(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 104392240 A
(43)申请公布日 2015.03.04

(21)申请号 201410587222.8

(22)申请日 2014.10.28

(71)申请人 中国疾病预防控制中心寄生虫病预防控制所
地址 200025 上海市黄浦区瑞金二路207号

(72)发明人 沈海默 陈韶红 陈家旭

(74)专利代理机构 上海世贸专利代理有限责任公司 31128
代理人 严新德

(51)Int.Cl.
G06K 9/62(2006.01)

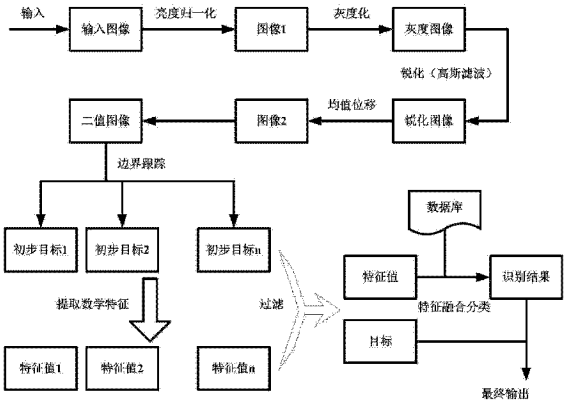
权利要求书6页 说明书17页 附图6页

(54)发明名称

一种基于多特征融合的寄生虫虫卵识别方法

(57)摘要

一种基于多特征融合的寄生虫虫卵的识别方法，包括一个对图像预处理的步骤，将显微照相设备获取的图像信息进行亮度归一化处理、基于高斯滤波的锐化处理，得到虫卵边缘锐化的图像；使用均值移位算法来对目标图片进行分割处理，获得判断为虫卵的区域；依据所建立的要识别寄生虫虫卵形状边缘区域信息，对每个候选的边缘区域进行二值化处理，采用边界跟踪算法按照虫卵区域的边界进行目标获取，得到分割后的虫卵图像；截取虫卵图像的指定特征值，存入预设特征数据库的；采用基于相对距离的KNN（k=3）算法，将所获取的特征值代入总数据库，基于KNN算法判断虫卵类别。本发明对虫卵的识别准确度超过90%，达到较理想的结果。



CN 104392240 A

1. 一种基于多特征融合的寄生虫虫卵的识别方法，包括一个利用显微照相设备获取寄生虫虫卵的图像的过程，其特征在于：所述的过程还包括如下步骤：

a) 一个对图像预处理的步骤，在所述的对图像预处理的步骤中，将显微照相设备获取的图像信息进行亮度归一化处理，对归一化的图像进行灰度化处理，生成归一化灰度图像，然后再对整张图片进行基于高斯滤波的锐化处理，得到虫卵边缘锐化的图像；

b) 一个对虫卵边缘锐化的图像进行均值移位寻找虫卵的步骤，在所述的对虫卵边缘锐化的图像进行均值移位寻找虫卵的步骤中，使用均值移位算法来对目标图片进行分割处理，得到上述图像的颜色特征向量，基于颜色特征向量规划并找到最佳目标区域，获得判断为虫卵的区域；

c) 一个基于上述判别为虫卵的区域对虫卵图像进行目标获取的步骤，在所述的目标获取的步骤中，依据所建立的要识别寄生虫虫卵形状边缘区域信息，对每个候选的边缘区域进行二值化处理，采用边界跟踪算法按照虫卵区域的边界进行目标获取，得到分割后的虫卵图像；

d) 一个对分割后的虫卵图像截取指定特征值，存入预设特征值数据库的步骤；

e) 一个分类识别的步骤，在一个分类识别的步骤中，采用基于相对距离的 KNN(k = 3) 算法，将所获取的特征值代入总数据库，基于 KNN 算法判断虫卵类别。

2. 如权利要求 1 所述的基于多特征融合的寄生虫虫卵的识别方法，其特征在于：在一个对虫卵边缘锐化的图像进行均值移位寻找虫卵的步骤中，使用均值移位算法来对目标进行分割处理，在使用均值移位算法来对目标进行分割处理的过程中，先对原图像进行 X×Y 的划分，得到 X×Y 个交点，并对这些交点进行合并处理，即某两个点对应的颜色值之间的欧氏距离小于某个阈值，所述的阈值为图像亮度最高的 5% 像素与亮度最低的 5% 像素的颜色平均值，则把它们合为一个点，这样得到 m 个点作为初始点集合，m 代表图片上 X×Y 共 n 个像素点的集合，每个像素点可以表示为自变量 $X_i\{i = 1\cdots n\}$，样本点平均值位移 M 的计算方法为：

$$M_{h,U}(x) = \frac{h^2}{d+2}\frac{\overline{\nabla} f_E(x)}{\overline{f}_U(x)}$$

在图片中心选择一个初始点，在以此点为中心的窗口 $S_h(x)$ 内计算平均值位移 $M_{h,U}(x)$，如果该值不小于某个阈值，就把窗口 $S_h(x)$ 平移 $M_{h,U}(x)$，然后重复在新的窗口中计算平均值位移，得到新的中心值，直到 $M_{h,U}(x)$ 小于某个阈值，停止平移，得到一个最大局部密度位置；重复上述步骤，得到 m 个对应最大局部密度位置的点，并对这些点进行合并处理，得到 n 个聚类的中心点，即原图像的主色，针对原图像中的每个像素点，根据欧氏距离判断归到哪个聚类中，用一维直方图表示主色信息，横坐标表示各主色，纵坐标表示各主色包含的像素数的比例，这样就得到该图像的颜色特征向量：

$Q = \{(P_i, W_i) i = 1, \cdots, n\}$，其中 $P_i = (L^*_i, a^*_i, b^*_i)$，$W_i \in (0,1]$，

上述公式中，W 为比例，P_i 为颜色值，颜色值用 LSH 分量表示法来表示，分别记为 L_i、a_i、b_i，颜色特征向量 Q 使用 EMD 算法规划最佳目标区域，EMD 函数的公式一般形式为

$$EMD(P,Q)=\min\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}d(p_i,q_j)f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n}f_{ij}}$$

其中与预期中心点相似度 EMD 最高的区域就是目标区。

3. 如权利要求 1 所述的基于多特征融合的寄生虫虫卵的识别方法，其特征在于：在一个基于上述判别为虫卵的区域对虫卵图像进行目标获取的步骤中，依据所建立的要识别寄生虫虫卵形状边缘区域信息，对每个候选的边缘区域进行二值化处理，采用边界跟踪算法进行目标获取，算法开始时按照从上到下的顺序搜索每个像素，设序列数组为 K，首先从左上方开始搜索第一个目标像素点，设为 k0，则像素 k0 是该区域最左上角的边界像素，也就是搜索的起点，设定搜索方向按逆时针，八邻域方向搜索，k0 设置为跟踪标志，并将 k0 做为序列数组的第一个元素插入，按逆时针方向搜索下一个目标像素，并设为 k，如果找不到，则 k 为孤立像素区域；若 k 等于搜索起始边界像素 k0，则按顺序继续判断其它邻近方向上是否还有未跟踪到的边界像素，若没有，则已回到起始点，算法结束，序列 K 中的边界像素点组成一条封闭区域，将目标区域包围在内。

4. 如权利要求 1 所述的一种基于多特征融合的寄生虫虫卵的识别方法，其特征在于：在对分割后的虫卵图像截取指定特征值、存入预设特征数据库的步骤中，先获取虫卵图像的特征值：

1) 求出边缘区域外接最小正方形区域，计数像素数即可得到长度、宽度，长度是指目标物外接矩形的长度，宽度是指目标物外接矩形的长度；

2) 计数目标区域、目标周边区域的像素点，可得面积和周长，其面积与最小外接正方形之比值即为椭圆度，椭圆度是指目标物面积与外接椭圆的面积之比；

3) 面积是指目标物面积，周长是指目标物周长；

4) 基于目标区域颜色构成信息，获取 RGB 分量；将图片转化为灰度即可获取灰度值的统计直方，其均值为灰度值；将目标转化为 HSV 空间，即可获取 HSL 分量；平均灰度是指灰度化后的目标物的颜色平均值；平均红色分量是指计算机对彩色的表达采用了 RGB 组合的方式，平均红色分量是指 R 部分的平均值；平均绿色分量是指计算机对彩色的表达采用了 RGB 组合的方式，平均绿色分量是指 G 部分的平均值；平均蓝色分量是指计算机对彩色的表达采用了 RGB 组合的方式，平均蓝色分量是指 B 部分的平均值；平均色度是指将 RGB 颜色模型转换成 HSL 颜色模型之后，H 部分的平均值；平均饱和度是指将 RGB 颜色模型转换成 HSL 颜色模型之后，S 部分的平均值；平均亮度是指将 RGB 颜色模型转换成 HSL 颜色模型之后 L 部分的平均值；灰度值的统计直方图是指对灰度值 0 ～ 255 的分布进行分阶段统计得到的向量；灰度标准差是指目标物各个局部颜色的差异；颜色权重是指计算机对彩色的表达采用了 RGB 组合的方式时根据像素点位置自动生成的平均色度与位置坐标的比值，该值仅用于纠错，不参与运算；

