

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

G06F 19/00

//A61B5/00, G06F159:00



[12] 发明专利说明书

[21] ZL 专利号 02103795.7

[45] 授权公告日 2004 年 8 月 18 日

[11] 授权公告号 CN 1162798C

[22] 申请日 2002.3.25 [21] 申请号 02103795.7

[71] 专利权人 北京工业大学

地址 100022 北京市朝阳区平乐园 100 号

[72] 发明人 沈兰荪 卫保国

审查员 张美菊

[74] 专利代理机构 北京思海天达知识产权代理有限公司

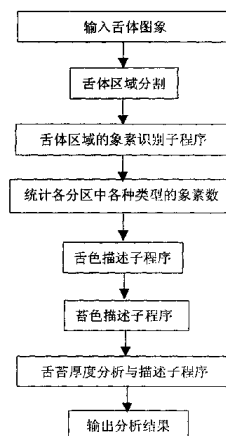
代理人 张 慧

权利要求书 4 页 说明书 14 页 附图 7 页

[54] 发明名称 基于多类支持向量机的中医舌色、苔色、舌苔厚度分析方法

[57] 摘要

一种基于多类支持向量机的中医舌色、苔色、舌苔厚度分析方法，应用于计算机医学图像处理领域。该方法是由数码相机进行图象采集，并通过 USB 接口将图象输入到计算机处理器中，在处理器中将舌体区域从舌图象中分割出来，特征在于该方法还依次包括下述步骤：在进行分级聚类分析生成聚类树并调整的基础上，采用基于聚类树的支持向量机(CTSVM)方法进行舌体区域像素识别，并把舌体区像素识别的类别数设为 15 种；将舌面分为五个区域，统计各区中数量最多的像素类别，及各种类型的象素数；进行舌色描述；进行舌苔色描述；进行舌苔厚度定量分析和描述；最后采用文字和伪彩色图显示分析结果。该方法快速、准确达到实用，符合中医专家的诊察原则。



ISSN 1008-4274

权利要求书

1、一种基于多类支持向量机的中医舌色、苔色、舌苔厚度分析方法，是由数码相机进行图象采集，并通过 USB 接口将图象输入到计算机处理器中，在处理器中采用通用技术，将舌体区域从舌图象中分割出来，本发明的特征在于该方法还依次包括下述步骤：

1) 在进行分级聚类分析生成聚类树并调整的基础上，采用基于聚类树的支持向量机方法进行舌体区域像素识别，以下将基于聚类树的支持向量机简称为 CTSVM；

分级聚类分析生成聚类树的方法步骤为：

- (1) 对于 k 类样本，计算得到 k 个类中心集合： $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k)$ ，
- (2) 第 1 级划分是每一个类中心各成一类，即 k 个叶节点，
- (3) 求每两个类中心之间的距离 D_{ij} ， $i=1, \dots, k$ ， $j=1, \dots, k$ ， $i \neq j$ ，
- (4) 将距离最近的两个中心合并，形成新的类中心，即中间节点，
- (5) 重复进行 (3) - (4) 步，直到第 k-1 级，所有的中心合并为 1 类，即根节点，
- (6) 根据需要，调整聚类树的结构，使得树的深度尽可能少，生成的聚类树与两类模式识别支持向量机结合，就构成了 CTSVM；

舌体区域像素识别的方法和的步骤为：

- (1) 计算输入图象中每个像素 3×3 邻域内的三色色度值 R、G、B 的均值，作为像素识别的输入特征，
- (2) 读取舌体中一个像素的 R、G、B 值，
- (3) 从所述的聚类树的根节点开始，根据该节点的 CTSVM 分类器，将该像素划分到某一分支，具体方法为：

计算该像素与该节点对应的每一个支持向量的核函数内积，并乘以相应的系数和标号，对所有这些乘积求和，然后求加上常数 b 之后的结果的符号。即求：

$$f(x) = \text{sign} \left(\sum_{x_i \in SV} \alpha_i y_i K(x_i, x) + b \right)$$

式中， x_i 是该分类器的支持向量， α_i 是与支持向量对应的系数， y_i 是支持

权 利 要 求 书

向量样本的标号， b 是常数，本发明中 b 取 40， $K(x_i, x)$ 是核函数内积，

$$K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$$

式中 $\|x-y\|$ 是向量 $x-y$ 的范数， σ 取 50，

$Sign()$ 是符号函数，如果自变量的值大于 0，则函数值为 1，小于 0，则为 -1。

然后根据 $f(x)$ 的值和该 CTSVM 的分类规则，将该像素划分到某一分支；

(4) 如果分支节点不是叶节点，则重复进行 (3) 中的过程：根据该节点的 CTSVM 分类器，将该像素划分到某一分支，如果分支节点是叶节点，则该叶节点所代表的类即为所识别的像素的类别，各节点的分类器是两类支持向量机分类器，分类过程按照常规的方法实现，

(5) 重复 (2) - (4) 的过程，直到舌体区域的全部像素分类完毕；

2) 统计各分区中各种类型的像素数；

3) 进行舌色描述，根据舌侧区与舌尖区像素的舌质类型描述舌质的颜色，计算舌尖、舌左侧、舌右侧等 3 个区域中属于各舌质类型的像素总数，像素数最多的舌质类型即为该舌图象的舌质特征；

4) 进行舌苔色描述，计算舌根、舌中、舌尖等 3 个区域中属于各舌苔类型的像素总数，在三个区域中，分别判断像素数最多的舌苔类型的像素数与整个舌面像素总数之比是否小于某一阈值，确定该区的舌苔特征；

5) 进行舌苔厚度定量分析和描述，根据像素及其邻域内像素的舌色、舌苔色类型和数量，分为“无苔”、“薄苔”、“中厚苔”、“厚苔”4 种类型，根据整个舌体区域内像素的厚度类型和相应的数量，计算舌苔厚度指数，根据厚度指数描述舌苔厚度；

6) 采用文字和伪彩色图显示分析结果。

2、根据权利要求 1 所述的基于多类支持向量机的中医舌色、苔色、舌苔厚度分析方法，其舌色、苔色、舌苔厚度的描述特征在于，像素识别完成后，统计各分区中各种类型的像素数后，令 N_i^j 代表 i 区中第 j 类像素个数，其中 $i=1, \dots, 5$ 分别表示舌根、舌中、舌尖、舌左侧、舌右侧等 5 个区域， $j=1, \dots, 15$ 表示舌质淡、舌质淡红、舌质红、舌质暗红、舌质绛红、舌质暗紫、薄白苔、

权 利 要 求 书

白苔、白厚苔、薄黄苔、黄苔、黄厚苔、灰苔、褐苔、黑苔等 15 种舌质和舌苔类型，在此基础上，根据中医专家的舌诊习惯，进行舌色、苔色、舌苔厚度的描述；

舌色的描述方法为根据舌侧区与舌尖区像素的舌质类型描述舌质的颜色，计算舌尖、舌左侧、舌右侧等 3 个区域中属于各舌质类型的像素总数，像素数最多的舌质类型即为该舌图象的舌质特征，具体过程为：

1) 计算总面积（像素的总数） A ，

2) 分别计算舌侧与舌尖区 6 种舌质类型的总数

$$SUM^j = N_3^j + N_4^j + N_5^j, \quad j=1, \dots, 6,$$

3) 令 $SUM^{j_{\max}} = \max(SUM^j, j=1, \dots, 6)$ ，则 j_{\max} 代表的舌质类型为该舌图象的舌质特征，

4) 计算暗紫舌质的面积比 $R_6 = \frac{SUM^6}{A}$ ，若 $R_6 > \theta_p$ 且 $j_{\max} \neq 6$ ，则增加描述“局部暗紫”，其中 θ_p 为根据实验设定的阈值，

5) 计算舌尖区“舌质红”的面积比 $R_3 = \frac{N_3^3}{A}$ ，若 $R_3 > \theta_r$ 且 $j_{\max} \neq 3$ ，则增加描述“舌尖红”，其中 θ_r 为根据实验设定的阈值；

苔色的描述方法为计算舌根、舌中、舌尖等 3 个区域中属于各舌苔类型的像素总数，在三个区域中，分别判断像素数最多的舌苔类型的像素数与整个舌面像素总数之比是否小于某一阈值，确定该区的舌苔特征，具体方法如下：

