



(12)发明专利

(10)授权公告号 CN 103902997 B

(45)授权公告日 2017.02.22

(21)申请号 201210572602.5

(22)申请日 2012.12.26

(65)同一申请的已公布的文献号

申请公布号 CN 103902997 A

(43)申请公布日 2014.07.02

(73)专利权人 西交利物浦大学

地址 215123 江苏省苏州市工业园区独墅湖高等教育区仁爱路111号

(72)发明人 张百灵 张云港

(74)专利代理机构 苏州创元专利商标事务所有
限公司 32103

代理人 范晴

(51)Int.Cl.

G06K 9/46(2006.01)

G06K 9/66(2006.01)

(56)对比文件

CN 102359938 A,2012.02.22,

CN 101388074 A,2009.03.18,

CN 101620669 A,2010.01.06,

CN 102289665 A,2011.12.21,

Sheng Yang等.A Fast Pedestrian

Detection Method Based on Simplified HOG Descriptor.《International Journal of Digital Content Technology and its Applications》.2012,第6卷(第4期),全文.

韦振中.基于核主成分分析的特征提取方法.《广西工学院学报》.2006,第17卷(第4期),第27—31页.

审查员 郭放

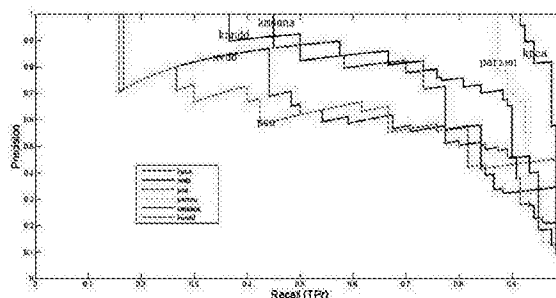
权利要求书2页 说明书7页 附图2页

(54)发明名称

用于生物细胞显微镜图像分类的特征子空间集成方法

(57)摘要

本发明公开了一种用于生物细胞显微镜图像分类的特征子空间集成方法,包括提取待分类的生物细胞显微镜图像的特征;使用核主成分分析法(KPCA)对提取的生物细胞显微镜图像的三种图像特征构建特征子空间模型,使每类生物细胞显微镜图像具有三个特征子空间;采用每一类已训练好的三个特征子空间对待分类的生物细胞显微镜图像重建三种图像特征,得到由每一类中各特征子空间对待分类图像特征的重建结果,通过各个特征子空间重建结果与原提取的图像特征向量的比较获得分类图像对每一类的分类置信度;将待分类图像将分类到置信度最高的类别当中。该方法可以有效地降低特征维数,提高集成分类器的多样性,进一步提升分类效果。



1. 一种用于生物细胞显微镜图像分类的特征子空间集成方法,其特征在於所述方法包括以下步骤:

(1) 通过以下方法提取待分类的生物细胞显微镜图像的特征:

i) 将生物细胞显微镜图像变换到不同的频率子带Sub-band中,然后对各个频率子带进行特征统计;以及

ii) 采用灰度共生矩阵的多个统计特征得到生物细胞显微镜图像的全局纹理特征;以及

iii) 通过完整局部二值模式提取生物细胞显微镜图像的局部纹理特征;

(2) 使用核主成分分析法KPCA对提取的生物细胞显微镜图像的三种图像特征构建特征子空间模型,使每类生物细胞显微镜图像具有三个特征子空间;

(3) 采用每一类已训练好的三个特征子空间对待分类的生物细胞显微镜图像重建三种图像特征,得到由每一类中各特征子空间对待分类图像特征的重建结果,通过各个特征子空间重建结果与原提取的图像特征向量的比较获得分类图像对每一类的分类置信度;将待分类图像将分类到置信度最高的类别当中,具体包括如下步骤:

1) 使用每一类已训练好的相应特征子空间对待分类的生物细胞显微镜图像的三种图像特征分别进行重建,得到由每一类中各特征子空间对待分类图像特征的重建结果,即 Y_1^m 、 Y_2^m 、 Y_3^m ;其中, $m=1, \dots, C$, C 为总类别个数;

2) 使用每一类已训练好的相应特征子空间对 Y_1 、 Y_2 和 Y_3 分别进行重建Preimage reconstruction,得到由每一类中各特征子空间对待分类图像特征的重建结果,即

$$Y_1^m = KS_1^m \cdot Y_1;$$

$$Y_2^m = KS_2^m \cdot Y_2;$$

$$Y_3^m = KS_3^m \cdot Y_3;$$

其中, $m=1, \dots, C$, C 为总类别个数; Y^m 是用第 m 类的特征子空间对 Y 的重建结果;分别计算各个特征子空间重建结果与原特征向量的距离:

$$D_1 = (Y_1 - Y_1^m)^2;$$

$$D_2 = (Y_2 - Y_2^m)^2;$$

$$D_3 = (Y_3 - Y_3^m)^2;$$

其中, D_1 、 D_2 、 D_3 为三个特征子空间重构结果 Y^m 与原始特征数据 Y 之间的距离;将计算得到的距离 D_1 、 D_2 、 D_3 映射成为分类概率:

$$p_n = \exp(-D_n), n=1, \dots, 3;$$

其中, $\exp()$ 为以自然对数 e 为底指数函数;