5) 获取特征值后，输入预设的文本格式数据库，以表格的形式加载后续的分类识别算法。

5. 如权利要求 1 所述的基于多特征融合的寄生虫虫卵的识别方法，其特征在于：在分

类识别的步骤中，采用基于相对距离的 KNN 算法，所述的 KNN 算法的步骤如下：

特征值数据库的每个样本应该对第 i 维属性值为 X[i]，计算最大值 Max[i]、最小值 Min[i]，再利用公式 X[i] = (X[i]-Mini[i])/(Max[i] - Min[i]) 进行归一化操作，样品各属性归一化后其值域为 [0,1]，然后根据特征值数据库构建数据集 D = {X1, …, XL}，其中 $X_i \in R^n$, i = 1…L；设样本共有 ClassNum 个类；设 C_i 表示第 i 类中的所有样本的集合，且 $C_i \cap C_j = \Phi$ (i, j = 1, …, ClassNum)，样本集也可表示为：$D = C_1 \cup C_2 \cup \cdots \cup C_r$；

设两个虫卵样本间的距离为 Dist，数据集 D 有 m 个属性，其数据集构成为 $R(A_1, A_2, \cdots, A_m)$，X 和 Y 分别为数据集 D 中的两个样本，则 X 与 Y 的距离度量公式为：

$$Dist(X,Y)=\sqrt{\sum_{i=1}^{m}(X.x_i-Y.y_i)^2}$$

测试样本中第 i 类的 K- 最近邻距离均值为：

$$Avgdis(i)=\frac{\sum_{j=1}^{k_i} Dist(X_j,Y)}{k_i} \quad X_j \in C_i \quad i=1,\cdots,ClassNum$$

K_i 为 C_i 中的样本个数，Y 为 Xj 的最近邻，测试样本 X 和训练样本 Y 之间的相对距离即为：D = Dist(X,Y)/Avgdis(i), $Y \in C_i$；

在 N = 3 时，只要计算数据集各样本到测算样本的距离，比较选取测试样本的 3 个最近邻，即可判别它的类别，分类结果由 score 来体现，设输入图片的特征为 (f1,f2,...,fn)，数据库中某样本的特征是 (x11,x12,...,x1n)，则 score = s(f1,x11)*s(f2,x12)*.....*s(fn,x1n)；

此处 s 函数构造为：令 maxV = max(f1,x11)；minV = min(f1,x11)，则 diff = (maxV-minV)/maxV；s = X*(pow(e,-diff)-1/e)+B，其中 e 是自然数，X = (A-B)*e/(e-1)，即可令 A 取到最大值，如果 diff = 0，则 s 是最大值 A；如果 diff = 1，则 s 是最小值 B。

6. 如权利要求 1 所述的基于多特征融合的寄生虫虫卵的识别方法，其特征在于：在使用均值移位算法对目标进行分割处理之前，图片需要预先计算彩色直方图。

7. 如权利要求 1 所述的基于多特征融合的寄生虫虫卵的识别方法，其特征在于：如因过度归一化等问题造成图像出现断点，导致遗失边界点无法构建区域框或区域像素集合，则用区域生长算法加以弥补，所述的区域生长算法的具体步骤是：先对每个需要分割的区域找一个种子像素作为生长的起点，然后将种子像素周围邻域中与种子像素具有相同或相似性质的像素合并到这一区域中，将这些新像素当作新的种子像素继续进行上面的过程，直到再没有满足条件的像素可被包括进来。

8. 如权利要求 1 所述的基于多特征融合的寄生虫虫卵的识别方法，其特征在于：所述的寄生虫为华支睾吸虫、带绦虫、鞭虫、蛲虫、蛔虫、钩虫、阔节裂头绦虫、日本血吸虫、布氏姜片吸虫、肺吸虫、或者曼氏迭宫绦虫。

9. 如权利要求 1 所述的基于多特征融合的寄生虫虫卵的识别方法，其特征在于：所述的待识别图具有旋转不变性、对光照无差异、对个体无差异。

10. 一种基于多特征融合的寄生虫虫卵的识别方法，其特征在于包括如下步骤：

a) 一个对图像预处理的步骤，在一个对图像预处理的步骤中，将显微照相设备获取的

图像信息进行亮度归一化处理，对归一化的图像进行灰度化处理，生成归一化灰度图像，然后再对整张图片进行基于高斯滤波的锐化处理，得到虫卵边缘锐化的图像；

b) 一个采用人工辅助识别寻找虫卵的步骤，在一个采用人工辅助识别寻找虫卵的步骤中，采用增强 Grab Cut 法对虫卵边缘锐化的图像进行分割，用户提供限定方框进行人工支持，得到虫卵图像的颜色特征向量，基于颜色特征向量规划并找到最佳目标区域，得到判断为虫卵的区域；

c) 一个基于上述判别为虫卵的区域对虫卵图像进行目标获取的步骤，在进行目标获取的步骤中，依据所建立的要识别寄生虫虫卵形状边缘区域信息，对每个候选的边缘区域进行二值化处理，采用边界跟踪算法按照虫卵区域的边界进行目标获取，得到分割后的虫卵图像；

d) 一个对分割后的虫卵图像截取指定特征值并存入预设特征数据库的步骤；

e) 一个分类识别的步骤，在一个分类识别的步骤中，采用基于相对距离的 KNN(k = 3) 算法，将所获取的特征值代入总数据库，基于 KNN 算法判断虫卵类别。

11. 如权利要求 10 所述的一种基于多特征融合的寄生虫虫卵的识别方法，其特征在于：

a) 在一个采用人工辅助识别寻找虫卵的步骤中，采用增强 Grab Cut 法对虫卵边缘锐化的图像进行分割，用户提供限定方框进行人工支持，方框以外的部分不处理，用户通过设置背景区域 TB 来初始化三分图 T，前景区域 TF 设置为空，未知区域 TU 设置为背景区域 TB 的补集，对于所有背景区域的像素，将它们的 Alpha 值设置为 0，即 a = 0；对于未知区域的像素点，将它们的 Alpha 值设置为 1，即 a = 1，分别用 a = 0 和 a = 1 这两个集合来初始化创建前景与背景的高斯混合模型，为未知区域中的每个像素点 n 设置高斯混合模型参数：

$kn = \arg\min Dn(a_n, k_n, \theta, Z_n)$，

由图像中各个像素的数据求得高斯混和模型参数

$\theta = \arg\min U(a, k_n, \theta, Z_n)$，

利用最小化能量公式来得到初始分割：

$\min_k E(a, k_n, \theta, Z_n)$，

重复执行 3 次，进行边界优化。

12. 如权利要求 10 所述的一种基于多特征融合的寄生虫虫卵的识别方法，其特征在于：基于上述判别为虫卵的区域对虫卵图像进行目标获取的步骤中，依据所建立的要识别寄生虫虫卵形状边缘区域信息，对每个候选的边缘区域进行二值化处理，采用边界跟踪算法进行目标获取，算法开始时按照从上到下的顺序搜索每个像素，设序列数组为 K，首先从左上方开始搜索第一个目标像素点，设为 k0，则像素 k0 是该区域最左上角的边界像素，也就是搜索的起点，设定搜索方向按逆时针，八邻域方向搜索，k0 设置为跟踪标志，并将 k0 做为序列数组的第一个元素插入，按逆时针方向搜索下一个目标像素，并设为 k，如果找不到，则 k 为孤立像素区域；若 k 等于搜索起始边界像素 k0，则按顺序继续判断其它邻近方向上是否还有未跟踪到的边界像素，若没有，则已回到起始点，算法结束，序列 K 中的边界像素点组成一条封闭区域，将目标区域包围在内。

13. 如权利要求 10 所述的一种基于多特征融合的寄生虫虫卵的识别方法，其特征在于：在一个对获取目标截取指定特征值存入特征数据库的步骤中，先获取虫卵图像的特征

值：

1）求出边缘区域外接最小正方形区域，计数像素数即可得到长度、宽度，长度是指目标物外接矩形的长度，宽度是指目标物外接矩形的长度；

2）计数目标区域、目标周边区域的像素点，可得面积和周长，其面积与最小外接正方形之比值即为椭圆度，椭圆度是指目标物面积与外接椭圆的面积之比；

3）面积是指目标物面积，周长是指目标物周长；

4）基于目标区域颜色构成信息，获取 RGB 分量；将图片转化为灰度即可获取灰度值的统计直方，其均值为灰度值；将目标转化为 HSV 空间，即可获取 HSL 分量；平均灰度是指灰度化后的目标物的颜色平均值；平均红色分量是指计算机对彩色的表达采用了 RGB 组合的方式，平均红色分量是指 R 部分的平均值；平均绿色分量是指计算机对彩色的表达采用了 RGB 组合的方式，平均绿色分量是指 G 部分的平均值；平均蓝色分量是指计算机对彩色的表达采用了 RGB 组合的方式，平均蓝色分量是指 B 部分的平均值；平均色度是指将 RGB 颜色模型转换成 HSL 颜色模型之后，H 部分的平均值；平均饱和度是指将 RGB 颜色模型转换成 HSL 颜色模型之后，S 部分的平均值；平均亮度是指将 RGB 颜色模型转换成 HSL 颜色模型之后 L 部分的平均值；灰度值的统计直方图是指对灰度值 0 ～ 255 的分布进行分阶段统计得到的向量；灰度标准差是指目标物各个局部颜色的差异；颜色权重是指计算机对彩色的表达采用了 RGB 组合的方式时根据像素点位置自动生成的平均色度与位置坐标的比值，该值仅用于纠错，不参与运算；