1) 计算舌面的总面积（像素的总数） A ，舌根、舌中、舌尖区的面积 $A_i, i=1, \dots, 3$ ，

2) 对于舌根、舌中、舌尖区，分别计算 $N_i^{j'_{\max}} = \max(N_i^j, j=7, \dots, 15)$ ，其中 $i=1, 2, 3$ ，则 j'_{\max} 代表的舌苔类型为该区的舌苔特征，

3) 计算 $i=1, 2, 3$ 区中 j'_{\max} 舌苔象素所占的面积比例 $R(j'_{\max}, i) = \frac{N_i^{j'_{\max}}}{A_i}$ ，

4) 分别对舌根、舌中、舌尖区进行描述，对于舌根区 $i=1$ ，若 $R(j'_{\max}, 1) < \theta_1$ 则描述为“舌根 少苔”，否则描述舌根为 j'_{\max} 类型的舌苔，对于舌中区 $i=2$ ，若 $R(j'_{\max}, 2) < \theta_2$ ，则描述为“舌中 少苔”，否则描述舌中为 j'_{\max} 类型的舌苔，对于舌尖区 $i=3$ ，若 $R(j'_{\max}, 3) < \theta_3$ ，则描述为“舌尖 少苔”，否则描述舌尖为 j'_{\max} 类

权 利 要 求 书

型的舌苔，其中 θ_1 、 θ_2 、 θ_3 为根据实验设定的阈值，

5) 对于舌根、舌中区，分别计算第二主要类型舌苔 $j_{\text{sec}}^i, i=1,2$ ，并计算 j_{sec}^i 类型像素面积与 j_{max}^i 类型的面积比 $R_{ms}^i = \frac{N_{j_{\text{sec}}^i}}{N_{j_{\text{max}}^i}}$ ，若 $R_{ms}^i > \theta_{\text{sec}}$ ，则第 i 区增加描述第二主要类型舌苔，其中 θ_{sec} 为根据实验设定的阈值；

舌苔厚度的分析和描述分为两部分：舌苔厚度分类及舌苔厚度定量化，具体步骤为：

- 1) 读取舌体像素及其 5×5 邻域的分类号，
- 2) 设 c_1 、 c_2 为根据实验设定的阈值，且 $c_1 > c_2$ ，计算该像素 5×5 邻域内属于舌质类型的像素数 S_{body} ，若 $S_{\text{body}} > c_1$ ，则该点的舌苔厚度类型定为“无苔”，
- 3) 若 $c_1 \geq S_{\text{body}} > c_2$ ，则该点的舌苔厚度类型定为“薄苔”，
- 4) 若 $S_{\text{body}} \leq c_2$ ，则依据该点的舌质舌苔类型 j 确定厚度类型，若为舌质即 $j < 7$ ，则为“薄苔”，若 $j=7、10$ ，则为“薄苔”，若 $j=8、11、13、14$ ，则为“中厚苔”，若 $j=9、12、15$ ，则为“厚苔”，
- 5) 重复1)-3)步，直到所有像素处理完毕，
- 6) 根据象素识别的结果，计算整幅图象的苔厚指数 T ，作为舌苔厚度分析的定量化结果，计算方法为：

$$T = \frac{\sum_k (w_{j^k})}{A}$$

式中 A 表示舌体区的像素总数， \sum_k 表示对舌体区域中所有象素求和， j^k 表示第 k 个象素的厚度类型，对应于“无苔”、“薄苔”、“中厚苔”、“厚苔”分别取0、1、2、3， w_{j^k} 为权值，根据实验和中医专家的经验确定，苔厚指数能够描述舌苔的总体厚度。

说明书

基于多类支持向量机的中医舌色、苔色、舌苔厚度分析方法

技术领域

本发明涉及计算机医学图像处理领域，设计了一种基于多类支持向量机的中医舌图像分析系统，对舌体图象进行舌色、苔色、舌苔厚度的定量分析及定性描述，辅助中医诊断。

背景技术

舌诊是中医四诊中望诊的重要内容。舌诊是靠医生对舌体进行观察，从而作出判断。这种方法带有很强的主观性、模糊性，诊察结果与医生的经验密切相关。目前人们争取对舌象特征进行客观定量的分析，克服传统目视观察的主观性，提高舌诊的可靠性，具有很大的优势。

采用计算机图象处理技术进行舌象特征分析，能够比较全面地反映舌象的情况，符合中医舌诊的习惯，便于舌诊资料的收集与保存，具有较好的实用价值。它是在一定的光照条件下，采集受试者的彩色舌图象，数字化后输入计算机，利用图象分析技术，对舌象特征进行自动分析，并将结果储存在计算机中。

已有的图像处理技术用于舌象分析时，大多只是采用舌图象颜色值简单统计分析。经过分析我们发现，舌色、苔色与厚度的分布具有局部性，在舌面上的不同区域往往表现出不同的特征，中医舌诊时先观察局部，再进行总体描述。描述时主要强调具有辩证意义的特征。

在图象处理或分析领域，由于图象拍摄的条件不同，处理或分析方法中的参数可能会根据实际情况调整。调整的方法是，从相同条件拍摄的图象中，取出一幅或几幅，通过实验确定参数的具体值。本方法中的各个参数基本都按照此方法确定。

已有的模式识别技术中的支持向量机（SVM）方法是目前解决小样本识别问题比较有效的方法，基本思想可用图 3 的两维情况说明。图 3 中，实心点和空心点代表两类样本，H 为分类线， H_1 、 H_2 分别为过各类中离分类线最

说明书

近的样本且平行于分类线的直线，它们之间的距离叫做分类间隔（margin）。所谓最优分类线就是要求分类线不但能将两类正确分开（训练错误率为 0），而且使分类间隔最大。

从图 3 中可以看出，(b) 中的分类间隔大，其分类错误的可能性较 (a) 中小，位于最优分类线上的样本称为支持向量，以上是线性可分情况的例子。对于非线性情况，通过非线性变换转换到高维的特征空间(Feature Space)，在高维空间中构造线性判别函数来实现原空间中的非线性判别函数。

舌象样本包括采集的数字化舌图象及舌诊专家给出的临床舌诊结论。样本数量越多，临床判断越可靠，则在此基础上建立的模型越符合实际，对新的舌象进行特征分析的准确度也就越高。

然而，有些类型的舌象是非常少见的（如灰黑苔、淡白舌、青紫舌等）。因此各类样本的分布不均匀，数量有限，如采用一般的模式识别方法，势必影响舌象特征自动分析的准确性和可靠性。因此采用支持向量机（SVM）模式识别方法进行像素识别，以提高分类的准确度，减小分类错误率。

经典的 SVM 一般针对两类识别问题，而舌体区域象素识别是一个多类问题。当用于多类模式识别时，通常将原问题分解为多个两类识别问题。现有的方法包括 1-a-a（one-against-all）和 1-a-1（one-against-one）等。其中 1-a-a 由 k 个两类 SVM (f^1, \dots, f^k) 组成， f^j 是将第 j 类样本的类别标号定为 +1，而将其它所有样本的类别标号定为 -1，1-a-1 则将每一类样本与其它各类分别构成两类识别问题，共构成 $\frac{k(k-1)}{2}$ 个两类 SVM。应用时，样本经过所有的两类 SVM，得到 $\frac{k(k-1)}{2}$ 个识别结果，其中最占优势的类别属性为最终识别结果。这两种方法的缺点是训练后得到的支持向量数较多，影响应用时的识别速度。

发明内容

为了解决前述问题，可自动、准确、快速分析舌体图象的舌色、苔色、舌苔厚度等特征，本发明设计了一种新的多类方法——基于聚类树的支持向量机（CTSVM）方法。这种方法采用先进行舌体区域的象素识别（局部分析），再进行总体描述的“两步法”分析步骤。像素识别时，结合了支持向量机能

说明书

够有效完成小样本识别问题，以及聚类树采用分级方法将多类问题转化为两类问题的特点，因而快速、准确达到实用。总体描述依据主要像素类别进行，符合中医专家的诊察原则。同时引入了苔厚指数，使得舌苔厚度定量化。采用这种方法能够用计算机获得舌图象中舌色、苔色、舌态厚度的定性描述和定量分析。本发明的技术思路特征在于：

