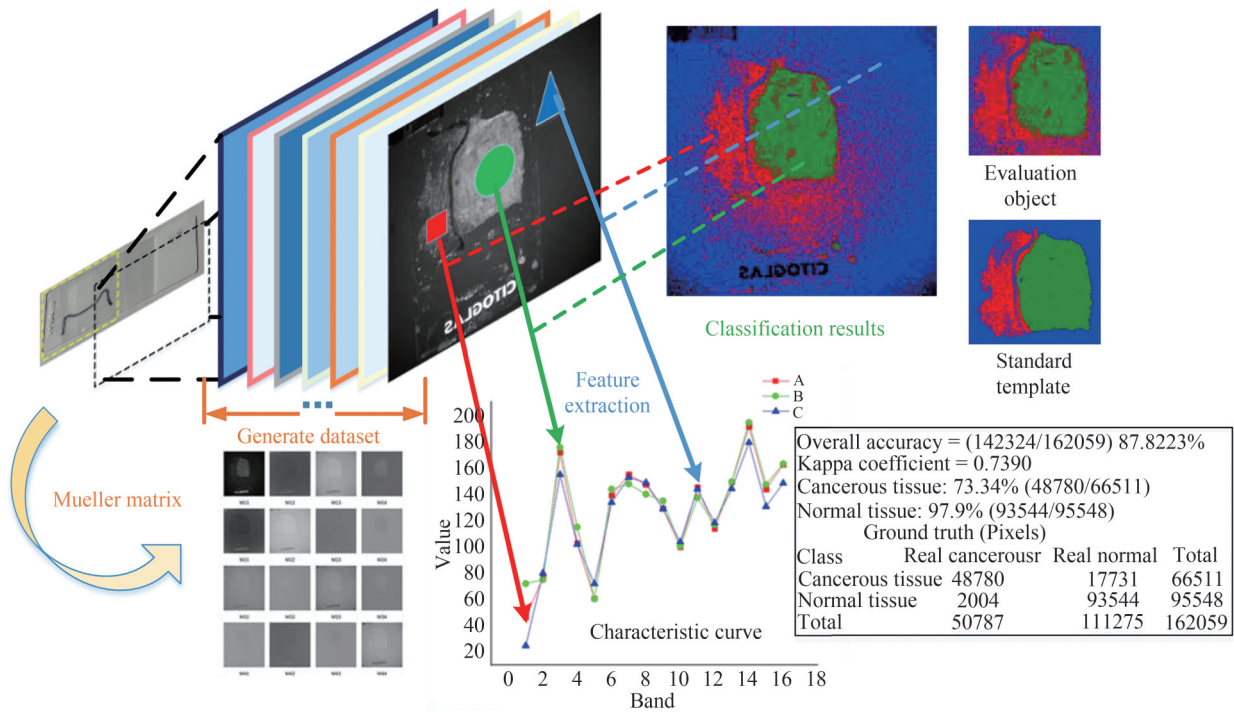


图 14 乳腺癌 44 号低倍率成像穆勒矩阵特征曲线分类

Fig. 14 Mueller matrix characteristic curve classification for low magnification imaging of breast cancer No.44

### 3 结论

本文从病理检测中所存在并且亟需解决的问题出发,改进原有偏振成像装置提高成像质量。分析偏振成像探测原理,依据穆勒矩阵阵元物理信息,扩充原有的分析维度生成数据立方体。赋予立方体纵向任意像素的特征曲线具有明确物理含义。选用未染色的肺癌和乳腺癌病理切片样本,提取样品纵向穆勒矩阵全



阵元曲线,并借助雷达图展示曲线多变的分析方式。随机采样10组平均特征曲线,同坐标展示同类组织曲线相类似性,不同组织特征曲线差异性。最终,借鉴夹角余弦分类方法用于抵消光路夹角问题带来的系统误差,依据混淆矩阵,肺癌分类获得89.59%的高精度,乳腺癌高低倍率成像分类获得77.5%和87.8%的精度。三者的错分主要来源于背景。分析可知光学平台遮光布结构特性,对组织背景干扰尤为明显。其次分类使用自身ROI作为谱线库,由于组织较薄不可避免夹带背景信息,加剧了组织和背景信息混叠,更干扰组织内分类状况。因此,针对背景问题做出改进,实现与组织信息区分开来正是下一步工作。