5）获取特征值后，输入预设的文本格式数据库，以表格的形式加载后续的分类识别算法。

14. 如权利要求 10 所述的一种基于多特征融合的寄生虫虫卵的识别方法，其特征在于：在一个分类识别的步骤中，采用基于相对距离的 KNN 算法，KNN 算法的步骤如下：

首先为避免由于属性值域不同而影响样本距离的计算，特征值数据库的每个样本应该对第 i 维属性值为 X[i]，计算最大值 Max[i]、最小值 Min[i]，再利用公式 X[i] = (X[i]-Mini[i])/(Max[i] - Min[i]) 进行归一化操作，样品各属性归一化后其值域为 [0,1]，然后根据特征值数据库构建数据集 D = {X1, …, XL}，其中 $X_i \in R^n$，i = 1…L；设样本共有 ClassNum 个类；设 C_i 表示第 i 类中的所有样本的集合，且 $Ci \cap C_j = \Phi$(i, j = 1, …, ClassNum)，样本集也可表示为：$D = C_1 \cup C_2 \cup \cdots \cup C_r$；

设两个虫卵样本间的距离为 Dist，数据集 D 有 m 个属性，其数据集构成为 $R(A_1, A_2, \cdots, A_m)$，X 和 Y 分别为数据集 D 中的两个样本，则 X 与 Y 的距离度量公式为：

$$Dist(X,Y)=\sqrt{\sum_{i=1}^{m}(X.x_i-Y.y_i)^2}$$

测试样本中第 i 类的 K- 最近邻距离均值为：

$$Avgdis(i)=\frac{\sum_{j=1}^{k_i}Dist(X_j,Y)}{k_i} \quad X_j \in C_i \quad i=1,\cdots,ClassNum$$

K_i 为 C_i 中的样本个数，Y 为 Xj 的最近邻，测试样本 X 和训练样本 Y 之间的相对距离即为：D = Dist(X,Y)/Avgdis(i)，$Y \in C_i$；

在N＝3时，只要计算数据集各样本到测算样本的距离，比较选取测试样本的3个最近邻，即可判别它的类别，分类结果由score来体现，设输入图片的特征为(f1,f2,...,fn)，数据库中某样本的特征是(x11,x12,...,x1n)，则score＝s(f1,x11)*s(f2,x12)*.....*s(fn,x1n)；

此处s函数构造为：令maxV＝max(f1,x11);minV＝min(f1,x11)，则diff＝(maxV-minV)/maxV;s＝X*(pow(e,-diff)-1/e)+B,其中e是自然数,X＝(A-B)*e/(e-1)，即可令A取到最大值，如果diff＝0,则s是最大值A;如果diff＝1,则s是最小值B。

15. 如权利要求10所述的一种基于多特征融合的寄生虫虫卵的识别方法，其特征在于：如因过度归一化等问题造成图像出现断点，导致遗失边界点，则用区域生长算法加以弥补，所述的区域生长算法的具体步骤是：先对每个需要分割的区域找一个种子像素作为生长的起点，然后将种子像素周围邻域中与种子像素具有相同或相似性质的像素合并到这一区域中，将这些新像素当作新的种子像素继续进行上面的过程，直到再没有满足条件的像素可被包括进来。

16. 如权利要求10所述的一种基于多特征融合的寄生虫虫卵的识别方法，其特征在于：所述的寄生虫为华支睾吸虫、带绦虫、鞭虫、蛲虫、蛔虫、钩虫、阔节裂头绦虫、日本血吸虫、布氏姜片吸虫、肺吸虫、或者曼氏迭宫绦虫。

17. 如权利要求10所述的一种基于多特征融合的寄生虫虫卵的识别方法，其特征在于：所述的待识别图具有有旋转不变性、对光照无差异、对个体无差异。

一种基于多特征融合的寄生虫虫卵识别方法

技术领域：

[0001] 本发明属于图像识别技术领域，尤其涉及一种虫卵识别方法，具体来说是一种基于多特征融合的寄生虫虫卵的识别方法。

背景技术：

[0002] 寄生虫病仍然是全球性的公共卫生问题之一，虫卵镜检是关键的防治技术之一，也是寄生虫形态特征分析和后续生物学研究的一个基础环节。寄生虫虫卵的识别无法象血细胞分析一样用自动化仪器进行，长期以来只能依赖人眼在显微镜下进行观察分辨。但在众多的寄生虫样本中对不同的虫卵进行鉴别是一项既繁琐的工作，同时还需要对技术人员进行专门的培训。目前采用标本人工涂片后在显微镜下肉眼辨别的方法，不仅操作繁琐、识别误差随检验人员的经验和状态而异，而且缺乏客观性和精确性，检验标本图像、数据和结果不便于存储、重现和检索，不能适应现代医疗信息化发展的需求。因此，有必要借助计算机技术来协助进行寄生虫虫卵的识别。

[0003] 1995 年大连理工大学电子系的孔祥维开展了显微镜下蠕虫卵微机检测与识别系统的研究，正确识别率接近 92.3%。1997 年赵亚娥也开展了针对 10 种寄生虫虫卵图像的自动识别研究，提取了虫卵区域的周长、面积、圆形度和密度四个特征进行了识别，识别正确率达到了 92%。中山大学傅承彬等 2002 年开发出对 7 种吸虫成虫标本并提取相应的 13 个形态学特征进行识别分类，识别准确率达 89.04%。但图像需要用利用 Photoshop、AutoCAD、等进行预处理。2004 年郭晓敏利用小波分类提取了虫卵图像的小波变化系数特征，并选用了概率神经网络来对虫卵进行了分类。2005 年李俊峰利用树型分层原理结合最小距离分类原则、Bayes 判别准则和人工神经网络等构建分类器进行识别，正确率达到 94.91%。2005 年湖南大学的彭社欣开发寄生虫识别系统，识别率可达到 93.0%。2007 年罗泽举、宋丽红等人提出一种新型图像特征提取方法并且采用 SVM 对血吸虫等九种寄生虫虫卵图片实现自动识别和分类，识别率达到 93.9%。

[0004] 在国外，1996 年丹麦哥本哈根兽医实验室 Sommer C. 利用计算其傅里叶变换的振幅进行分类，准确识别率为 81.5% ；Sommer C. 于 1998 年提取了三种牛线虫虫卵图像的大小、纹理和形状特征用于分类识别，使得平均正确识别率达到 91.2%；1999 年韩国首尔国立大学 Yang ySl131 等人采用 7 种共 52 张人体寄生虫虫卵图像，并利用人工神经网络识别方法对提取的 4 种形态学特征进行分类检测和识别，识别准确率达到 86 %。Yang 等于 2001 年增加了虫卵的种类和图片数量，并利用上述方法进行分类检测和识别后得到的正确识别率提高到 90.3%。2000 年希腊雅典国家科技大学的 G. Theodoropoulos[141 等人对寄生于家畜中的五种线虫幼虫图像进行数字图像识别处理，提取的 7 个有效特征参数进行分类，正确识别率为 91.9%。2007 年巴西圣保罗大学的 Jane S. Fraga 等人利用 Bayes 分类器实现了对家禽感染寄生虫的识别，识别率达到 85.75%。同年苏丹的 S. Raviraja 运用统计学的方法来分类感染疟疾病原体的血液图片的分类。

[0005] 虽然国内外有关研究人员都在尝试利用计算机进行寄生虫病原体的自动识别研

究，但利用计算机对寄生虫卵图像进行自动识别仍有不少困难，主要体现在以下几方面：

[0006] a) 寄生虫的种类多，使得在图像预处理时很难找到能适合所有虫卵的方法，寄生虫卵的形态颜色各异，使得选取区分各种虫卵的特征很困难；

[0007] b) 由于图像拍摄装置的差异、拍摄环境的不同，即使是同一种虫卵，拍摄出来的图像在背景和虫卵本身的颜色等方面也可能存在差异，这也会影响识别效果；

[0008] c) 寄生虫卵本身在不同的时期也会有不同的形态，有的甚至相差很大，如蛔虫卵在未受精未脱蛋白膜时期和已受精已脱蛋白膜时期就明显不同。

[0009] 从现有研究资料看，我国外开展寄生虫虫卵数字图像自动识别研究不到十五年，远未达到临床应用或自动化仪器识别的程度，带有浓厚的“纯研究”色彩，问题主要表现在以下三个方面：

[0010] a) 能够识别的种类较少，往往局限于某几种、某类虫卵或成虫的实验室研究，适应面太窄的系统在临床上应用价值不高。

[0011] b) 识别系统处理过程中需要人工干预的部分多，如有的识别系统需要先用通用软件测量出特征参数并录入数据库后再调用出来进行分类识别，并非一体化的识别系统；有的系统需要用鼠标选定目标或确定边界跟踪分割的起始点，离声称的“自动”识别距离甚远。