1、将图象处理方法用于舌象特征分析时，不同于一般的方法，而是采用了先进行舌体区域的像素识别（局部分析），再进行总体描述的“两步法”分析步骤。由于中医舌诊时先观察局部，再进行总体描述。因而该方法与中医的诊断习惯相吻合，能够有效识别舌象特征。

2、我们采用了先进行舌体区域的像素识别（局部分析），再进行总体描述的分析步骤。也即对舌质、舌苔的特征进行局部分析。将舌体上的像素根据颜色属性，采用多类支持向量机划分为不同的类别。我们把舌体区像素识别的类别数设为 15 种，包括舌质淡、舌质淡红、舌质红、舌质暗红、舌质绛红、舌质暗紫、薄白苔、白苔、白厚苔、薄黄苔、黄苔、黄厚苔、灰苔、褐苔、黑苔等。其中，前六种是舌质类型，后九种是舌苔类型。像素识别时，一般先需要给出已知类别的像素，称为样本。如果样本相对很少，这样的识别问题称为小样本问题。

针对像素识别这个多类、小样本的分类问题，采用了一种新的基于聚类树的多类支持向量机（CTSVM）方法。CTSVM 的主要思想是先对训练样本集的各类中心进行分级聚类分析，生成称为聚类树的多级分类器，再据此构成 $(k-1)$ 个两类 SVM。利用它可以把一个复杂的多类别分类问题转化为若干个简单的分类问题来解决。聚类树的每个非叶子节点代表一个支持向量分类器。聚类树的结构可以根据人工经验确定，也可以采用分级聚类方法得到。所谓聚类，就是根据“物以类聚，人以群分”的原则，把相似性大的样本聚集为一个类型。该方法能够有效的将多类识别问题转化为个数较少、因而识别速度较快的两类分类问题，并且可以克服舌象分析中样本较少对识别率的影响，从而能够较为准确的进行像素识别，并通过分级聚类并结合中医专家

说明书

经验的方法得到聚类树，参见图 4。

聚类树的形式根据训练样本集的分布情况和最近距离准则自适应确定。将这种聚类树与两类模式识别 SVM 结合，就构成了 CTSVM。在 CTSVM 中，每个非叶子节点（根节点和中间节点）都是一个两类支持向量机。

我们根据上述方法，结合中医专家的诊察经验确定的聚类树如图 4 所示。图中，H~M 舌质类型，N~V 是舌苔类型。图中，X 称为根节点，A~G 称为非叶子节点。与一个节点相连的上层节点称为父节点，相应的，该节点称为其父节点的子节点。图 3 中，A 有两个子节点 C 和 D，C 有三个子节点 H、I、J，像素识别时首先划分距离最大的两大类，再逐级进行较细的划分。即由根节点出发，用与该节点相应的分类器进行分类，将它归属与某一子节点，直至到达某个叶节点，即得到该像素的最终识别结果。在本发明中，与每一个非叶子节点相对应的分类器都采用支持向量机。如果一个非叶子节点有两个子节点，则直接采用两类支持向量机即可；如果一个非叶子节点有三个子节点，在该节点就有三个两类支持向量机完成共同像素识别任务，每个两类支持向量机将其中一个子节点代表的类与其他两个子节点代表的类分开。

3、将舌面分为五个区域，统计各区中数量最多的像素类别，对该区舌色、苔色的描述主要根据该类别，结合中医专家的经验进行。这符合中医舌诊描述时主要强调具有辩证意义的特征的原则。

4、舌苔厚度分类时，将舌体区域的每一个象素分为“无苔”、“薄苔”、“中厚苔”、“厚苔”4 种类型。先逐象素舌苔厚度分类，其依据是“舌苔的厚薄以见底、不见底为标准”。再进行舌苔厚度定量化，在逐象素识别的基础上，引入苔厚指数，对整幅舌图象的舌苔厚度进行定量化描述。

5、为了使分析结果易于理解和临床应用，采用文字和伪彩色图显示分析结果。将象素识别结果表示成舌色、苔色分布伪彩色图。将逐象素的厚度分类结果表示成舌苔厚度伪彩色分布图。采用文字对舌色、苔色舌苔厚度进行定性描述，并给出舌苔厚度指数值。

本发明的技术方案见图 1、图 2、图 5、图 6、图 8、图 9、图 10，方法

说明书

步骤为由数码相机进行图象采集，并通过 USB 接口将图象输入到计算机处理器中，在处理器中采用通用技术，将舌体区域从舌图象中分割出来，特征在于该方法还依次包括下述步骤：

1) 在进行分级聚类分析生成聚类树并调整的基础上，采用基于聚类树的支持向量机 (CTSVM) 方法进行舌体区域像素识别；

分级聚类分析生成聚类树的方法步骤见图 5：

- (1) 对于 k 类样本，计算得到 k 个类中心集合： $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k)$ ，
- (2) 第 1 级划分是每一个类中心各成一类，即 k 个叶节点，
- (3) 求每两个类中心之间的距离 D_{ij} ， $i=1, \dots, k$ ， $j=1, \dots, k$ ， $i \neq j$ ，
- (4) 将距离最近的两个中心合并，形成新的类中心，即中间节点，
- (5) 重复进行(3)-(4)步，直到第 $k-1$ 级，所有的中心合并为 1 类，即根节点，
- (6) 根据需要，调整决策树的结构，使得树的深度尽可能少；

舌体区域像素识别的方法和的步骤见图 6：

- (1) 计算输入图象中每个像素 3×3 邻域内的三色色度值 R 、 G 、 B 的均值，作为像素是别的输入特征，
- (2) 读取舌体中一个像素的 R 、 G 、 B 值，
- (3) 从根节点开始，根据该节点的 SVM 分类器，将该像素划分到某一分支，
- (4) 如果分支节点不是叶节点，则重复进行 (3) 中的过程：根据该节点的 SVM 分类器，将该像素划分到某一分支，如果分支节点是叶节点，则该叶节点所代表的类即为所识别的像素的类别，各节点的分类器是两类支持向量机分类器，分类过程按照常规的方法实现，
- (5) 重复 (2) - (4) 的过程，直到舌体区域的全部像素分类完毕。

可以看出，像素识别时，首先划分距离最大的两大类，再逐级进行较细的划分，直至叶节点，就得到像素得类别属性。由于减少了两类判别的次数，同时兼顾样本的分布情况，本方法能够在加快识别速度的同时保证识别的准确率。

2) 统计各分区中各种类型的象素数；

说明书

- 3) 进行舌色描述, 见图 8;
- 4) 进行苔色描述, 见图 9;
- 5) 进行舌苔厚度定量分析和描述, 见图 10;
- 6) 采用文字和伪彩色图显示分析结果。为了使分析结果易于理解和临床应用, 采用文字和伪彩色图显示分析结果。将像素识别结果表示成舌色、苔色分布伪彩色图。将逐像素的厚度分类结果表示成舌苔厚度伪彩色分布图。采用文字对舌色、苔色舌苔厚度进行定性描述, 并给出舌苔厚度指数值。图 11 是显示分析结果的一个例子。

另外舌色、苔色、舌苔厚度的描述特征还在于, 像素识别完成后, 统计各分区中各种类型的像素数后, 令 N_{ij} 代表 i 区中第 j 类象其中 $i=1, \dots, 5$ 分别表示舌根、舌中、舌尖、舌左侧、舌右侧等 5 个区域, $j=1, \dots, 15$ 表示舌质淡、舌质淡红、舌质红、舌质暗红、舌质绛红、舌质暗紫、薄白苔、白苔、白厚苔、薄黄苔、黄苔、黄厚苔、灰苔、褐苔、黑苔等 15 种舌质和舌苔类型, 并在此基础上, 根据中医专家的舌诊习惯, 进行舌色、苔色、舌苔厚度的描述。