3) 将计算得到的距离映射成为分类概率:

$$p_n = \exp(-D_n), n=1, \dots, 3;$$

其中, $\exp()$ 为以自然对数 e 为底指数函数, D_n 为上一步骤计算得到的距离;

集成三个特征子空间得到的分类概率,使待分类的生物细胞显微镜图像 x 得到对每一类的最终分类置信度:

$$y_m(x) = \frac{\prod_m p_m(x|w_T)}{\prod_m p_m(x|w_T) + \prod_m p_m(x|w_C)};$$

其中, $\prod_m p_m(x|w_T)$ 为判断当前分类为真的概率, $\prod_m p_m(x|w_C)$ 为判断为其他类别的概率;

经过所有特征子空间的分类概率集成,待分类图像对于每一类都计算得到一个分类置信度 $y(x)$, 根据分类置信度进行分类, 将待分类图像分类到置信度最高的类别当中。

2. 根据权利要求1所述的集成方法, 其特征在于所述方法步骤(2)中核主成分分析法KPCA按照如下步骤进行:

1) 将所提取的 n 个图像, 且每一图像由 m 维特征向量描述的一批数据记为一个 $(m \times n)$ 维

数据矩阵 $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$; 计算核矩阵 $K_{\mu\nu} = (\Phi(x_\mu) \cdot \Phi(x_\nu))$; 其中, x 表示矩阵 A 中的行

向量或列向量, Φ 表示映射核函数, 核函数 $K(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{\sigma^2})$; 其中 $x_1, x_2, \dots, x_M$

为训练样本, 用 $\{x_i\}$ 表示输入空间; σ 为函数的宽度参数;

2) 计算 K 的特征值 $\lambda_1, \dots, \lambda_n$ 和对应的特征向量 $v_1, \dots, v_n$; 将特征值按降序排序, 得 $\lambda_1' > \dots > \lambda_n'$; 并对特征向量进行相应调整得 $v_1', \dots, v_n'$;

3) 通过施密特正交化方法单位正交化特征向量, 得到 $a_1, \dots, a_n$;

4) 计算特征值的累积贡献率 $B_1, \dots, B_n$, 根据给定的提取效率 p , 如果 $B_t \geq p$, 则提取 t 个主分量 $a_1, \dots, a_t$;

5) 计算核矩阵 K 在提取出的特征向量上的投影 $Y = K|a$, 其中 $a = (a_1, \dots, a_t)$; 所得的投影 Y 即为图像特征数据经KPCA处理后所得的特征子空间。

用于生物细胞显微镜图像分类的特征子空间集成方法

技术领域

[0001] 本发明属于智能图像分析领域,尤其是生物细胞显微镜图像的分类研究,具体涉及一种用于生物细胞显微镜图像分类的特征子空间集成方法。

背景技术

[0002] 细胞显微镜成像技术对于理解细胞的功能以及构造方式是至关重要的,对于许多疾病的诊断、分类,细胞组织镜下活检已经成为了诊断的金标准。然而,大量的病理活检成像对于医学图像分析师来说会造成巨大的工作负担;人工检测有时因为种种外部因素(如疲劳、经验不足等)也很可能造成错检和漏检,因此,自动的细胞显微镜图像检测技术成为近年来研究的热点。

[0003] 目前广泛使用的亚细胞蛋白质位置检测方法是荧光显微法。近年来,有很多基于荧光显微镜图像的细胞检测方法,以正细胞图像分割为基础,显型检测问题就成为了一个多类的图像分类问题,包含两个主要的步骤:特征表示和分类。

[0004] 一般来说,大部分的基于图像的细胞结构分析使用多种图像特征的组合来表示图像,例如,形态学、边缘、纹理、几何特征、矩,以及小波特征。近年来,图像特征描述方法的进步产生了一些“拿来即用”的特征提取方法,这些方法可以直接应用于生物图像分析领域。

[0005] 另一方面,大量的机器学习和智能计算方法已经进入到生物体图像分析领域,例如人工神经网络、支持向量机等。这些工具已经广泛地用于荧光显微镜图像的分类以及亚细胞蛋白质的分类中。然而,生物显微镜图像具有一个显著的特点:图像具有较大的类间相似性和类内相异性,这决定了使用常用的分类器难以达到很好的分类边界。此外,由于多种图像特征的组合使用,特征的维数不断增加,使得特征向量的维数超过了训练集的样本数。本发明因此而来。

发明内容

[0006] 本发明目的在于提供一种用于生物细胞显微镜图像分类的特征子空间集成方法,解决了现有技术中图像分类效果差、分类精度差等问题。

[0007] 为了解决现有技术中的这些问题,本发明提供的技术方案是:

[0008] 一种用于生物细胞显微镜图像分类的特征子空间集成方法,其特征在于所述方法包括以下步骤:

[0009] (1)通过以下方法提取待分类的生物细胞显微镜图像的特征:

[0010] i)将生物细胞显微镜图像变换到不同的频率子带(Sub-band)中,然后对各个频率子带进行特征统计;以及

[0011] ii)采用灰度共生矩阵的多个统计特征得到生物细胞显微镜图像的全局纹理特征;以及

[0012] iii)通过完整局部二值模式提取生物显微镜图像的局部纹理特征;

[0013] (2)使用核主成分分析法(KPCA)对提取的生物细胞显微镜图像的三种图像特征构

建特征子空间模型,使每类生物细胞显微镜图像具有三个特征子空间;

[0014] (3)采用每一类已训练好的三个特征子空间对待分类的生物细胞显微镜图像重建三种图像特征,得到由每一类中各特征子空间对待分类图像特征的重建结果,通过各个特征子空间重建结果与原提取的图像特征向量的比较获得分类图像对每一类的分类置信度;将待分类图像将分类到置信度最高的类别当中。

[0015] 优选的,所述方法步骤(2)中核主成分分析法(KPCA)按照如下步骤进行:

[0016] 1)将所提取的n个图像(每一图像由m维特征向量描述)的一批数据记为一个 $(m \times$

n)维数据矩阵 $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$;计算核矩阵 $K_{\mu\nu} = (\Phi(x_\mu) \cdot \Phi(x_\nu))$;其中,x表示矩阵A中

的行(列)向量, $\Phi_{Y \times Y}$ 表示映射核函数,核函数 $K(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{\sigma^2})$;其中 $x_1, x_2, \dots, x_M$ 为

训练样本,用 $\{x_i\}$ 表示输入空间; σ 为函数的宽度参数;

[0017] 2)计算K的特征值 $\lambda_1, \dots, \lambda_n$ 和对应的特征向量 $v_1, \dots, v_n$;将特征值按降序排序,得 $\lambda'_1 > \dots > \lambda'_n$;并对特征向量进行相应调整得 $v'_1, \dots, v'_n$;

[0018] 3)通过施密特正交化方法单位正交化特征向量,得到 $a_1, \dots, a_n$;

[0019] 4)计算特征值的累积贡献率 $B_1, \dots, B_n$,根据给定的提取效率p,如果 $B_t \geq p$,则提取t个主分量 $a_1, \dots, a_t$;

[0020] 5)计算核矩阵K在提取出的特征向量上的投影 $Y = K|a$,其中 $a = (a_1, \dots, a_t)$;所得的投影Y即为图像特征数据经KPCA处理后所得的特征子空间。

[0021] 优选的,所述方法中进行待分类生物细胞显微镜图像的分类方法具体步骤包括:

[0022] 1)使用每一类已训练好的相应特征子空间对待分类的生物细胞显微镜图像的三种图像特征分别进行重建,得到由每一类中各特征子空间对待分类图像特征的重建结果,即 $Y_1^{m'}$ 、 $Y_2^{m'}$ 、 $Y_3^{m'}$;其中, $m=1, \dots, C$,C为总类别个数;

[0023] 2)使用每一类已训练好的相应特征子空间对Y1、Y2和Y3分别进行重建(Preimage reconstruction),得到由每一类中各特征子空间对待分类图像特征的重建结果,即

[0024] $Y_1^{m'} = KS_1^m \cdot Y1$;

[0025] $Y_2^{m'} = KS_2^m \cdot Y2$;

[0026] $Y_3^{m'} = KS_3^m \cdot Y3$;

[0027] 其中, $m=1, \dots, C$,C为总类别个数; $Y^{m'}$ 是用第m类的特征子空间对Y的重建结果。

[0028] 分别计算各个特征子空间重建结果与原特征向量的距离:

[0029] $D_1 = (Y1 - Y_1^{m'})^2$;

[0030] $D_2 = (Y2 - Y_2^{m'})^2$;

[0031] $D_3 = (Y3 - Y_3^{m'})^2$;

[0032] 其中, D_1, D_2, D_3 为三个特征子空间重构结果 $Y^{m'}$ 与原始特征数据Y之间的距离。

[0033] 将计算得到的距离 D_1, D_2, D_3 映射成为分类概率:

[0034] $p_n = \exp(-D_n)$, $n=1, \dots, 3$;

[0035] 其中, $\exp()$ 为以自然对数e为底指数函数。

[0036] 集成三个特征子空间得到的分类概率 p_1, p_2, p_3 ,使待分类图像 x 得到对每一类的最终分类置信度:

$$[0037] \quad y_m(x) = \frac{\prod_m p_m(x|w_T)}{\prod_m p_m(x|w_T) + \prod_m p_m(x|w_O)};$$

[0038] 其中, $\prod_m p_m(x|w_T)$ 为判断当前分类为真的概率, $\prod_m p_m(x|w_O)$ 为判断为其他类别的概率;经过所有特征子空间的分类概率集成,待分类图像对于每一类都计算得到一个分类置信度 $y(x)$,根据分类置信度进行分类,将待分类图像分类到置信度最高的类别当中。