[0012] c) 提取的图像特征不能准确的反映图像特点，使各种识别对象的特征值范围重叠较多，不得不采用复杂度很大的分类算法来提高识别率，与目前医院使用的五分类血细胞分析仪的高效运行相比，国外最快识别处理时间为 15s 的系统还是有待于提高。

[0013] 所以，开发研究一种能适应常见人体寄生虫虫卵识别分类、各种识别处理步骤一体化及处理速度较快的虫卵图像自动识别系统是很有必要的，可以适应未来自动化仪器中的临床应用。

发明内容：

[0014] 本发明提供了一种基于多特征融合的寄生虫虫卵的识别方法，所述的这种基于多特征融合的寄生虫虫卵的识别方法要解决现有技术中的寄生虫虫卵的识别率低，识别种类少，鉴别时间长的技术问题。

[0015] 本发明一种基于多特征融合的寄生虫虫卵的识别方法，包括如下步骤：

[0016] a) 一个对图像预处理的步骤，在所述的对图像预处理的步骤中，将显微照相设备获取的图像信息进行亮度归一化处理，对归一化的图像进行灰度化处理，生成归一化灰度图像，然后再对整张图片进行基于高斯滤波的锐化处理，得到虫卵边缘锐化的图像；

[0017] b) 一个对虫卵边缘锐化的图像进行均值移位寻找虫卵的步骤，在一个对虫卵边缘锐化的图像进行均值移位寻找虫卵的步骤中，使用均值移位算法来对目标图片进行分割处理，得到上述图像的颜色特征向量，基于颜色特征向量规划并找到最佳目标区域，获得判断为虫卵的区域；

[0018] c) 一个基于上述判别为虫卵的区域对虫卵图像进行目标获取的步骤，在所述的目标获取的步骤中，依据所建立的要识别寄生虫虫卵形状边缘区域信息，对每个候选的边缘区域进行二值化处理，采用边界跟踪算法按照虫卵区域的边界进行目标获取，得到分割后的虫卵图像；

[0019] d) 一个对分割后的虫卵图像截取指定特征值，存入预设特征数据库的步骤；

[0020] e) 一个基于多种算法的分类识别的步骤，在一个基于多种算法的分类识别的步骤中，采用基于相对距离的KNN(k = 3) 算法，将所获取的特征值代入总数据库，基于KNN算法判断虫卵类别。

[0021] 进一步的，在一个对虫卵边缘锐化的图像进行均值移位寻找虫卵的步骤中，使用均值移位算法来对目标进行分割处理，在使用均值移位算法来对目标进行分割处理的过程中，先对原图像进行 X×Y 的划分，得到 X×Y 个交点，并对这些交点进行合并处理，即某两个点对应的颜色值之间的欧氏距离小于某个阈值，所述的阈值为图像亮度最高的 5％像素与亮度最低的 5％像素的颜色平均值，则把它们合为一个点，这样得到 m 个点作为初始点集合，m 代表图片上 X×Y 共 n 个像素点的集合，每个像素点可以表示为自变量 $X_i\{i = 1\cdots n\}$，样本点平均值位移 M 的计算方法为：

[0022] $$M_{h,U}(x)=\frac{h^2}{d+2}\frac{\overline{\nabla} f_E(x)}{\overline{f}_U(x)}$$

[0023] 在图片中心选择一个初始点，在以此点为中心的窗口 $S_h(x)$ 内计算平均值位移 $M_{h,U}(x)$，如果该值不小于某个阈值，就把窗口 $S_h(x)$ 平移 $M_{h,U}(x)$，然后重复在新的 窗口中计算平均值位移，得到新的中心值，直到 $M_{h,U}(x)$ 小于某个阈值，停止平移，得到一个最大局部密度位置；重复上述步骤，得到 m 个对应最大局部密度位置的点，并对这些点进行合并处理，得到 n 个聚类的中心点，即原图像的主色，针对原图像中的每个像素点，根据欧氏距离判断归到哪个聚类中，用一维直方图表示主色信息，横坐标表示各主色，纵坐标表示各主色包含的像素数的比例，这样就得到该图像的颜色特征向量：

[0024] $Q = \{(P_i, W_i)\, i = 1, \cdots, n\}$，其中 $P_i = (L^*_i, a^*_i, b^*_i)$，$W_i \in (0, 1]$，上述公式中，W 为比例，$P_i$ 为颜色值，与传统RGB分量表示法不同，此处颜色值用LSH分量表示法来表示，分别记为 L_i、a_i、b_i。基于颜色特征向量 Q 使用传统EMD算法即可规划最佳目标区域，EMD函数的公式一般形式为

[0025] $$EMD(P,Q)=\min\frac{\sum_{i=1}^{m}\sum_{j=1}^{n} d(p_i,q_j) f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}}$$

[0026] 其中与预期中心点相似度EMD最高的区域就是目标区；

[0027] 进一步的，在一个基于上述判别为虫卵的区域对虫卵图像进行目标获取的步骤中，基于上述EMD函数计算得到的目标区，即依据所建立的要识别寄生虫虫卵形状边缘区域信息，对每个候选的边缘区域进行二值化处理，采用边界跟踪算法进行目标获取，算法开始时按照从上到下的顺序搜索每个像素，设序列数组为 K，首先从左上方开始搜索第一个目标像素点，设为 k0，则像素 k0 是该区域最左上角的边界像素，也就是搜索的起点，设定搜索方向按逆时针，八邻域方向搜索，k0 设置为跟踪标志，并将 k0 做为序列数组的第一个元素插入，按逆时针方向搜索下一个目标像素，并设为 k，如果找不到，则 k 为孤立像素区域；若 k 等于搜索起始边界像素 k0，则按顺序继续判断其它邻近方向上是否还有未跟踪到的边界像

素，若没有，则已回到起始点，算法结束，序列 K 中的边界像素点组成一条封闭区域，将目标区域包围在内。

[0028] 进一步的，在一个对分割后的虫卵图像截取指定特征值，存入预设特征数据库的步骤中，先获取虫卵图像的特征值：

[0029] 1) 求出边缘区域外接最小正方形区域，计数像素数即可得到长度 (length)、宽度 (width)，长度是指目标物外接矩形的长度，宽度是指目标物外接矩形的长度；

[0030] 2) 计数目标区域、目标周边区域的像素点，可得面积和周长，其面积与最小外接正方形之比值即为椭圆度 (ovality)，椭圆度是指目标物面积与外接椭圆的面积之比；

[0031] 3) 面积 (area) 是指目标物面积，周长 (perimeter) 是指目标物周长；

[0032] 4) 基于目标区域颜色构成信息，获取 RGB 分量；将图片转化为灰度即可获取灰度值的统计直方，其均值为灰度值；将目标转化为 HSV 空间，即可获取 HSL 分量；平均灰度 (grey) 是指灰度化后的目标物的颜色平均值；平均红色分量是指计算机对彩色的表达采用了 RGB 组合的方式，平均红色分量 (red) 是指 R 部分的平均值；平均绿色分量 (green) 是指计算机对彩色的表达采用了 RGB 组合的方式，平均绿色分量是指 G 部分的平均值；平均蓝色分量 (blue) 是指计算机对彩色的表达采用了 RGB 组合的方式，平均蓝色分量是指 B 部分的平均值；平均色度 (color) 是指将 RGB 颜色模型转换成 HSL 颜色模型之后，H 部分的平均值；平均饱和度 (saturation) 是指将 RGB 颜色模型转换成 HSL 颜色模型之后，S 部分的平均值；平均亮度 (bright) 是指将 RGB 颜色模型转换成 HSL 颜色模型之后 L 部分的平均值；灰度值的统计直方图是指对灰度值 0 ～ 255 的分布进行分阶段统计得到的向量；灰度标准差 (greyscale) 是指目标物各个局部颜色的差异；颜色权重 (weighted) 是指计算机对彩色的表达采用了 RGB 组合的方式时根据像素点位置自动生成的平均色度与位置坐标的比值，该值仅用于纠错，不参与运算；

[0033] 5) 获取特征值后，输入预设的文本格式数据库，以表格的形式加载后续的分类识别算法。

[0034] 进一步的，在一个基于多种算法的分类识别的步骤中，采用基于相对距离的 KNN 算法，所述的 KNN 算法的步骤如下：

[0035] 首先为避免由于属性值域不同而影响样本距离的计算，特征值数据库的每个样本应该对第 i 维属性值为 X[i]，计算最大值 Max[i]、最小值 Min[i]，再利用公式 X[i] = (X[i]-Mini[i])/(Max[i] - Min[i]) 进行归一化操作，样品各属性归一化后其值域为 [0,1]，然后根据特征值数据库构建数据集 D = {X1, …, XL}，其中 $X_i \in R^n$，i = 1…L；设样本共有 ClassNum 个类；设 C_i 表示第 i 类中的所有样本的集合，且 $C_i \cap C_j = \Phi$ (i, j = 1, …, ClassNum)，样本集也可表示为：$D = C_1 \cup C_2 \cup \cdots \cup C_r$；

[0036] 设两个虫卵样本间的距离为 Dist，数据集 D 有 m 个属性，其数据集构成为 $R(A_1, A_2, \cdots, A_m)$，X 和 Y 分别为数据集 D 中的两个样本，则 X 与 Y 的距离度量公式为：