其中舌色的描述方法为根据舌侧区与舌尖区像素的舌质类型描述舌质的颜色。由于舌质主要分布在舌侧与舌尖, 因而只在舌尖、舌左侧、舌右侧等 3 个区域考虑舌质类型。计算舌尖、舌左侧、舌右侧等 3 个区域中属于各舌质类型 (即属于舌质淡、舌质淡红、舌质红、舌质暗红、舌质绛红、舌质暗紫等类型) 的像素总数, 像素数最多的舌质类型即为该舌图象的舌质特征。根据中医的经验, 舌质是否局部有暗紫和舌尖是否为红舌很重要, 因此对这两种情况单独处理。方法是: 计算舌尖、舌左侧、舌右侧等 3 个区域属于“舌质暗紫”的像素数目与整个舌面像素总数之比, 若这个比值大于某个预设的值, 则认定“局部暗紫”; 计算舌尖区属于“舌质红”的像素数目与整个舌面像素总数之比, 若这个比值大于某个预设的值, 则认定“舌质红”。具体过程为:

- 1) 计算舌面的总面积 (像素的总数) A ;
- 2) 分别计算舌侧与舌尖区 6 种舌质类型的总数

说明书

$$SUM^j = N_3^j + N_4^j + N_5^j, \quad j=1, \dots, 6;$$

3) 令 $SUM^{j_{\max}} = \max(SUM^j, j=1, \dots, 6)$, 则 j_{\max} 代表的舌质类型为该舌图象的舌质特征;

4) 计算暗紫舌质的面积比 $R_6 = \frac{SUM^6}{A}$, 若 $R_6 > \theta_p$ 且 $j_{\max} \neq 6$, 则增加描述“局部暗紫”, 其中 θ_p 为根据实验设定的阈值;

5) 计算舌尖区“舌质红”的面积比 $R_3 = \frac{N_3^3}{A}$, 若 $R_3 > \theta_r$ 且 $j_{\max} \neq 3$, 则增加描述“舌尖红”, 其中 θ_r 为根据实验设定的阈值。

其中苔色的描述, 由于舌侧很少有舌苔分布, 因而对舌苔的描述分为舌根、舌中、舌尖 3 个区域。计算舌根、舌中、舌尖等 3 个区域中属于各舌苔类型 (即属于薄白苔、白苔、白厚苔、薄黄苔、黄苔、黄厚苔、灰苔、褐苔、黑苔等类型) 的像素总数, 在三个区域中, 分别判断像素数最多的舌苔类型的像素数与整个舌面像素总数之比是否小于某一阈值, 如果小于, 则认定该区少苔, 否则该舌苔类型即为该区的舌苔特征。本发明中这三个阈值由试验获得。根据中医的经验, 舌根、舌中区往往呈现多种舌苔, 因此也对舌根、舌中区像素数次多的舌苔类型加以判断。方法是: 计算舌根、舌中区像素数次多的舌苔类型的像素数目与像素数最多的舌苔类型的像素数目之比, 若这个比值大于某个预设的值, 则该区增加描述第二主要类型舌苔。具体方法如下:

1) 计算舌面的总面积 (象素的总数) A , 舌根、舌中、舌尖区的面积 $A_i, i=1, \dots, 3$;

2) 对于舌根、舌中、舌尖区, 分别计算 $N_i^{j_{\max}} = \max(N_i^j, j=7, \dots, 15)$, 其中 $i=1, 2, 3$, 则 j_{\max}^i 代表的舌苔类型为该区的舌苔特征;

3) 计算 $i=1, 2, 3$ 区中 j_{\max}^i 舌苔象素所占的面积比例 $R(j_{\max}^i, i) = \frac{N_i^{j_{\max}}}{A_i}$;

4) 分别对舌根、舌中、舌尖区进行描述。对于舌根区 $i=1$, 若 $R(j_{\max}^1, 1) < \theta_1$ 则描述为“舌根 少苔”; 否则描述舌根为 j_{\max}^1 类型的舌苔。对于舌中区 $i=2$, 若 $R(j_{\max}^2, 2) < \theta_2$, 则描述为“舌中 少苔”; 否则描述舌中为 j_{\max}^2 类型的舌苔, 对于舌尖区 $i=3$, 若 $R(j_{\max}^3, 3) < \theta_3$, 则描述为“舌尖 少苔”; 否则描述舌尖为 j_{\max}^3 类

说 明 书

型的舌苔。其中 θ_1 、 θ_2 、 θ_3 为根据实验设定的阈值；

5) 对于舌根、舌中区，分别计算第二主要类型舌苔 $j_{sec}^i, i=1,2$ ，并计算 j_{sec}^i 类型像素面积与 j_{max}^i 类型的面积比 $R_{ms}^i = \frac{N_i^{j_{sec}^i}}{N_i^{j_{max}^i}}$ 。若 $R_{ms}^i > \theta_{sec}$ ，则第 i 区增加描述第二

主要类型舌苔。其中 θ_{sec} 为根据实验设定的阈值。

其中舌苔厚度的分析和描述分为两部分：(1) 舌苔厚度分类。将舌体区域的每一个像素分为“无苔”、“薄苔”、“中厚苔”、“厚苔”4种类型。逐像素舌苔厚度分类的依据是“舌苔的厚薄以见底、不见底为标准”。(2) 舌苔厚度定量化。在逐像素识别的基础上，对整幅舌图象的舌苔厚度进行定量化描述。其中，“见底不见底”以像素邻域内舌质像素的多少判定。如果像素邻域内舌质像素很多，则见底程度大。像素的舌苔类型也反映了舌苔的厚薄。属于薄白苔、薄黄苔反映该像素处为薄苔；属于白苔、黄苔、灰苔、褐苔表明该像素处为中厚苔；属于白厚苔、黄厚苔、黑苔表明该像素处为厚苔。具体步骤为：

- 1) 读取舌体像素及其 5×5 邻域的分类号；
- 2) 设 c_1 、 c_2 为根据实验设定的阈值 $c_1 > c_2$ ，且。计算该像素 5×5 邻域内属于舌质类型的像素数 S_{body} ，若 $S_{body} > c_1$ ，则该点的舌苔厚度类型定为“无苔”；
- 3) 若 $c_1 \geq S_{body} > c_2$ ，则该点的舌苔厚度类型定为“薄苔”，
- 4) 若 $S_{body} \leq c_2$ ，则依据该点的舌质舌苔类型 j 确定厚度类型，若为舌质($j < 7$)，则为“薄苔”，若 $j=7$ (薄白苔)、 10 (薄黄苔)，则为“薄苔”，若 $j=8$ (白苔)、 11 (黄苔)、 13 (灰苔)、 14 (褐苔)，则为“中厚苔”；若 $j=9$ (白厚苔)、 12 (黄厚苔)、 15 (黑苔)，则为“厚苔”；
- 5) 重复(1)-(3)步，直到所有像素处理完毕；
- 6) 根据象素识别的结果，计算整幅图象的苔厚指数 T ，作为舌苔厚度分析的定量化结果，计算方法为：

$$T = \frac{\sum_k (w_{j,k})}{A}$$

说明书

式中 A 表示舌体区的象素总数, \sum_k 表示对舌体区域中所有象素求和, j^k 表示第 k 个象素的厚度类型, 对应于“无苔”、“薄苔”、“中厚苔”、“厚苔”分别取 0、1、2、3, w_{j^k} 为权值, 根据实验和中医专家的经验确定, 苔厚指数能够描述舌苔的总体厚度。

由图 11 可以看出, 本方法能够通过计算机获得的定量分析与定性描述结果, 且结果与中医的目视诊察结果一致, 有助于舌诊的客观化、标准化。

附图说明

图 1 是中医舌象舌色、苔色、苔厚分析系统框图

1、数码相机, 2、USB 接口, 3、计算机处理器, 4、输出缓存, 5、舌象分析, 6、显示器, 7、分析结果;

图 2 是中医舌象分析方法主程序流程图;

图 3 支持向量机原理说明图

(a) 分类间隔较小的分类面, (b) 具有最大分类间隔的最优分类面;

图 4 用于分级聚类分析生成的聚类树示意图

H、舌质暗红, I、舌质淡红, J、舌质淡, K、舌质绛紫, L、舌质暗紫, M、舌质红, N、白苔, O、白厚苔, P、薄白苔, Q、黄苔, R、薄黄苔, S、黄厚苔, T、褐苔, U、灰苔, V、黑苔;