[0039] 本发明技术方案中使用曲线波变换(Curvelet Transform)、灰度共生矩阵的统计特征和局部纹理特征组合的图像描述方法,得到了较好的图像分类效果。采用特征子空间集成机制(Feature Subspace Ensemble)将解决单一分类器所存在的问题,提高分类的效果。针对特征维数问题,采用特征子空间(Feature Subspace)方法可以有效地降低特征维数,同时,还可以提高集成分类器的多样性,进一步提升分类效果。

[0040] 相对于现有技术中的方案,本发明的优点是:

[0041] 本发明技术方案为用于生物细胞显微镜图像分类的特征子空间集成方法,提出特征子空间集成分类器的模型,对于提取的各个图像特征建立特征子空间模型,待分类图像的图像特征将用这些特征子空间模型进行重建,计算重建的特征和原始特征之间的距离并映射成为分类概率,集成各特征子空间得到的分类概率得到最终的综合分类置信度用于图像分类。实验表明,此设计体系可显著提高生物显微镜图像分类的准确率。

附图说明

[0042] 下面结合附图及实施例对本发明作进一步描述:

[0043] 图1为乳腺癌细胞显微镜图像,(a)为局限性肿瘤,肿瘤细胞局限于组织某一区域内(箭头所指位置);(b)为弥漫性肿瘤,细胞组织被肿瘤完全占位;(c)为正常组织细胞;

[0044] 图2为本发明技术方案用于提取图像局部纹理特征的LBP模型;

[0045] 图3为不同数据描述方法的结果比较;包括KPCA、支持向量数据描述(svdd),主成分分析(pca),帕尔森窗口(parzen),k均值聚类(kmeans)以及k近邻(knndd);

[0046] 图4为本发明不同数据描述方法的ROC曲线

[0047] 图5为本发明(FLD)与其他分类器的分类稳定度比较结果。

具体实施方式

[0048] 以下结合具体实施例对上述方案做进一步说明。应理解,这些实施例是用于说明本发明而并不限于限制本发明的范围。实施例中采用的实施条件可以根据具体厂家的条件做进一步调整,未注明的实施条件通常为常规实验中的条件。

[0049] 实施例

[0050] 本实施例用于生物细胞显微镜图像分类的特征子空间集成方法,包括以下步骤:

[0051] (1)通过以下方法提取待分类的生物细胞显微镜图像的特征:

[0052] i)将生物细胞显微镜图像变换到不同的频率子带(Sub-band)中,然后对各个频率子带进行特征统计;以及

[0053] ii) 采用灰度共生矩阵的多个统计特征得到生物细胞显微镜图像的全局纹理特征;以及

[0054] iii) 通过完整局部二值模式提取生物显微镜图像的局部纹理特征;

[0055] (2) 使用核主成分分析法 (KPCA) 对提取的生物细胞显微镜图像的三种图像特征构建特征子空间,使每类生物细胞显微镜图像具有三个特征子空间;

[0056] (3) 采用每一类已训练好的三个特征子空间对待分类的生物细胞显微镜图像重建三种图像特征,得到由每一类中各特征子空间对待分类图像特征的重建结果,通过各个特征子空间重建结果与原提取的图像特征向量的比较获得分类图像对每一类的分类置信度;将待分类图像将分类到置信度最高的类别当中。

[0057] 所要解决的问题是构建可靠准确的生物显微镜图像分类体系,采用技术方案具体描述如下:

[0058] (1) 生物细胞显微镜图像特征描述

[0059] 如图1所示为生物显微镜图像示例。对于如图1所示的乳腺癌细胞显微镜图像,(a) 图为局限性肿瘤,肿瘤细胞局限于组织某一区域内(箭头所指位置);(b) 图为弥漫性肿瘤,细胞组织被肿瘤完全占位;(c) 图为正常组织细胞。本发明采用多种图像特征提取方法进行特征提取。使用的特征提取方法包括:

[0060] 1) 曲线波变换 (Curvlet Transform)

[0061] 曲线波变换是最近提出的一种非自适应变换,和小波变换相比,曲线波变换具有提取图像方向性特征的能力,例如图像中的边缘。曲线波变换将生物细胞显微镜图像变换到不同的频率子带 (Sub-band) 中,然后对各个频率子带进行特征统计,均值、方差和熵用作子带的统计特征。当每幅图像变换后具有n个子带时,将得到一个3n维度的曲线波特征向量。

[0062] 2) 灰度共生矩阵 (Gray Level Co-occurrence Matrix) 统计特征

[0063] 生物细胞显微镜图像的全局纹理特征使用灰度共生矩阵的多个统计特征组合得到。灰度共生概率提供了一种生成图像特征的二阶方法。共提取22个基于灰度共生矩阵的统计特征用于描述生物细胞显微镜图像的全局纹理特征,见表1。