[0037] $$Dist(X,Y)=\sqrt{\sum_{i=1}^{m}(X.x_i-Y.y_i)^2}$$

[0038] 测试样本中第 i 类的 K- 最近邻距离均值为：

[0039] $$Avgdis(i)=\frac{\sum_{j=1}^{k_i} Dist(X_j,Y)}{k_i} \quad X_j \in C_i \quad i=1,\cdots,ClassNum$$

[0040] K_i 为 C_i 中的样本个数，Y 为 Xj 的最近邻，测试样本 X 和训练样本 Y 之间的相对距离即为：D = Dist(X,Y)/Avgdis(i)，Y ∈ C_i；

[0041] 在 N = 3 时，只要计算数据集各样本到测算样本的距离，比较选取测试样本的 3 个最近邻，即可判别它的类别，分类结果由 score 来体现，设输入图片的特征为 (f1,f2,...,fn)，数据库中某样本的特征是 (x11,x12,...,x1n)，则 score = s(f1,x11)*s(f2,x12)*.....*s(fn,x1n)；

[0042] 此处 s 函数构造为：令 maxV = max(f1,x11)；minV = min(f1,x11)，则 diff = (maxV-minV)/maxV；s = X*(pow(e,-diff)-1/e)+B，其中 e 是自然数，X = (A-B)*e/(e-1)，即可令 A 取到最大值，如果 diff = 0，则 s 是最大值 A；如果 diff = 1，则 s 是最小值 B。

[0043] 进一步的，在使用均值移位算法对目标进行分割处理之前，图片需要预先计算彩色直方图。

[0044] 进一步的，如因过度归一化等问题造成图像出现断点，导致遗失边界点无法构建区域框或区域像素集合，则用区域生长算法加以弥补，所述的区域生长算法的具体步骤是：先对每个需要分割的区域找一个种子像素作为生长的起点，然后将种子像素周围邻域中与种子像素具有相同或相似性质的像素合并到这一区域中，将这些新像素当作新的种子像素继续进行上面的过程，直到再没有满足条件的像素可被包括进来。

[0045] 进一步的，所述的寄生虫为华支睾吸虫、带绦虫、鞭虫、蛲虫、蛔虫、钩虫、阔节裂头绦虫、日本血吸虫、布氏姜片吸虫、肺吸虫、或者曼氏迭宫绦虫。

[0046] 进一步的，所述的待识别图具有旋转不变性、对光照无差异、对个体无差异。

[0047] 具体的，所述的特征数据库是指图像虫卵部分的特征值，如长度 (length)、宽度 (width)、椭圆度 (ovality)、面积 (area)、周长 (perimeter)、平均灰度 (grey)、平均红色分量 (red)、平均绿色分量 (green)、平均蓝色分量 (blue)、平均色度 (color)、 平均饱和度 (saturation)、平均亮度 (bright)、灰度标准差 (greyscale)、颜色权重 (weighted)，数据库格式参见附表。

[0048] 具体的，所述的总数据库是指由已知参数的寄生虫特征数据库组成的总库。

[0049] 本发明还提供了一种基于多特征融合的寄生虫虫卵的识别方法，其特征在于包括如下步骤：

[0050] a) 一个对图像预处理的步骤，在一个对图像预处理的步骤中，将显微照相设备获取的图像信息进行亮度归一化处理，对归一化的图像进行灰度化处理，生成归一化灰度图像，然后再对整张图片进行基于高斯滤波的锐化处理，得到虫卵边缘锐化的图像；

[0051] b) 一个采用人工辅助识别寻找虫卵的步骤，在一个采用人工辅助识别寻找虫卵的步骤中，采用增强 Grab Cut 法对虫卵边缘锐化的图像进行分割，用户提供限定方框进行人工支持，得到虫卵图像的颜色特征向量，基于颜色特征向量规划并找到最佳目标区域，得到判断为虫卵的区域；

[0052] c) 一个基于上述判别为虫卵的区域对虫卵图像进行目标获取的步骤，在进行目标获取的步骤中，依据所建立的要识别寄生虫虫卵形状边缘区域信息，对每个候选的边缘区

域进行二值化处理，采用边界跟踪算法按照虫卵区域的边界进行目标获取，得到分割后的虫卵图像；

[0053] d) 一个对分割后的虫卵图像截取指定特征值并存入预设特征数据库的步骤；

[0054] e) 一个基于多种算法的分类识别的步骤，在一个基于多种算法的分类识别的步骤中，采用基于相对距离的KNN(k = 3)算法，将所获取的特征值代入由11种寄生虫特征数据库组成的总数据库，基于KNN算法判断虫卵类别。

[0055] 进一步的，在一个采用人工辅助识别寻找虫卵的步骤中，采用增强Grab Cut法对虫卵边缘锐化的图像进行分割，用户提供限定方框进行人工支持，方框以外的部分不处理，用户通过设置背景区域TB来初始化三分图T，前景区域TF设置为空，未知区域TU设置为背景区域TB的补集，对于所有背景区域的像素，将它们的Alpha值设置为0，即a = 0；对于未知区域的像素点，将它们的Alpha值设置为1，即a = 1，分别用a = 0和a = 1这两个集合来初始化创建前景与背景的高斯混合模型，为未知区域中的每个像素点n设置高斯混合模型参数：

[0056] $kn = \arg\min Dn(a_n, k_n, \theta, Z_n)$，

[0057] 由图像中各个像素的数据求得高斯混和模型参数

[0058] $\theta = \arg\min U(a, k_n, \theta, Z_n)$，

[0059] 利用最小化能量公式来得到初始分割：

[0060] $\min_k E(a, k_n, \theta, Z_n)$，

[0061] 重复执行3次，进行边界优化。

[0062] 进一步的，基于上述判别为虫卵的区域对虫卵图像进行目标获取的步骤中，依据所建立的要识别寄生虫虫卵形状边缘区域信息，对每个候选的边缘区域进行二值化处理，采用边界跟踪算法进行目标获取，算法开始时按照从上到下的顺序搜索每个像素，设序列数组为K，首先从左上方开始搜索第一个目标像素点，设为k0，则像素k0是该区域最左上角的边界像素，也就是搜索的起点，设定搜索方向按逆时针，八邻域方向搜索，k0设置为跟踪标志，并将k0做为序列数组的第一个元素插入，按逆时针方向搜索下一个目标像素，并设为k，如果找不到，则k为孤立像素区域；若k等于搜索起始边界像素k0，则按顺序继续判断其它邻近方向上是否还有未跟踪到的边界像素，若没有，则已回到起始点，算法结束，序列K中的边界像素点组成一条封闭区域，将目标区域包围在内。

[0063] 进一步的，在一个对获取目标截取指定特征值存入特征数据库的步骤中，先获取虫卵图像的特征值：

[0064] 1) 求出边缘区域外接最小正方形区域，计数像素数即可得到长度(length)、宽度(width)，长度是指目标物外接矩形的长度，宽度是指目标物外接矩形的长度；

[0065] 2) 计数目标区域、目标周边区域的像素点，可得面积和周长，其面积与最小外接正方形之比值即为椭圆度(ovality)，椭圆度是指目标物面积与外接椭圆的面积之比；

[0066] 3) 面积(area)是指目标物面积，周长(perimeter)是指目标物周长；

[0067] 4) 基于目标区域颜色构成信息，获取RGB分量；将图片转化为灰度即可获取灰度值的统计直方，其均值为灰度值；将目标转化为HSV空间，即可获取HSL分量；平均灰度(grey)是指灰度化后的目标物的颜色平均值；平均红色分量是指计算机对彩色的表达采用了RGB组合的方式，平均红色分量(red)是指R部分的平均值；平均绿色分量(green)是

指计算机对彩色的表达采用了RGB组合的方式，平均绿色分量是指G部分的平均值；平均蓝色分量(blue)是指计算机对彩色的表达采用了RGB组合的方式，平均蓝色分量是指B部分的平均值；平均色度(color)是指将RGB颜色模型转换成HSL颜色模型之后，H部分的平均值；平均 饱和度(saturation)是指将RGB颜色模型转换成HSL颜色模型之后，S部分的平均值；平均亮度(bright)是指将RGB颜色模型转换成HSL颜色模型之后L部分的平均值；灰度值的统计直方图是指对灰度值0～255的分布进行分阶段统计得到的向量；灰度标准差(greyscale)是指目标物各个局部颜色的差异；颜色权重(weighted)是指计算机对彩色的表达采用了RGB组合的方式时根据像素点位置自动生成的平均色度与位置坐标的比值，该值仅用于纠错，不参与运算；

[0068] 5)获取特征值后，输入预设的文本格式数据库，以表格的形式加载后续的分类识别算法。

[0069] 进一步的，在一个基于多种算法的分类识别的步骤中，采用基于相对距离的KNN算法，KNN算法的步骤如下：

[0070] 首先为避免由于属性值域不同而影响样本距离的计算，特征值数据库的每个样本应该对第i维属性值为X[i]，计算最大值Max[i]、最小值Min[i]，再利用公式X[i]＝(X[i]-Mini[i])/(Max[i]-Min[i])进行归一化操作，样品各属性归一化后其值域为[0,1]，然后根据特征值数据库构建数据集D＝{X1，…，XL}，其中$X_i \in R^n$，i＝1…L；设样本共有ClassNum个类；设C_i表示第i类中的所有样本的集合，且$C_i \cap C_j = \Phi$(i，j＝1，…，ClassNum)，样本集也可表示为：$D = C_1 \cup C_2 \cup \cdots \cup C_r$；