图 5 是图 4 聚类树的生成方法子程序流程图;

图 6 是舌体区域像素识别方法子程序流程图;

图 7 是舌面分区示意图;

图 8 是舌色描述方法子程序流程图;

图 9 是苔色描述方法子程序流程图;

图 10 是舌苔厚度分析与描述方法子程序流程图;

图 11 是舌图象的分析结果

(a) 舌体图像, (b) 舌色、苔色伪彩色图, (c) 舌苔厚度伪彩色图, (d) 文字描述的自动分析结果;

图 12 是在计算机上运行的舌象分析的程序主流程图;

说明书

图 13 是在计算机上运行的舌体区域像素识别子程序流程图。

具体实施方式

在图 1 的中医舌象分析系统框图中，数码相机和 USB 接口都是市售的，主要完成采集舌图象，将舌体及色标的光学信号转换为图象电信号输入到计算机，便于计算机处理、传输等操作；计算机处理主要是通过 USB 接口软件对舌图像进行读/写处理，处理后的舌图像输出到缓存器，便于显示。显示器是图象的输出设备，人眼可以通过显示器观看原始舌图象和分析结果。舌体分析是对计算机读入的舌图象进行舌色、苔色、苔厚的定量分析与定性描述，并输出分析结果。

原始舌图象可以通过数码相机实时采集到的图象，也可以是实现通过数码相机采集到保存在计算机硬盘里的图象。

采用支持向量机进行像素识别，涉及的关键问题包括输入特征空间的选择、学习集的构成、核函数的选取、惩罚因子的选取等。学习集指类别已知的样本的集合。

本系统中，输入特征空间为 RGB 彩色空间，为避免噪声的影响，对每一个像素，取其 3×3 邻域的 RGB 均值作为特征矢量。在若干典型图象中选出一系列图象子块，由中医专家逐块确定类别后构成学习集。样本（图象子块）根据图象的具体情况而定，没有统一的大小。

本发明中 SVM 的惩罚因子 C 体现了对学习集的信任度。C 越大，信任度越高，C 越大，信任度越高。舌体像素识别的研究中，由于分类的类别数较多，且各类之间的可分性较差，需要通过实验选取一个合适的惩罚因子。本发明中，通过实验选取 $C=400$ 。

不同的核函数对应学习机在输入空间中具有不同类型的非线性决策面。我们试验了常用的多项式、径向基函数（RBF）、神经网络等核函数形式。最终应用的核函数为 RBF，其形式为：

$$K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$$

式中 σ 是一个重要的参数，决定着 RBF 核函数的具体形式，通过实验选取 σ

说明书

=50。

本发明中为了使分析结果易于理解和临床应用，要进行舌质与舌苔的定性、定量描述。将舌面划分为舌根、舌中、舌尖、舌左侧、舌右侧等 5 个区域。划分方法是以舌尖至人字形界沟中点划分为 5 等分，前 1/5 称舌尖，中 2/5 称舌中，后 2/5 称舌根。另以舌中线与舌边的中点划一线，线外部分称舌侧。如图 7 所示。统计各分区中各种类型的像素数，根据中医专家的舌诊习惯，进行基于多类支持向量机的中医舌色、苔色、舌苔厚度的计算机分析方法。

舌象分析主要通过软件来实现。在计算机中完成以下程序（主程序见图 12）：

1、读入舌体区域图象数据，读入 CTSVM 的支持向量及相应的系数，初始化参数 θ_p 、 θ_r 、 θ_1 、 θ_2 、 θ_3 、 θ_{sec} 、 c_1 、 c_2 、 w_0 、 w_1 、 w_2 、 w_3 。其中 θ_p 、 θ_r 分别为舌色分析时判断“局部暗紫”和“舌尖红”的阈值，本发明中取 $\theta_p=0.01$ ， $\theta_r=0.06$ 。 θ_1 、 θ_2 、 θ_3 舌根、舌中、舌尖三区的阈值，本发明中取 $\theta_1=0.5$ 、 $\theta_2=0.2$ 、 $\theta_3=0.4$ 。 c_1 、 c_2 为苔厚分析时判断舌苔的有无和舌苔类型时采用的阈值。本发明中取 $c_1=20$ 、 $c_2=8$ 。 w_0 、 w_1 、 w_2 、 w_3 为计算苔厚指数时的权系数，本发明中确定为 $w_0=0$ 、 $w_1=0.2$ 、 $w_2=0.7$ 、 $w_3=1.3$ 。

2、进入像素识别子程序，采用基于聚类树的支持向量机（CTSVM）方法进行舌体区域像素识别。在此设已经通过训练得到图 4 所示的聚类树。树中的每一个节点都是一个支持向量机分类器，支持向量及相应的系数已经通过训练获得。对一个像素，支持向量机的分类过程如下：

1) 计算该像素 3×3 邻域内的 RGB 均值 $x=(R,G,B)$ ；

2) 从根节点开始，计算该像素与该节点对应的每一个支持向量的核函数内积，并乘以相应的系数和标号，对所有这些乘积求和，然后求加上常数 b 之后的结果的符号。即求：

$$f(x) = \text{sign} \left(\sum_{x_i \in SV} \alpha_i y_i K(x_i, x) + b \right)$$

说明书

式中, x_i 是该分类器的支持向量, α_i 是与支持向量对应的系数, y_i 是支持向量样本的标号, b 是常数, 本发明中 b 取 40, $K(x_i, x)$ 是核函数内积,

$$K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$$

式中 $\|\cdot\|$ 是向量的范数, σ 取 50。

$Sign()$ 是符号函数, 如果自变量的值大于 0, 则函数值为 1, 小于 0, 则为 -1。

如果 $f(x)$ 大于 0, 则将它归类到左边的分支, 否则归类到右边的分支;

3) 重复 2) 中的过程, 直至到达某个分支节点。例如某个像素的识别过程为: $X \rightarrow A \rightarrow D \rightarrow K$, 则该像素被判别为“舌质绛红”;

4) 对每一个像素, 执行 2)、3), 直到舌体区域中所有像素都判别完毕, 每个像素都分配了一个类别号。

3、舌面分区。从左至右、从上到下扫描舌体区域, 得到舌体区域的外接矩形。设矩形左右两个边在舌图象中的坐标为 l 、 r , 上下两个边在舌图象中的坐标为 t 、 b , 则根据前述舌面分区的原则 (将舌面划分为舌根、舌中、舌尖、舌左侧、舌右侧等 5 个区域, 划分方法是以舌尖至人字形界沟中点划分为 5 等分, 前 1/5 称舌尖, 中 2/5 称舌中, 后 2/5 称舌根。另以舌中线与舌边的中点划一线, 线外部分称舌侧。), 得到以下的分区结果:

设 (x, y) 表示舌图象中像素的纵、横坐标, 则

舌左侧: 若 $x < \frac{3l+r}{4}$, 且 (x, y) 在舌体区域,

舌右侧: 若 $x \geq \frac{l+3r}{4}$, 且 (x, y) 在舌体区域,

舌根区: 若 $\frac{3l+r}{4} \leq x < \frac{l+3r}{4}, y < \frac{3t+2b}{4}$, 且 (x, y) 在舌体区域,

舌中区: 若 $\frac{3l+r}{4} \leq x < \frac{l+3r}{4}, \frac{3t+2b}{4} \leq y < \frac{t+4b}{4}$,

舌尖区: 若 $\frac{3l+r}{4} \leq x < \frac{l+3r}{4}, y \geq \frac{t+4b}{4}$, 且 (x, y) 在舌体区域。

4、统计各分区中各种类型的象素数 N'_i 。

5、舌色分析与描述。根据舌侧与舌尖的舌质类型描述舌质, 具体方法为:

说明书

1) 计算舌面的总面积 (象素的总数) A ;

2) 分别计算舌侧与舌尖区 6 种舌质类型的总数:

$$SUM^j = N_3^j + N_4^j + N_5^j, \quad j=1, \dots, 6;$$

3) 令 $SUM^{j_{\max}} = \max(SUM^j, j=1, \dots, 6)$, 则 j_{\max} 代表的舌质类型为该舌图象的舌质特征, 例如 $j_{\max} = 2$ 时描述为“舌质淡红”;