Indices	Features	Indices	Features
[0064]	1 Energy	12	Sum of Squares
	2 Entropy	13	Sum Average
	3 Dissimilarity	14	Sum Variance
	4 Contrast	15	Sum Entropy
	5 Inverse Difference	16	Difference Variance
	6 Correlation	17	Difference Entropy
	7 Homogeneity	18	Information Measure of Correlation (1)
	8 Autocorrelation	19	Information Measure of Correlation (2)
	9 Cluster Shade	20	Maximal Correlation Coefficient
	10 Cluster Prominence	21	Inverse Difference Normalized
	11 Maximum Probability	22	Inverse Difference Moment Normalized

[0065] 表1用于描述生物显微镜图像的22个灰度共生矩阵统计特征

[0066] 3) 局部纹理特征

[0067] Completed Local Binary Pattern (CLBP) 用于提取生物显微镜图像的局部纹理特征。LBP的工作流程见图2,为用于提取图像局部纹理特征的LBP模型。将提取三个通道的LBP,分别为CLBP_S,CLBP_M和CLBP_C,三个通道的特征最终生成一个3维的交叉直方图用于

描述图像局部纹理特征。

[0068] (2) 基于核主成分分析的特征子空间集成分类

[0069] 1、核主成分分析特征子空间的构造

[0070] 本发明使用核主成分分析KPCA(Kernel Principle Component Analysis)构造图像特征子空间。假设 $x_1, x_2, \dots, x_M$ 为训练样本,用 $\{x_i\}$ 表示输入空间。KPCA方法的基本思想是通过某种隐式方式将输入空间映射到某个高维空间,并且在特征空间中实现PCA。

[0071] 对于上述所提取的图像的每一种特征,由KPCA处理过程得到每一特征的特征子空间,处理过程如下:

[0072] 1)、将所提取的 n 个图像(每一图像由 m 维特征向量描述)的一批数据写成一个 $(m \times n)$ 维数据矩阵

$$[0073] \quad A = \begin{pmatrix} a_{11} & K & a_{1n} \\ M & O & M \\ a_{m1} & L & a_{mn} \end{pmatrix}.$$

[0074] 2)、计算核矩阵:

$$[0075] \quad K_{\mu\nu} = (\Phi(x_\mu) \cdot \Phi(x_\nu));$$

[0076] 其中, x 表示矩阵 A 中的行(列)向量, Φ_{YXY} 表示映射核函数,本发明选用高斯径向

$$\text{核函数 } K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{\sigma^2}\right).$$

[0077] 3)、计算 K 的特征值 $\lambda_1, \dots, \lambda_n$ 和对应的特征向量 $v_1, \dots, v_n$ 。

[0078] 4)、特征值按降序排序(通过选择排序)得 $\lambda'_1 > \dots > \lambda'_n$ 并对特征向量进行相应调整得 $v'_1, \dots, v'_n$ 。

[0079] 5)、通过施密特正交化方法单位正交化特征向量,得到 $a_1, \dots, a_n$ 。

[0080] 6)、计算特征值的累积贡献率 $B_1, \dots, B_n$,根据给定的提取效率 p ,如果 $B_t \geq p$,则提取 t 个主分量 $a_1, \dots, a_t$ 。

[0081] 7)、计算核矩阵 K 在提取出的特征向量上的投影 $Y = K|a$,其中 $a = (a_1, \dots, a_t)$ 。

[0082] 所得的投影 Y 即为特征数据经KPCA处理后所得的特征子空间。

[0083] 对本发明使用的三种特征提取方法分别使用KPCA进行特征子空间生成后,每一类图像得到三个特征子空间,记为 KS_n^m ,其中, m 为类标号, $n(1 \leq n \leq 3)$ 为特征子空间标号。

[0084] 2、特征子空间集成分类

[0085] 本发明使用特征子空间集成方法对一幅待分类图像分类的具体步骤如下:

[0086] 1)对待分类图像分别提取三种图像特征(curvelet、GLCM、CLBP),分别记为 Y_1 、 Y_2 和 Y_3 。

[0087] 2)使用每一类已训练好的相应特征子空间对 Y_1 、 Y_2 和 Y_3 分别进行重建(Preimage reconstruction),得到由每一类中各特征子空间对待分类图像特征的重建结果,即

$$[0088] \quad Y_1^{m'} = KS_1^m \cdot Y_1;$$

$$[0089] \quad Y_2^{m'} = KS_2^m \cdot Y_2;$$

$$[0090] \quad Y_3^{m'} = KS_3^m \cdot Y_3;$$

[0091] 其中, $m=1, \dots, C$, C 为总类别个数; $Y^{m'}$ 是用第 m 类的特征子空间对 Y 的重建结果。

[0092] 分别计算各个特征子空间重建结果与原特征向量的距离:

$$[0093] \quad D_1 = (Y_1 - Y_1^m)^2;$$

$$[0094] \quad D_2 = (Y_2 - Y_2^m)^2;$$

$$[0095] \quad D_3 = (Y_3 - Y_3^m)^2;$$

[0096] 其中, D_1, D_2, D_3 为三个特征子空间重构结果 Y^m 与原始特征数据 Y 之间的距离。

[0097] 将计算得到的距离 D_1, D_2, D_3 映射成为分类概率:

$$[0098] \quad p_n = \exp(-D_n), n=1, \dots, 3;$$

[0099] 其中, $\exp()$ 为以自然对数 e 为底指数函数, D_n 为上一步骤计算得到的距离。

[0100] 集成三个特征子空间得到的分类概率 p_1, p_2, p_3 , 使待分类图像 x 得到对每一类的最终分类置信度:

$$[0101] \quad y_m(x) = \frac{\prod_m p_m(x|w_T)}{\prod_m p_m(x|w_T) + \prod_m p_m(x|w_G)}$$