[0071] 设两个虫卵样本间的距离为Dist，数据集D有m个属性，其数据集构成为$R(A_1, A_2, \cdots, A_m)$，X和Y分别为数据集D中的两个样本，则X与Y的距离度量公式为：

[0072] $$Dist(X,Y)=\sqrt{\sum_{i=1}^{m}(X.x_i-Y.y_i)^2}$$

[0073] 测试样本中第i类的K-最近邻距离均值为：

[0074] $$Avgdis(i)=\frac{\sum_{j=1}^{k_i} Dist(X_j,Y)}{k_i} \quad X_j \in C_i \quad i=1,\cdots,ClassNum$$

[0075] K_i为C_i中的样本个数，Y为Xj的最近邻，测试样本X和训练样本Y之间的相对距离即为：D＝Dist(X,Y)/Avgdis(i)，$Y \in C_i$；

[0076] 在N＝3时，只要计算数据集各样本到测算样本的距离，比较选取测试样本的3个最近邻，即可判别它的类别，分类结果由score来体现，设输入图片的特征为(f1,f2,...,fn)，数据库中某样本的特征是(x11,x12,...,x1n)，则score＝s(f1,x11)*s(f2,x12)*.....*s(fn,x1n)；

[0077] 此处s函数构造为：令maxV＝max(f1,x11)；minV＝min(f1,x11)，则diff＝(maxV-minV)/maxV；s＝X*(pow(e,-diff)-1/e)+B，其中e是自然数，X＝(A-B)*e/(e-1)，即可令A取到最大值，如果diff＝0，则s是最大值A；如果diff＝1，则s是最小值B。

[0078] 进一步的，如因过度归一化等问题造成图像出现断点，导致遗失边界点，则用区域生长算法加以弥补，所述的区域生长算法的具体步骤是：先对每个需要分割的区域找一个

种子像素作为生长的起点，然后将种子像素周围邻域中与种子像素具有相同或相似性质的像素合并到这一区域中，将这些新像素当作新的种子像素继续进行上面的过程，直到再没有满足条件的像素可被包括进来。

[0079] 进一步的，所述的寄生虫为华支睾吸虫、带绦虫、鞭虫、蛲虫、蛔虫、钩虫、阔节裂头绦虫、日本血吸虫、布氏姜片吸虫、肺吸虫、或者曼氏迭宫绦虫。

[0080] 进一步的，所述的待识别图具有有旋转不变性、对光照无差异、对个体无差异。

[0081] 本发明提供一个全自动将虫卵从待识别图像中分割出来的技术来获取特征值。由于寄生虫虫卵的不规则的形态结构、空间方位和标本杂质较多等原因，全自动识别难度大，本发明同时提供了基于 Grab Cut 数字抠图方法的人工辅助手段。由于本发明的难点在于自动从虫卵图片中分割出目标物，所以要求系统对输入图像具有旋转不变性（即对拍摄角度不敏感），对光照差异、个体差异具有较强的适应性，对虫卵种类具有可扩展性。

[0082] 人体寄生虫可分为单细胞的原虫 (protozoon)、多细胞的蠕虫 (helminth) 和节肢动物 (arthorpod) 三大类，而我国常见的主要是十多种蠕虫，又主要分为线虫 (nematode)、I 吸虫 (trematode) 和绦虫 (eestode) 和棘头虫 (aeanthocphala) 四类。本发明以华支睾吸虫、带绦虫、鞭虫、蛲虫、蛔虫、钩虫、阔节裂头绦虫、日本血吸虫、布氏姜片吸虫、肺吸虫、曼氏迭宫绦虫等 11 种我国常见人体寄生虫虫卵为预定的识别对象。

[0083] 本发明提供了一个基于 KNN 算法的虫卵分类器，利用特征值分类。随着计算机视觉与各种先进医疗成像设备的不断发展，单一的图像特征很难全面、精确地表达医学图像的内容，多特征融合已成为提取医学图像有效特征的必然途径。在尽量保留原始信息的基础上，克服了原始数据量大而不稳定的特点，提取的融合特征可以有效地用于图像识别。人体寄生虫的识别取决于虫卵的形状、大小、内含物、颜色和特殊结构（如卵壳厚薄、特异形态）等特征。本发明将上述视觉特征转化为可通过相关算法定量提取的周长、面积、圆形度、颜色和纹理 15 个分类特征。

[0084] 本发明和已有技术相比，其技术效果是明显的。(1) 本系统能够对图片的旋转不敏感（即对拍摄角度不仅敏感），对光照差异、虫卵个体差异有一定的容忍度；(2) 在数学 特征提取阶段，本系统紧密结合了虫卵的生物学特征，具有很强的针对性，较高的区分度，从而识别更加准确；(3) 本系统提供了全自动分割和半自动分割两种选择，其中半自动分割算法对复杂背景的虫卵图像具有很好的适应性；全自动分割具有很高的自动化程度，对后续的自动测试、统计、分析等工作提供了便捷。(4) 本系统目前对 11 中虫卵的识别准确度超过 90％，达到较理想的结果。

附图说明：

[0085] 图 1 是基于全自动分割的虫卵图像识别流程。

[0086] 图 2 是边缘锐化后的图片。

[0087] 图 3 是基于均值位移算法找到目标区域并计算区域框。

[0088] 图 4 是基于目标区域和边界跟踪算法提取目标。

[0089] 图 5 是特征值数据库结构和采集用例。

[0090] 图 6 是基于人工辅助的虫卵图像识别流程。

[0091] 图 7 是边缘锐化后的图片。

[0092] 图 8 是基于 Grab Cut 法的人工画框辅助分割。

[0093] 图 9 是基于人工辅助和边界跟踪算法提取目标。

[0094] 图 10 是特征值数据库结构和采集用例。

具体实施方式：

[0095] 下述实施例采用的是一个血吸虫卵的血液涂片的显微照片。

[0096] 实施例 1

[0097] 一种基于多特征融合的寄生虫虫卵的全自动识别方法，包括如下流程（流程图见图 1）：

[0098] a) 一个对图像预处理的步骤，将显微照相设备获取的图像信息进行亮度归一化处理，对归一化的图像进行灰度化处理，生成归一化灰度图像，然后再对整张图片进行基于高斯滤波的锐化处理，得到虫卵边缘锐化的图像；

[0099] b) 一个对虫卵边缘锐化的图像进行均值移位寻找虫卵的步骤，使用均值移位算法来对目标图片进行分割处理，得到该图像的颜色特征向量，基于颜色向量规划并找到最佳目标区域，得到判断为虫卵的区域；

[0100] c) 一个基于上述判别为虫卵的区域对虫卵图像进行目标获取的步骤。基于上述形状分割信息，即依据所建立的要识别寄生虫虫卵形状边缘区域信息，对每个候 选的边缘区域进行二值化处理，采用边界跟踪算法按照虫卵区域的边界进行目标获取，得到分割后的虫卵图像；

[0101] d) 一个对分割后的虫卵图像截取指定特征值，存入预设特征数据库的步骤；

[0102] e) 一个基于多种算法的分类识别的步骤，在本发明所述的一个基于多种算法的分类识别的步骤中，采用基于相对距离的 KNN(k = 3) 算法，将所获取的特征值代入包含 11 种寄生虫特征数据库的总数据库，基于 KNN 算法判断虫卵类别；

[0103] 进一步的，在所述的对图像预处理的步骤中，将显微照相设备获取的图像信息进行亮度归一化处理，对归一化的图像进行灰度化处理，生成归一化灰度图像，然后再对整张图片进行基于高斯滤波的锐化处理，得到虫卵边缘锐化的图像（用例图见图 2）；

[0104] 进一步的，一个对虫卵边缘锐化的图像进行均值移位的步骤，在一个对虫卵边缘锐化的图像进行均值移位的步骤中，使用均值移位算法来对目标进行分割处理，在使用均值移位算法来对目标进行分割处理的过程中，先对原图像进行 X×Y 的划分，得到 X×Y 个交点，并对这些交点进行合并处理，即某两个点对应的颜色值之间的欧氏距离小于某个阈值（为保证软件运行速度，本发明中此阈值定义为图像亮度最高的 5% 像素与亮度最低的 5% 像素的颜色平均值），则把它们合为一个点，这样得到 m 个点作为初始点集合，m 代表图片上 X×Y 共 n 个像素点的集合，每个像素点可以表示为自变量 $X_i\{i = 1\cdots n\}$，样本点平均值位移 M 的计算方法为：

[0105] $$M_{h,U}(x) = \frac{h^2}{d+2}\frac{\overline{\nabla} f_E(x)}{\overline{f}_U(x)}$$

[0106] 在图片中心选择一个初始点，在以此点为中心的窗口 $S_h(x)$ 内计算平均值位移 $M_{h,U}(x)$，如果该值不小于某个阈值，就把窗口 $S_h(x)$ 平移 $M_{h,U}(x)$，然后重复在新的窗口中计算