4) 计算暗紫舌质的面积比 $R_6 = \frac{SUM^6}{A}$, 若 $R_6 > 0.01$ 且 $j_{\max} \neq 6$, 则增加描述“局部暗紫”;

5) 计算舌尖区“舌质红”的面积比 $R_3 = \frac{N_3^3}{A}$, 若 $R_3 > 0.05$ 且 $j_{\max} \neq 3$, 则增加描述“舌尖红”。

6、苔色分析与描述。对舌苔的描述分为舌根、舌中、舌尖 3 个区域, 具体步骤如下:

1) 计算舌面的总面积 (象素的总数) A , 舌根、舌中、舌尖区的面积 $A_i, i=1, \dots, 3$;

2) 对于舌根、舌中、舌尖区, 分别计算 $N_i^{j'_{\max}} = \max(N_i^j, j=7, \dots, 15)$, 其中 $i=1, 2, 3$, 则 j'_{\max} 代表的舌苔类型为该区的舌苔特征;

3) 计算 $i=1, 2, 3$ 区中 j'_{\max} 舌苔象素所占的面积比例 $R(j'_{\max}, i) = \frac{N_i^{j'_{\max}}}{A_i}$;

4) 对于舌根区 $i=1$, 若 $R(j'_{\max}, i) < 0.5$, 则描述为“舌根 少苔”; 对于舌中区 $i=2$, 若 $R(j'_{\max}, i) < 0.4$, 则描述为“舌中 少苔”; 对于舌尖区 $i=3$, 若 $R(j'_{\max}, i) < 0.2$, 则描述为“舌尖 少苔”;

5) 若 $R(j'_{\max}, 1) \geq 0.5$, 则描述舌根为 j'_{\max} 类型的舌苔。例如 $j'_{\max} = 11$ 时, 描述为“舌根 黄苔”。若 $R(j'_{\max}, 2) \geq 0.4$, 则描述舌中为 j'_{\max} 类型的舌苔, 例如 $j'_{\max} = 8$ 时, 描述为“舌中 白苔”。若 $R(j'_{\max}, 3) \geq 0.2$, 则描述舌尖为 j'_{\max} 类型的舌苔, 例如 $j'_{\max} = 7$ 时, 描述为“舌尖 薄白苔”;

6) 对于舌根、舌中区, 分别计算第二主要类型舌苔 $j'_{\text{sec}}, i=1, 2$, 并计算 j'_{sec} 类型象素面积与 j'_{\max} 类型的面积比 $R'_{\text{ms}} = \frac{N_i^{j'_{\text{sec}}}}{N_i^{j'_{\max}}}$ 。若 $R'_{\text{ms}} > 0.3$, 则第 i 区增加描述第二

主要类型舌苔。

说明书

7、苔厚的定量分析与定性描述。具体步骤为：

- 1) 读取舌体像素及其 5×5 邻域的分类号；
- 2) 设 c_1 、 c_2 为根据实验设定的阈值 $c_1 > c_2$ ，且。计算该像素 5×5 邻域内属于舌质类型的像素数 S_{body} ，若 $S_{body} > c_1$ ，则该点的舌苔厚度类型定为“无苔”；
- 3) 若 $c_1 \geq S_{body} > c_2$ ，则该点的舌苔厚度类型定为“薄苔”；
- 4) 若 $S_{body} \leq c_2$ ，则依据该点的舌质舌苔类型 j 确定厚度类型。若为舌质 ($j < 7$)，则为“薄苔”；若 $j=7$ (薄白苔)、 10 (薄黄苔)，则为“薄苔”；若 $j=8$ (白苔)、 11 (黄苔)、 13 (灰苔)、 14 (褐苔)，则为“中厚苔”；若 $j=9$ (白厚苔)、 12 (黄厚苔)、 15 (黑苔)，则为“厚苔”；
- 5) 重复 1) ~ 4) 步，直到所有像素处理完毕；
- 6) 根据象素识别的结果，计算整幅图象的苔厚指数 T ，作为舌苔厚度分析的定量化结果。计算方法为：

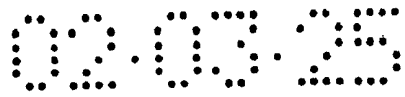
$$T = \frac{\sum_k (w_{j^k})}{A}$$

式中 A 表示舌体区的象素总数， \sum_k 表示对舌体区域中所有象素求和， j^k

表示第 k 个象素的厚度类型，对应于“无苔”、“薄苔”、“中厚苔”、“厚苔”分别取 0、1、2、3， w_{j^k} 为权值，根据实验和中医专家的经验确定。本发明中确定为 $w_0=0$ 、 $w_1=0.2$ 、 $w_2=0.7$ 、 $w_3=1.3$ 。苔厚指数能够描述舌苔的总体厚度。

8、分析结果的显示。为了使分析结果易于理解和临床应用，采用文字和伪彩色图显示分析结果。将象素识别结果表示成舌色、苔色分布伪彩色图。将逐象素的厚度分类结果表示成舌苔厚度伪彩色分布图。采用文字对舌色、苔色舌苔厚度进行定性描述，并给出舌苔厚度指数值。