着眼于临床辅助人工检测,本文侧重于在结构层面提出正常和癌变组织区分的思路。由于样本实验篇幅有限,但本文提供了从采集构造分析数据以及最终完成分类判别精度的整套思路,并且结果显示对癌变、正常组织及背景区分度好精度较高,低倍率成像依旧能完成特征提取。跳过染色过程,对临床缩短病理检测时间有十分重要的意义。本文方法展现了对不同癌变的适用性,也为后续继续开展大样本谱线库提取研究指明了方向。更进一步表明全阵元曲线在偏振检测领域有很大的挖掘空间以及应用潜力。

### 参考文献

- [1] The latest WHO cancer report: breast cancer becomes the “largest cancer in the world”[J]. *China Medical News*, 2021, 36(2): 18.  
WHO最新癌症报告:乳腺癌成“全球第一大癌”[J]. *中华医学信息导报*, 2021, 36(2): 18.
- [2] YUAN Huiyun, JIANG Yufei, TAN Yuting, et al. Current status and time trends of cancer incidence and mortality worldwide[J]. *Cancer Research on Prevention and Treatment*, 2021, 48(6): 642-646.  
袁蕙芸, 蒋宇飞, 谭玉婷, 等. 全球癌症发病与死亡流行现状和变化趋势[J]. *肿瘤防治研究*, 2021, 48(6): 642-646.
- [3] HE C, HE H, CHANG J T, et al. Polarisation optics for biomedical and clinical applications: a review[J]. *Light: Science & Applications*, 2021, 10(1): 194.
- [4] YAO Y, ZHANG F D, WANG B, et al. Polarization imaging-based radiomics approach for the staging of liver fibrosis[J]. *Biomedical Optics Express*, 2022, 13(3): 1564-1580.
- [5] HE H H, LIAO R, ZENG N, et al. Mueller matrix polarimetry—an emerging new tool for characterizing the microstructural feature of complex biological specimen[J]. *Journal of Lightwave Technology*, 2019, 37(11): 2534-2548.
- [6] AZZAM R M. Photopolarimetric measurement of the Mueller matrix by Fourier analysis of a single detected signal[J]. *Optics Letters*, 1978, 2(6): 148.
- [7] ZAKERI A, BAYGI MHM, MADANIPOUR K. Polarization characterization of biological tissues using Stokes vector decomposition[J]. *Journal of Modern Optics*, 2013, 60(12): 987-992.
- [8] SI L, HUANG T Y, WANG X J, et al. Deep learning Mueller matrix feature retrieval from a snapshot Stokes image[J]. *Optics Express*, 2022, 30(6): 8676-8689.
- [9] TIAN Bowen, HUANG Danfei, ZHONG Aiqi, et al. Detection of pathological sections of breast cancer based on polarization imaging[J]. *Science Technology and Engineering*, 2020, 20(16): 6369-6374.  
田博文, 黄丹飞, 钟艾琦, 等. 偏振成像的乳腺癌病理切片检测方法[J]. *科学技术与工程*, 2020, 20(16): 6369-6374.
- [10] PIERRANGELO A, NAZAC A, BENALI A, et al. Polarimetric imaging of uterine cervix: a case study[J]. *Optics Express*, 2013, 21(12): 14120-14130.
- [11] HE H H, CHANG J T, HE C, et al. Transformation of full  $4 \times 4$  Mueller matrices: a quantitative technique for biomedical diagnosis[C]. *SPIE*, 2016, 9707: 9707OK.
- [12] WANG Yan. Research on Mueller matrix based on polarization detection[D]. Dalian: Liaoning Normal University, 2019.  
王妍. 基于偏振探测的穆勒矩阵研究[D]. 大连: 辽宁师范大学, 2019.
- [13] SONG Junhong, XIAO Zuojiang, LI Yingchao, et al. Influence of concentration variation of oil mist particles on scatter[J]. *Acta Optica Sinica*, 2021, 41(23): 9-18.  
宋俊宏, 肖作江, 李英超, 等. 油雾粒子浓度变化对散射穆勒矩阵的影响[J]. *光学学报*, 2021, 41(23): 9-18.
- [14] LU Min, WANG Zhile, ZHANG Shuqing, et al. Polarization scattering characterization and discrimination principle of surface defects[J]. *Acta Optica Sinica*, 2021, 41(12): 252-258.  
陆敏, 王治乐, 张树青, 等. 表面缺陷的偏振散射表征方法及区分原理[J]. *光学学报*, 2021, 41(12): 252-258.
- [15] GUO Yihong. Quantitatively characterizing the anisotropic microstructure using the Mueller matrix decomposition technique[D]. Beijing: Tsinghua University, 2014.  
郭亦鸿. 利用穆勒矩阵分解定量测量各向异性介质微观结构[D]. 北京: 清华大学, 2014.
- [16] CHANG Jintao. A study on design of polarimetric measurement system for biomedical applications[D]. Beijing: Tsinghua University, 2016.  
常金涛. 生物体系偏振测量系统的设计[D]. 北京: 清华大学, 2016.