平均值位移，得到新的中心值，直到 $M_{h,U}(x)$ 小于某个阈值，停止平移，得到一个最大局部密度位置；重复上述步骤，得到 m 个对应最大局部密度位置的点，并对这些点进行合并处理，得到 n 个聚类的中心点，即原图像的主色，针对原图像中的每个像素点，根据欧氏距离判断归到哪个聚类中，用一维直方图表示主色信息，横坐标表示各主色，纵坐标表示各主色包含的像素数的比例，这样就得到该图像的颜色特征向量：

[0107] $P = \{(P_i, W_i) i = 1, \cdots, n\}$，其中 $P_i = (L^*_i, a^*_i, b^*_i)$，$W_i \in (0, 1]$。

[0108] 基于颜色向量 p 和 q 规划最佳目标区域，公式为

[0109] $$EMD(P,Q) = \min \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} d(p_i, q_j) f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}}$$

[0110] 其中与预期中心点相似度 EMD 最高的区域就是目标区（用例图见图 3）。

[0111] 传统上图像分割主要包括并行边界分割、串行边界分割、并行区域分割和串行区域分割四类。过往研究指出并行边界分割对于灰度均匀变化的图像不适用，而串行区域分割算法（主要有区域生长法和松弛迭代法）虽然很热门，但一般认为其计算量较大，不符合本发明的限定条件。闭值分割的需要先设置一个阈值，然后把图像中的像素点和闽值相比较，把像素划分为目标和背景并加以分割。但在寄生虫虫卵检测照片中，由于杂质的污染，目标与背景之间由自然形成的灰度差不能满足分割的要求。

[0112] 而均值移位算法能够在寄生虫虫卵图片这类复杂概率分布中，沿着最短路径使得每一个像素点找到密度函数的局部极大值点。经测试，利用均值移位算法的统计鲁棒性和沿着密度梯度方向快速收敛的特性以及彩色直方图算法对目标形状的匹配，可以解决了非刚性目标形态多变、跟踪复杂程度高的问题。

[0113] 本发明中均值移位是指：一种基于非参数的核密度估计理论，是利用梯度法迭代计算概率密度函数的极值点的方法。该算法具有无参数、快速模式匹配的特点，是一种有效的目标跟踪算法。

[0114] 本发明中区域生长 (region growing) 是指：将具有相似性质的像素集合起来构成区域。具体步骤是：先对每个需要分割的区域找一个种子像素作为生长的起点，然后将种子像素周围邻域中与种子像素具有相同或相似性质的像素合并到这一区域中。将这些新像素当作新的种子像素继续进行上面的过程，直到再没有满足条件的像素可被包括进来。

[0115] 进一步的，基于上述判别为虫卵的区域对虫卵图像进行目标获取的步骤，基于上述形状分割信息，即依据所建立的要识别寄生虫虫卵形状边缘区域信息，对每个候选的边缘区域进行二值化处理，采用边界跟踪算法进行目标获取，算法开始时按照从上到下的顺序搜索每个像素，设序列数组为 K，首先从左上方开始搜索第一个目标像素点，设为 k0，则像素 k0 是该区域最左上角的边界像素，也就是搜索的起点，设定搜索方向按逆时针，八邻域方向搜索，k0 设置为跟踪标志，并将 k0 做为序列数组的第一个元素插入，按逆时针方向搜索下一个目标像素，并设为 k，如果找不到，则 k 为孤立像素区域；若 k 等于搜索起始边界像素 k0，则按顺序继续判断其它邻近方向上是否还有未跟踪到的边 界像素，若没有，则已回到起始点，算法结束，序列 K 中的边界像素点组成一条封闭区域，将目标区域包围在内（用例图见图 4）；

[0116] 进一步的，在一个对获取目标截取指定特征值并存入特征数据库的步骤中，在所述的对获取目标截取指定特征值，存入特征数据库的步骤中，先获取虫卵图像的特征值（参见附表）：

[0117] 1）求出边缘区域外接最小正方形区域，计数像素数即可得到长度（length）、宽度（width），长度是指目标物外接矩形的长度，宽度是指目标物外接矩形的长度；

[0118] 2）计数目标区域、目标周边区域的像素点，可得面积和周长，其面积与最小外接正方形之比值即为椭圆度（ovality），椭圆度是指目标物面积与外接椭圆的面积之比；

[0119] 3）面积（area）是指目标物面积，周长（perimeter）是指目标物周长；

[0120] 4）基于目标区域颜色构成信息，获取 RGB 分量；将图片转化为灰度即可获取灰度值的统计直方，其均值为灰度值；将目标转化为 HSV 空间，即可获取 HSL 分量；平均灰度（grey）是指灰度化后的目标物的颜色平均值；平均红色分量是指计算机对彩色的表达采用了 RGB 组合的方式，平均红色分量（red）是指 R 部分的平均值；平均绿色分量（green）是指计算机对彩色的表达采用了 RGB 组合的方式，平均绿色分量是指 G 部分的平均值；平均蓝色分量（blue）是指计算机对彩色的表达采用了 RGB 组合的方式，平均蓝色分量是指 B 部分的平均值；平均色度（color）是指将 RGB 颜色模型转换成 HSL 颜色模型之后，H 部分的平均值；平均饱和度（saturation）是指将 RGB 颜色模型转换成 HSL 颜色模型之后，S 部分的平均值；平均亮度（bright）是指将 RGB 颜色模型转换成 HSL 颜色模型之后 L 部分的平均值；灰度值的统计直方图是指对灰度值 0 ～ 255 的分布进行分阶段统计得到的向量；灰度标准差（greyscale）是指目标物各个局部颜色的差异；颜色权重（weighted）是指计算机对彩色的表达采用了 RGB 组合的方式时根据像素点位置自动生成的平均色度与位置坐标的比值，该值仅用于纠错，不参与运算；

[0121] 5）获取特征值后，输入预设的文本格式数据库，以表格的形式加载后续的分类识别算法，数据库结构如下表所示。

[0122]

length	width	area	grayscale	red	green	blue	color	saturation	bright	ovality	perimeter	grey	weighted
495	440	198400	129.7311	125.0611	140.8316	123.3005	114.0764	0.126617	123	0.945418	680	1291.363	68.75533
484	540	209408	75.81782	78.96598	82.71077	65.7767	75.31776	0.214129	65	0.977195	608	1014.349	50.46343
528	495	212524	73.70431	77.17218	81.04894	62.89181	75.01997	0.23503	62	0.957164	488	691.4633	168.3718
595	500	249712	66.8332	70.87652	73.55594	56.06613	69.38618	0.238685	56	0.972942	688	720.0351	216.4935
524	468	226792	67.19308	71.80312	74.09469	55.68144	69.34189	0.257707	55	0.996875	672	461.1149	117.8513

[0123] 进一步的,在一个基于多种算法的分类识别的步骤中,采用基于相对距离的 KNN 算法,具体 KNN 算法的步骤如下 :

[0124] 首先为避免由于属性值域不同而影响样本距离的计算,特征值数据库的每个样本

应该对第 i 维属性值为 X[i]，计算最大值 Max[i]、最小值 Min[i]，再利用公式 X[i] = (X[i] -Mini[i])/(Max[i] - Min[i]) 进行归一化操作，样品各属性归一化后其值域为 [0,1]，然后根据特征值数据库构建数据集 D = {X1, …, XL}，其中 $X_i \in R^n$，i = 1…L；设样本共有 ClassNum 个类；设 C_i 表示第 i 类中的所有样本的集合，且 $C_i \cap C_j = \Phi$ (i, j = 1, …, ClassNum)，样本集也可表示为：$D = C_1 \cup C_2 \cup \cdots \cup C_r$；

[0125] 设两个虫卵样本间的距离为 Dist，数据集 D 有 m 个属性，其数据集构成为 $R(A_1, A_2, \cdots, A_m)$，X 和 Y 分别为数据集 D 中的两个样本，则 X 与 Y 的距离度量公式为：

[0126] $$Dist(X,Y)=\sqrt{\sum_{i=1}^{m}(X.x_i-Y.y_i)^2}$$

[0127] 测试样本中第 i 类的 K- 最近邻距离均值为：

[0128] $$Avgdis(i)=\frac{\sum_{j=1}^{k_i} Dist(X_j,Y)}{k_i} \quad X_j \in C_i \quad i=1,\cdots,ClassNum$$

[0129] K_i 为 C_i 中的样本个数，Y 为 Xj 的最近邻，测试样本 X 和训练样本 Y 之间的相对距离即为：D = Dist(X,Y)/Avgdis(i)，$Y \in C_i$；

[0130] 在 N = 3 时，只要计算数据集各样本到测算样本的距离，比较选取测试样本的 3 个最近邻，即可判别它的类别，分类结果由 score 来体现，设输入图片的特征为 (f1,f2,...,fn)，数据库中某样本的特征是 (x11,x12,...,x1n)，则 score = s(f1,x11)*s(f2,x12)*.....*s(fn,x1n)；

[0131] 此处 s 函数构造为：令 maxV = max(f1,x11)；minV = min(f1,x11)，则 diff = (maxV-minV)/maxV；s = X*(pow(e,-diff)-1/e)+B，其中 e 是自然数，X = (A-B)*e/(e-1)，即可令 A 取到最大值，如果 diff = 0，则 s 是最大值 A；如果 diff = 1，则 s 是最小值 B。