[0102] 其中, $\prod_m p_m(x|w_T)$ 为判断当前分类为真的概率, $\prod_m p_m(x|w_G)$ 为判断为其他类别的概率。经过所有特征子空间的分类概率集成, 待分类图像对于每一类都计算得到一个分类置信度 $y(x)$, 此分类置信度将直接用于分类, 待分类图像将分类到置信度最高的类别当中。

[0103] 本发明应用于一个具有3个类别的乳腺癌细胞显微图像开放测试图像库的参数设定如下:

[0104] 特征提取: Curvelet 采用5层分解, 共82个子带, 每个子带3个特征, 共246维;

[0105] 灰度共生矩阵共提取表1中22个特征, 每个维度为10, 共220维;

[0106] 局部纹理特征 CLBP 直方图的维数为200。

[0107] 特征子空间构成: Kernel PCA 的核函数采用高斯核函数 $K(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{\sigma^2})$,

其中 σ 取值为0.5。经KPCA投影后选取前35个主分量构成特征子空间。

[0108] 图像特征重建: 从特征子空间重建图像特征由以下规则完成:

[0109] 满足上式的最小值 $\hat{x}$ 可由梯度下降法迭代计算得到。

[0110] 实验图像数据库为具有3类的乳腺癌细胞显微镜图像, 共有360幅图像, 其中, 弥漫型癌细胞类有102幅图像; 局限性癌细胞类有118幅图像; 正常乳腺细胞图像有140幅。

[0111] 实验开展方式为执行10次10折的交叉验证(10-fold cross validation), 每次将每类图像分为10份, 9份轮流作为训练集用于得到特征子空间, 1份用作测试集用于得到模型的分类结果。10次10折交叉验证的分类结果取平均得到实验结果。

[0112] 图3为采用本发明技术方案的准确率(Precision)--召回率(Recall)曲线, Precision就是检索出来的图像中有多少是准确的, Recall就是所有该类的图像有多少被检索出来了, 理想的分类器应该使得两者都高, 即在图中曲线的拐点越靠近右上角越好。

[0113] 作为比较, 在实验中使用了其他数据描述方法用于生成特征子空间, 比较的方法分别为: 支持向量数据描述(svdd), 主成分分析(pca), 帕尔森窗口(parzen), k均值聚类(kmeans)以及k近邻(knndd)。从图3不同数据描述方法的结果比较可以看出, 以kpca作为特征子空间构造方法取得了明显的分类优势。

[0114] 图4给出了不同数据描述方法的受试者工作特征曲线(receiver operating characteristic curve,简称ROC曲线)。ROC曲线是反映敏感性和特异性连续变量的综合指标,是用构图法揭示敏感性和特异性的相互关系,它通过将连续变量设定出多个不同的临界值,从而计算出一系列敏感性和特异性,再以敏感性为纵坐标、(1-特异性)为横坐标绘制成曲线,曲线下面积越大,诊断准确性越高。在ROC曲线上,最靠近坐标图左上方的点为敏感性和特异性均较高的临界值。从图4可以看出,以kpca为基础的特征子空间方法取得了较好的分类敏感性和特异性。

[0115] 图5为本发明提出的特征子空间集成方法与其他常见分类器的分类结果比较。本发明提出的方法(Feature Subspace)得到了较好的分类精度(98.67%),其他比较的分类器为支持向量机(SVM),多层感知器神经网络(MLP),Fisher线性判别(FLD)和朴素贝叶斯分类器(**Naïve Bayes,NB**)。从图中可以看出,本发明提出的方法较其他分类器取得了更好的分类精度。

[0116] 上述实例只为说明本发明的技术构思及特点,其目的在于让熟悉此项技术的人是能够了解本发明的内容并据以实施,并不能以此限制本发明的保护范围。凡根据本发明精神实质所做的等效变换或修饰,都应涵盖在本发明的保护范围之内。