图 11 是显示分析结果的一个例子。



说明书附图

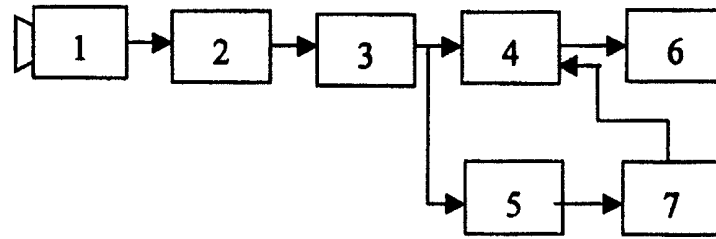


图 1

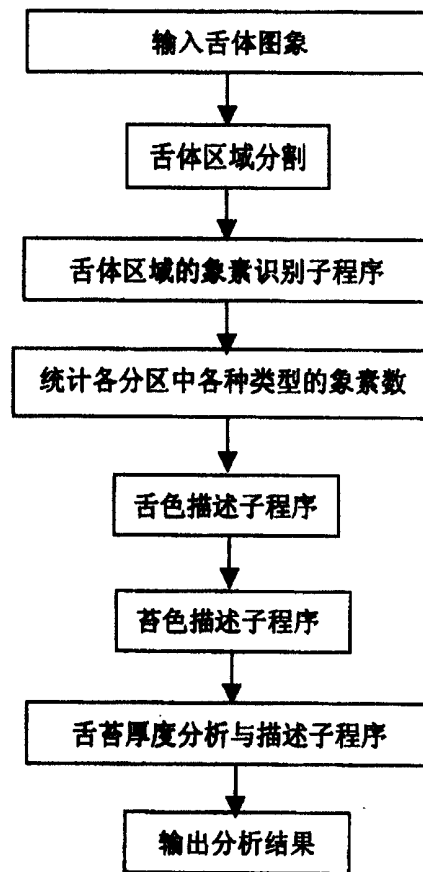


图 2

说明书附图

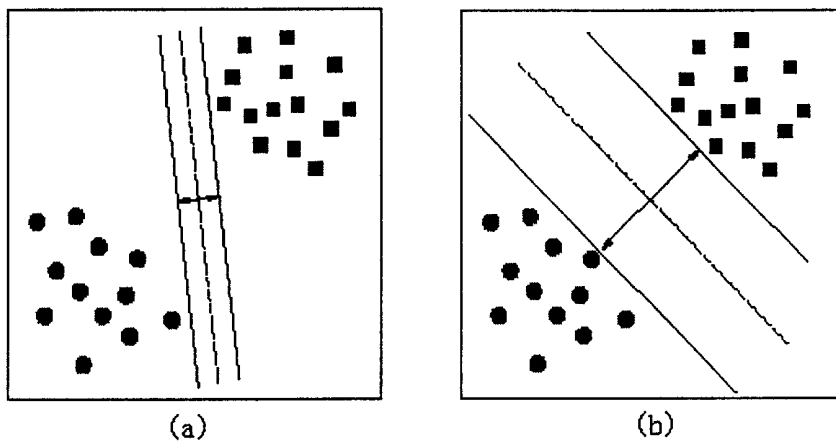


图 3

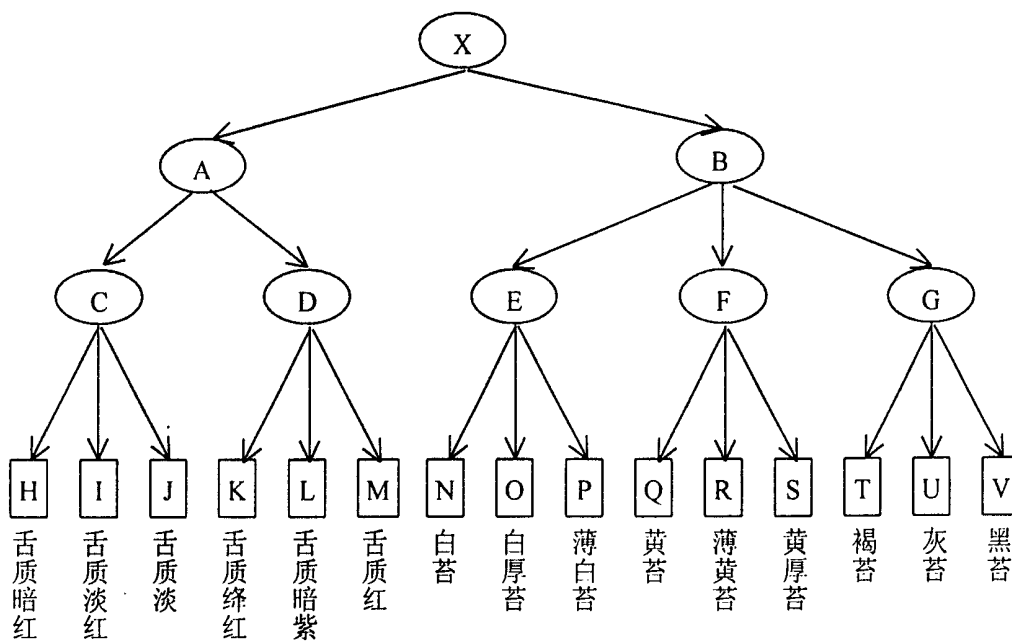


图 4

说明书附图

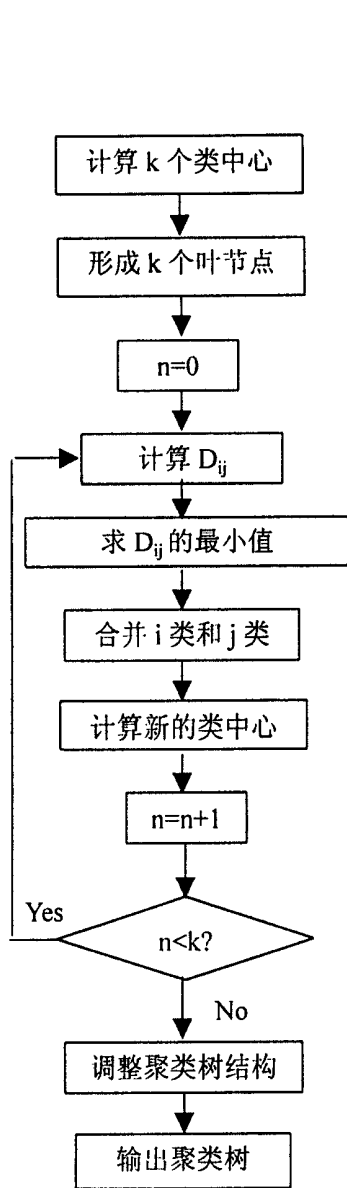


图 5

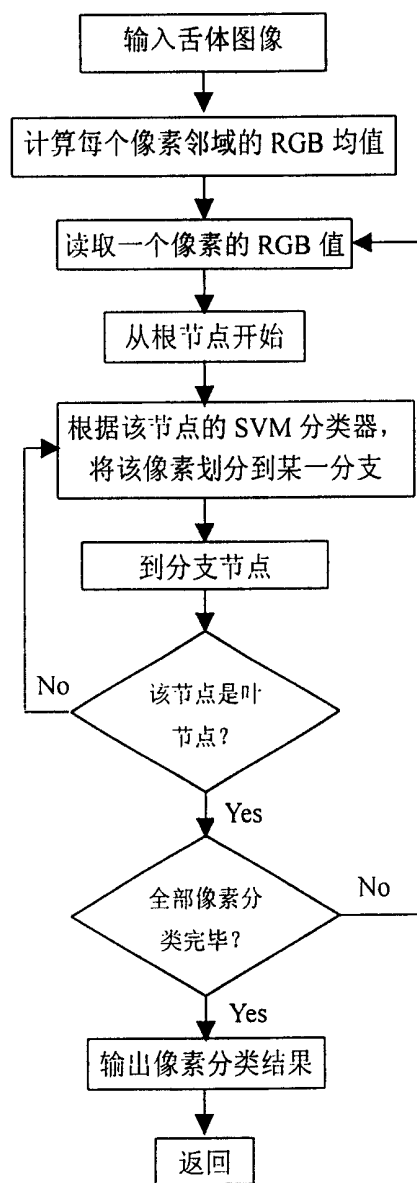


图 6

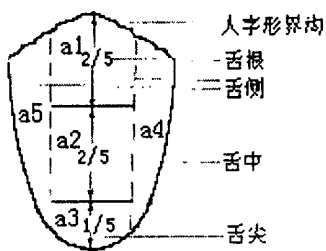


图 7

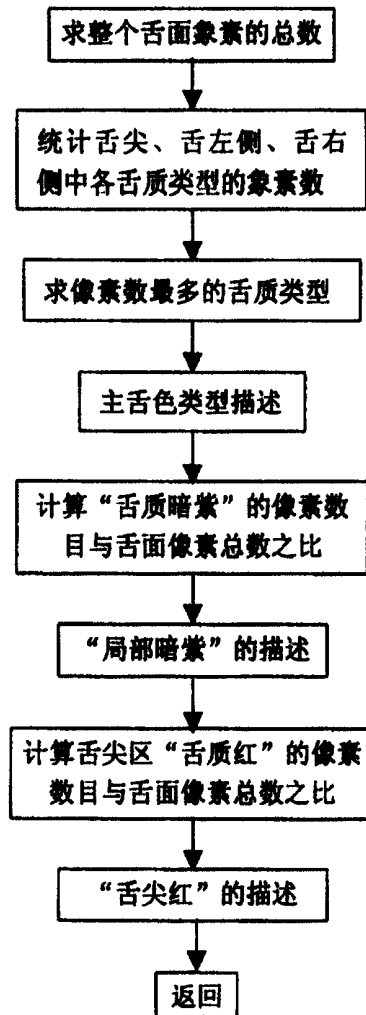