## Cancer Tissue Recognition by Muller Matrix Imaging Full Array Curve

GAO Junzhao<sup>1,2</sup>, HUANG Dangfei<sup>1,2</sup>, ZHANG Lechao<sup>1,2</sup>, SONG Dong<sup>3</sup>, HONG Jinghui<sup>3</sup>,  
ZHANG Lili<sup>1,2</sup>, TANG Hongyu<sup>1,2</sup>, ZHOU Yao<sup>1,2</sup>

(1 College of Optoelectronic Engineering, Changchun University of Technology, Changchun 130022, China)

(2 Zhongshan Research Institute of Changchun University of Technology, Zhongshan 528437, China)

(3 The First Hospital of Jilin University, Changchun 130021, China)

**Abstract:** As the second most common cause of death in the world, cancer has become a persistent public health challenge facing the world. At present, the most mature treatment for cancer is surgical resection, and the determination of whether the resection is clean depending on the detection of artificial pathological sections. During the operation, it is necessary to repeatedly confirm whether the diseased tissue is removed cleanly, and each test requires the preparation of pathological sections for manual detection. Pathological section staining is tedious and time-consuming, and the process of manual section detection and confirmation is time-consuming and inefficient. Therefore, there is an urgent need for an auxiliary detection method that skips the staining process, and the method is accurate, fast, highly sensitive, non-contact, non-invasive, and highly safe.

Polarization is a new information dimension independent of light intensity, wavelength, and phase. Polarization imaging uses the incident light to irradiate the sample to detect the outgoing light and obtain the physical information of the sample. Polarization imaging is a non-invasive, label free, non-contact optical detection technology. The detected polarization information is often represented by Mueller matrix. Mueller matrix characterizes the polarization characteristics of complex media, including rich macro and micro structural information. Mueller matrix imaging can encode rich information on microstructure even under low resolution and wide field conditions.

Polarization detection technology has the advantages such as large information, strong compatibility, wide scale, but the microstructure of each element value of Mueller matrix is not directly related to the same product, and it is difficult to characterize due to the mixing of a large number of influencing factors. It is difficult to obtain specific indicators of sample microstructure. Today, the focus on information extraction of Mueller matrix is only on the processing level of individual array elements with obviously different information, which is very dependent on the information content of sample array elements. This research mainly uses the existing clear classification of physical relations to establish the relationship between the internal information of Mueller matrix, analyze the full array element of Mueller matrix vertically with pixel as the unit, and further extract the overall difference information.

In this paper, we will start with the reconstruction of the original backward Mueller matrix imaging system light source and the incident direction to improve the imaging quality. We use the measurement method of double wave plates rotating 30 times to generate Mueller matrix, which is different from the previous data analysis and processing methods. With the help of the classified physical information of Mueller matrix, the new arrangement order is linked, and the Mueller matrix data cube is generated by analogy with the spectral data processing method in the remote sensing field. The full array element curve is extracted longitudinally in pixels. The whole array element curve calls the overall polarization characteristics, which is different from the description form of two-dimensional space image. The expression form is flexible and changeable, which can more intuitively describe the multi-level analysis ability of information. The included angle cosine method is used to eliminate the system error caused by the included angle of the light path, and the target classification is realized by traversing the entire two-dimensional space. This method is fast and easy to implement. On this basis, the samples were selected from the pathological sections of unstained lung cancer and breast cancer. According to the above methods, information visualization analysis, feature extraction and classification, cross scale experiments were carried out, and a confusion matrix was used to show the classification and accuracy. The results show that the pixel level full array information observation mode has more flexible target selection, more variable observation forms and strong anti-interference ability, and good accuracy in distinguishing cancerous change, non cancerous change and background. The accuracy of lung cancer and breast cancer can reach 89.59% and 87.82%, and the accuracy of high magnification breast cancer is 77.52%. Low magnification

imaging with cross scale characteristics can still complete feature extraction. Skipping the staining process and assisting in manual detection is of great significance for shortening the time of pathological detection in clinical practice.

**Key words:** Polarization imaging; Mueller matrix; Cancerous tissue; Feature extraction; Pathological section detection

**OCIS Codes:** 260.2130; 110.5405; 170.4580; 170.0110