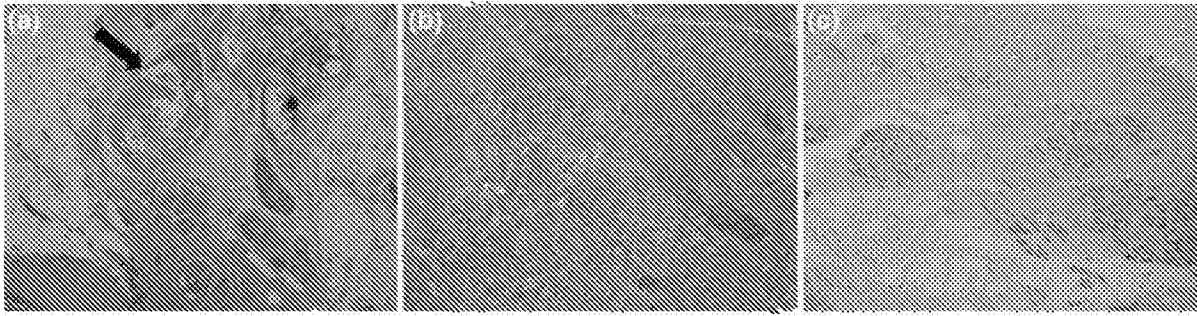


图1

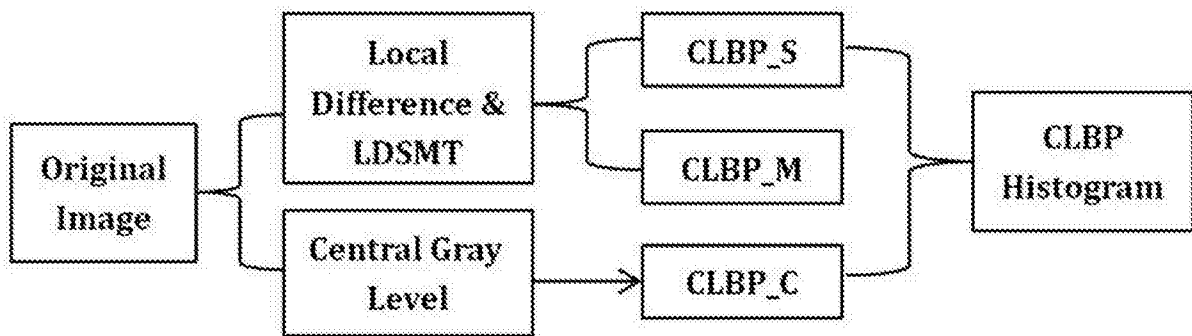


图2

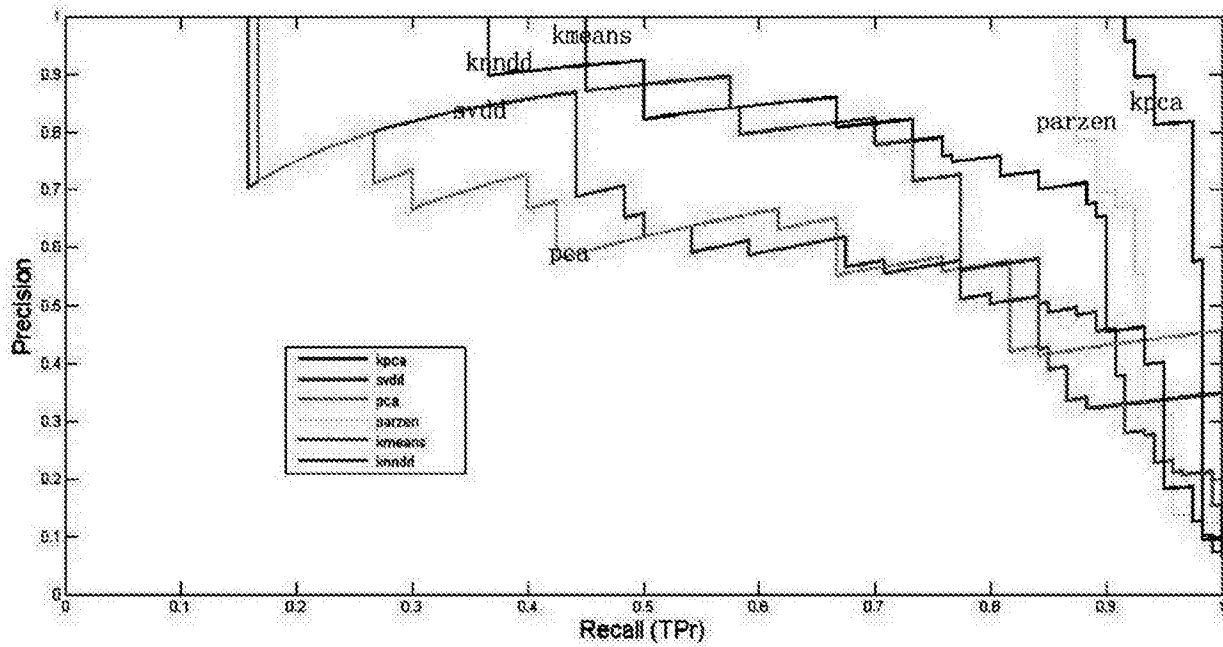