图 8

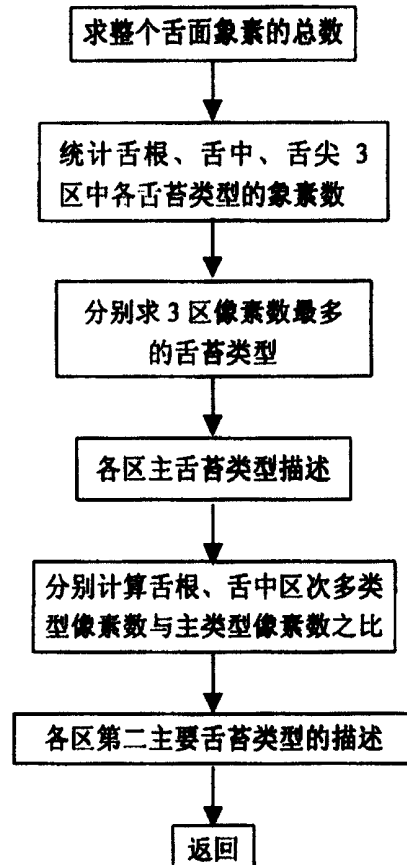


图 9

说明书附图

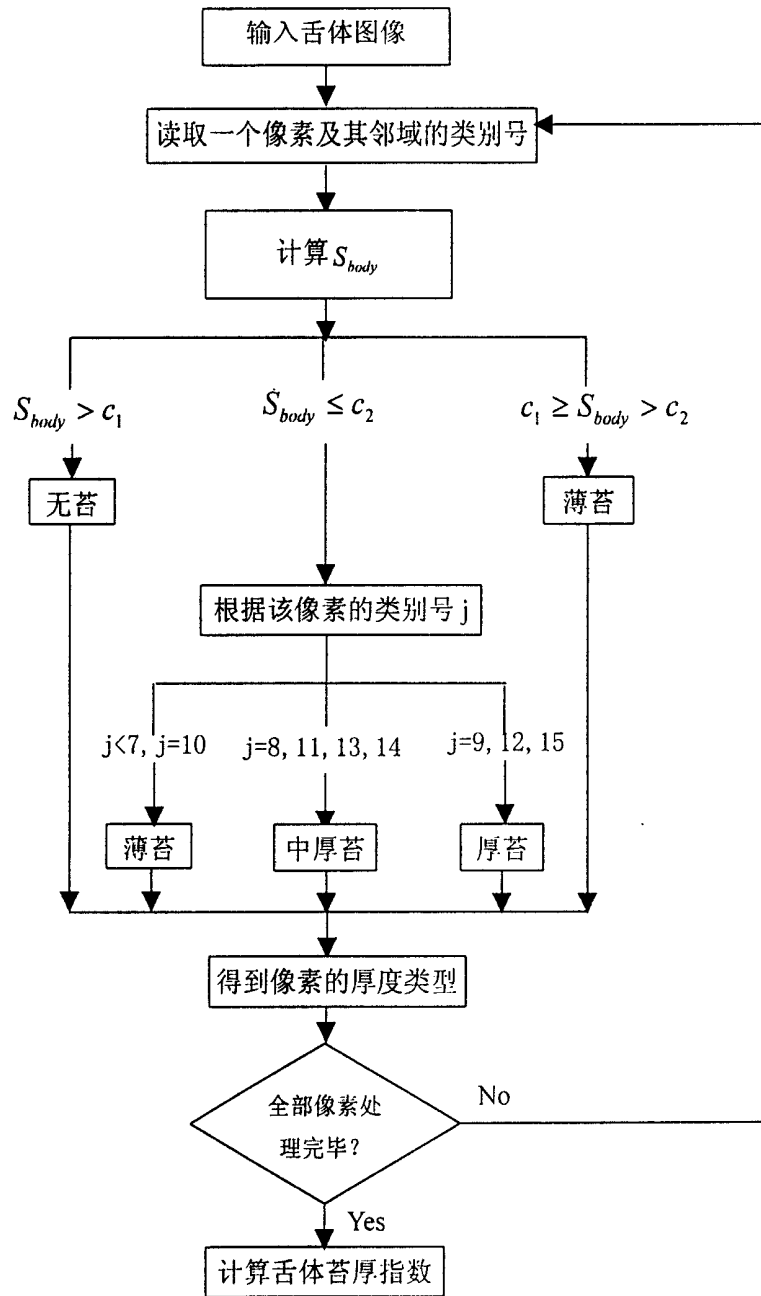
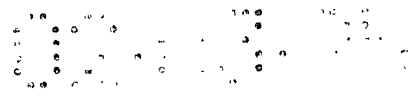
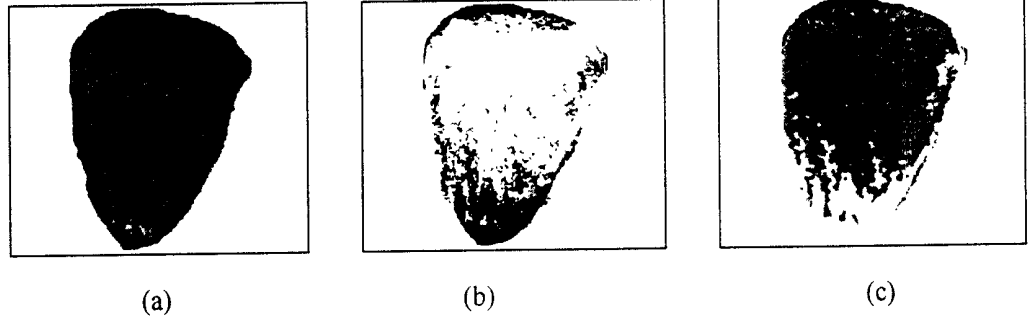


图 10



说明书附图



自动分析结果:

- [舌质] 舌质暗红、尖部红
- [舌苔] 舌根薄黄苔
舌中薄白苔、薄黄苔
舌尖苔少
- [苔厚指数] 0.29

(d)

图 11



说明书附图

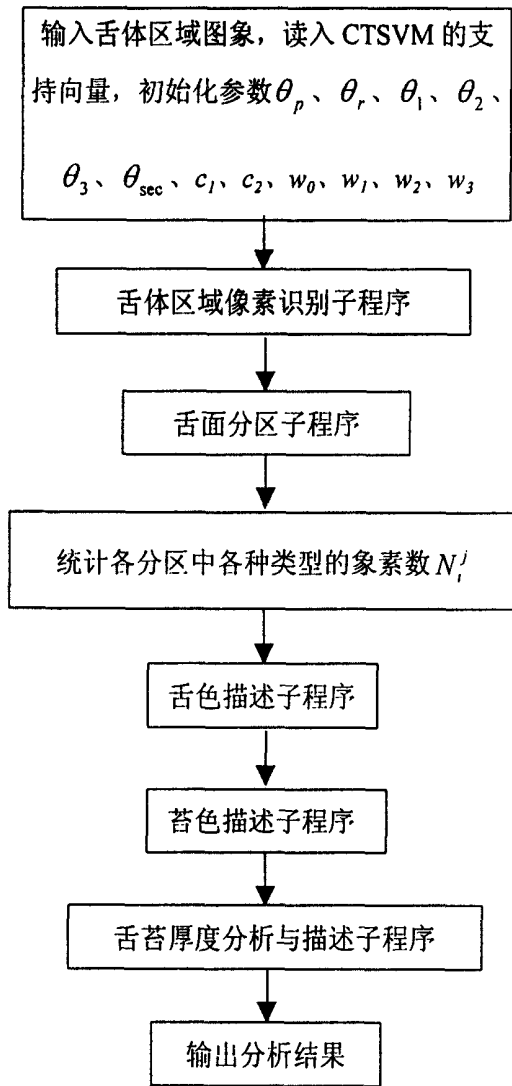


图 12

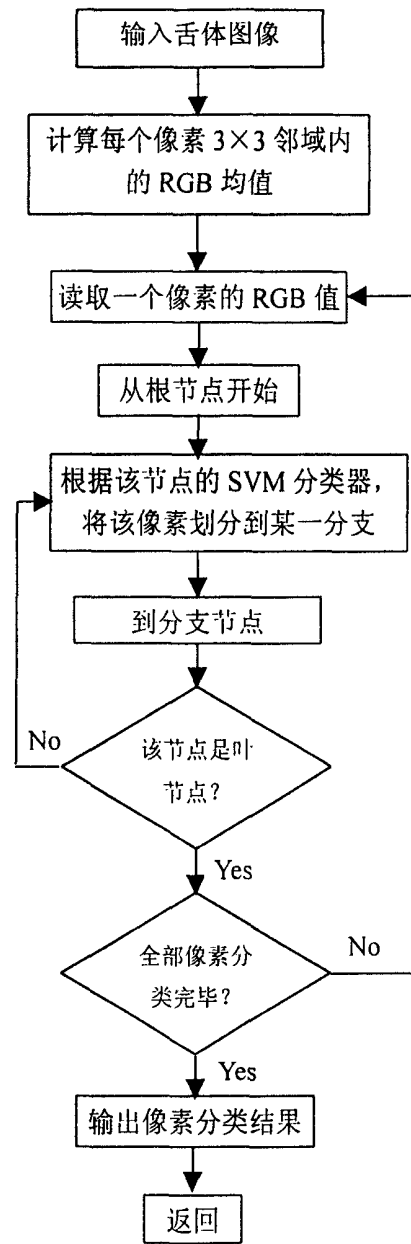


图 13