图3

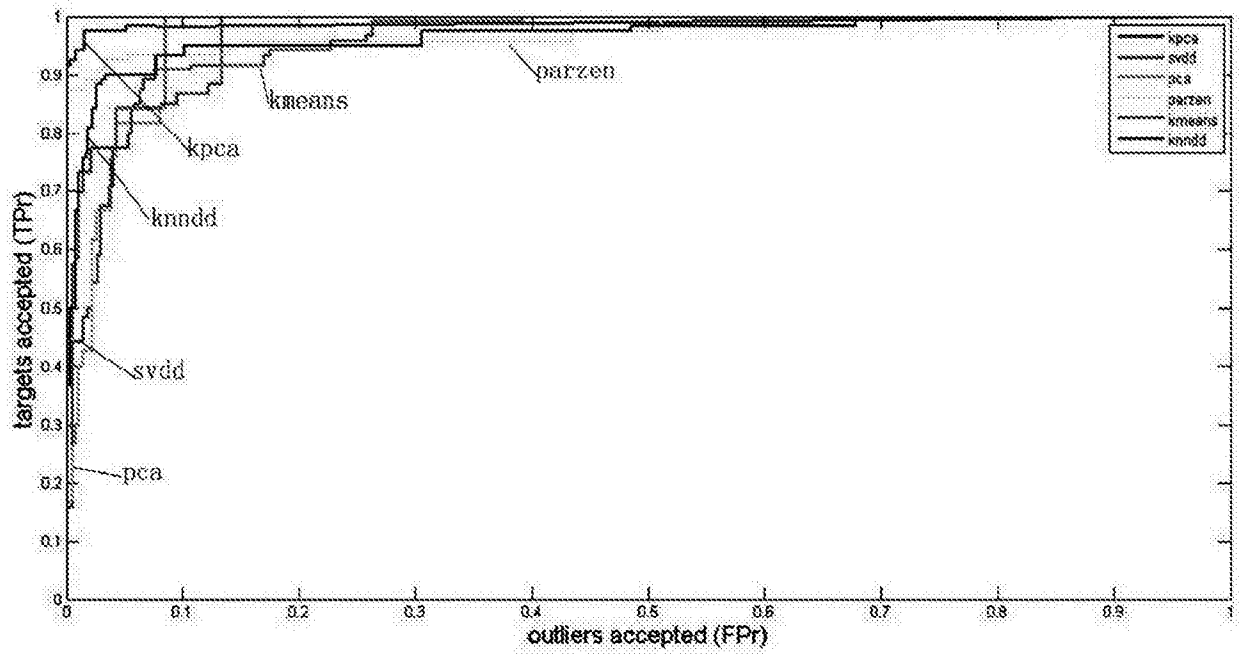


图4

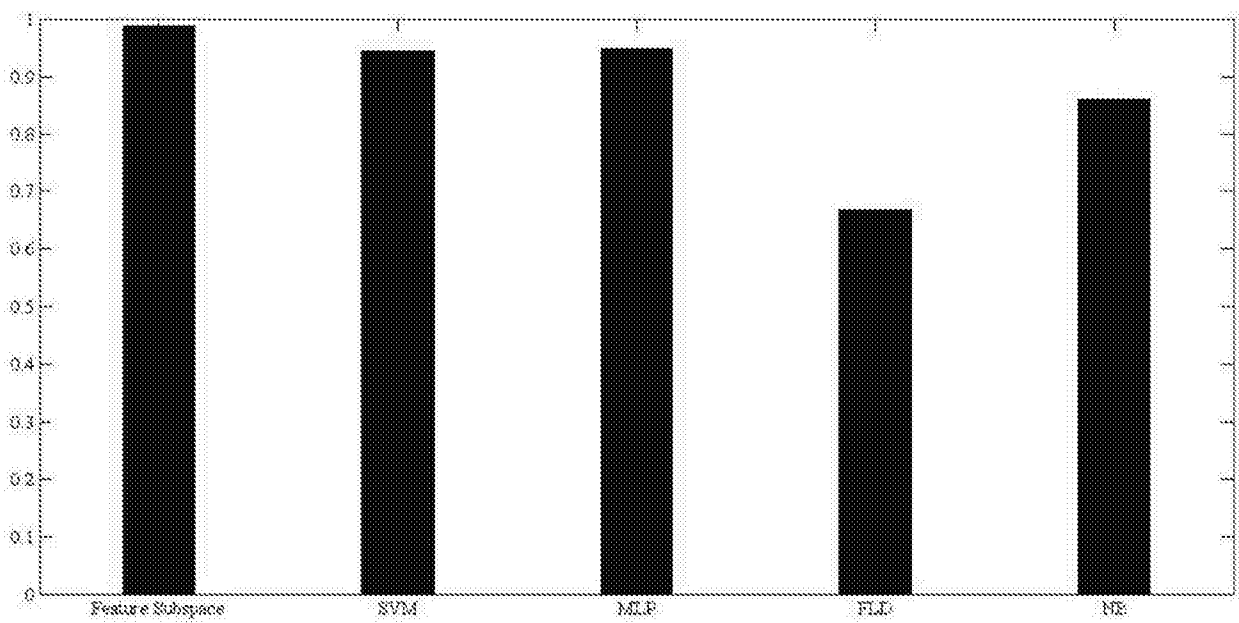


图5