[0132] 进一步的，在使用均值移位算法对目标进行分割处理之前，图片需要预先计算彩色直方图。此处使用 VC++ 自带的彩色直方计算标准算法模块，使用默认参数。

[0133] 进一步的，如因过度归一化等问题造成图像出现断点，导致遗失边界点无法构建区域框或区域像素集合，则用区域生长算法加以弥补，所述的区域生长算法的具体步骤是：先对每个需要分割的区域找一个种子像素作为生长的起点，然后将种子像素周围邻域中与种子像素具有相同或相似性质的像素合并到这一区域中，将这些新像素当作新的种子像素继续进行上面的过程，直到再没有满足条件的像素可被包括进来。

[0134] 实施例 2

[0135] 一种基于多特征融合的寄生虫虫卵的人工辅助识别方法，其特征在于包括如下流程（流程图见图 6）：

[0136] a) 一个对图像预处理的步骤。将显微照相设备获取的图像信息进行亮度归一化处理，对归一化的图像进行灰度化处理，生成归一化灰度图像，然后再对整张图片进行基于高斯滤波的锐化处理，得到虫卵边缘锐化的图像；

[0137] b) 一个采用增强 Grab Cut 法对虫卵边缘锐化的图像进行人工辅助识别寻找虫卵的步骤。在目标分割的步骤中，采用增强 Grab Cut 法，即用户提供限定方框进行人工支持，

更精确的划分前景和背景，得到该图像的颜色特征向量，基于颜色向量规划并找到最佳目标区域，得到判断为虫卵的区域；

[0138] c) 一个基于上述判别为虫卵的区域对虫卵图像进行目标获取的步骤。基于上述形状分割信息，即依据所建立的要识别寄生虫虫卵形状边缘区域信息，对每个候选的边缘区域进行二值化处理，采用边界跟踪算法按照虫卵区域的边界进行目标获取，得到分割后的虫卵图像；

[0139] d) 一套对分割后的虫卵图像截取指定特征值并存入预设特征数据库的步骤；

[0140] e) 一个基于多种算法的分类识别的步骤，在本发明所述的一个基于多种算法的分类识别的步骤中，采用基于相对距离的 KNN(k = 3) 算法，将所获取的特征值代入总数据库，基于 KNN 算法判断虫卵类别；

[0141] 进一步的，在所述的对图像预处理的步骤中，将显微照相设备获取的图像信息进行亮度归一化处理，对归一化的图像进行灰度化处理，生成归一化灰度图像，然后再对整张图片进行基于高斯滤波的锐化处理，得到虫卵边缘锐化的图像（用例图见图 7）；

[0142] 进一步的，一个采用增强Grab Cut法对虫卵边缘锐化的图像进行人工辅助识别寻找虫卵的步骤。用户提供限定方框进行人工支持，方框以外的部分不处理，所以比全自动计算均值位移算法更快更精准。用户需要通过设置背景区域 TB 来初始化三分图 T，前景区域 TF 设置为空，未知区域 TU 设置为背景区域 TB 的补集，对于所有背景区域的像素，将它们的 Alpha（透明度）值设置为 0，即 a = 0；对于未知区域的像素点，将它们的 Alpha 值设置为 1，即 a = 1，分别用 a = 0 和 a = 1 这两个集合来初始化创建前景与背景的高斯混合模型，为未知区域中的每个像素点 n 设置高斯混合模型参数：

[0143] $kn = \arg\min Dn(a_n, k_n, \theta, Z_n)$，

[0144] 由图像中各个像素的数据求得高斯混和模型参数

[0145] $\theta = \arg\min U(a, k_n, \theta, Z_n)$，

[0146] 利用最小化能量公式来得到初始分割：

[0147] $\min_k E(a, k_n, \theta, Z_n)$，

[0148] 重复执行 3 次，进行边界优化（用例图见图 8）；

[0149] 一些杂质较多的图片用边界跟踪算法不能保证产生闭合的边界，并且算法也可能失控而偏离图像边界，特别是对某些边界较薄、多重边界且边界轮廓线附近灰度变化不太明显的图像更是如此。所以本发明提供人工干预作为辅助手段。前景背景提取的集合被分别初始化为三分图的未知区域部分和背景区域部分。初始化中的用户交互部分将影响到最终的分割结果。根据所给定的初始信息，来为初始的前景抠图区域与背景抠图区域分别创建高斯混合模型的 K 个组件。

[0150] 本发明中增强 Grab Cut 方法是指基于限定寄生虫虫卵图片，在 Graph Cuts 方法的基础上对算法适应性作了以下 3 方面的增强：第一，利用高斯混合模型 (Gaussian Mixture Model，GMM) 取代直方图来描述前景与背景像素的分布，由对灰度图像的处理上升到对彩色图像的处理；第二，利用迭代方法求取高斯混合模型中的各个参数替代了一次最小化估计来完成能量最小化的计算过程；第三，通过非完全标记方法，减少了用户在交互过程中的工作量，用户只需利用矩形框标记出背景区域即可。

[0151] 进一步的，基于上述判别为虫卵的区域对虫卵图像进行目标获取的步骤。基于上

述形状分割信息，即依据所建立的要识别寄生虫虫卵形状边缘区域信息，对每个候选的边缘区域进行二值化处理，采用边界跟踪算法进行目标获取，算法开始时按照从上到下的顺序搜索每个像素，设序列数组为 K，首先从左上方开始搜索第一个目标像素点，设为 k0，则像素 k0 是该区域最左上角的边界像素，也就是搜索的起点，设定搜索方向按逆时针，八邻域方向搜索，k0 设置为跟踪标志，并将 k0 做为序列数组的第一个元素插入，按逆时针方向搜索下一个目标像素，并设为 k，如果找不到，则 k 为孤立像素区域；若 k 等于搜索起始边界像素 k0，则按顺序继续判断其它邻近方向上是否还有未跟踪到的边界像素，若没有，则已回到起始点，算法结束，序列 K 中的边界像素点组成一条封闭区域，将目标区域包围在内（用例图见图 9）；

[0152] 进一步的，一个对获取目标截取指定特征值，存入特征数据库的步骤，在所述的对获取目标截取指定特征值，存入特征数据库的步骤中，先需要获取的特征值名称、获取方法：

[0153] 1）求出边缘区域外接最小正方形区域，计数像素数即可得到长度（length）、宽度（width），长度是指目标物外接矩形的长度，宽度是指目标物外接矩形的长度；

[0154] 2）计数目标区域、目标周边区域的像素点，可得面积和周长，其面积与最小外接正方形之比值即为椭圆度（ovality），椭圆度是指目标物面积与外接椭圆的面积之比；

[0155] 3）面积（area）是指目标物面积，周长（perimeter）是指目标物周长；

[0156] 4）基于目标区域颜色构成信息，获取 RGB 分量；将图片转化为灰度即可获取灰度值的统计直方，其均值为灰度值；将目标转化为 HSV 空间，即可获取 HSL 分量；平均灰度（grey）是指灰度化后的目标物的颜色平均值；平均红色分量是指计算机对彩色的表达采用了 RGB 组合的方式，平均红色分量（red）是指 R 部分的平均值；平均绿色分量（green）是指计算机对彩色的表达采用了 RGB 组合的方式，平均绿色分量是指 G 部分的平均值；平均蓝色分量（blue）是指计算机对彩色的表达采用了 RGB 组合的方式，平均蓝色分量是指 B 部分的平均值；平均色度（color）是指将 RGB 颜色模型转换成 HSL 颜色模型之后，H 部分的平均值；平均饱和度（saturation）是指将 RGB 颜色模型转换成 HSL 颜色模型之后，S 部分的平均值；平均亮度（bright）是指将 RGB 颜色模型转换成 HSL 颜色模型之后 L 部分的平均值；灰度值的统计直方图是指对灰度值 0 ～ 255 的分布进行分阶段统计得到的向量；灰度标准差（greyscale）是指目标物各个局部颜色的差异；颜色权重（weighted）是指计算机对彩色的表达采用了 RGB 组合的方式时根据像素点位置自动生成的平均色度与位置坐标的比值，该值仅用于纠错，不参与运算；

[0157] 5）获取特征值后，输入预设的文本格式数据库，以表格的形式加载后续的分类识别算法，数据库范例如下表。

[0158]

length	width	area	grayscale	red	green	blue	color	saturation	bright	quality	perimeter	error	weighted
496	440	198400	129.7311	125.0611	140.8316	123.3006	114.0764	0.126617	123	0.945418	680	1291.363	68.75533
484	540	209408	75.81782	78.96598	82.71077	65.7767	75.31776	0.214129	65	0.977195	608	1014.349	50.46343
528	496	212624	73.70431	77.17218	81.04894	62.89181	75.01997	0.23503	62	0.957164	488	691.4633	158.3718
596	500	249712	66.8332	70.87652	73.55594	56.06613	69.38518	0.238585	56	0.972942	688	720.0351	216.4935
624	468	226792	67.19308	71.80312	74.09469	55.68144	69.34189	0.257707	55	0.995875	672	461.1149	117.8513

[0159] 特征值采集的运行用例见图 10 ；

[0160] 进一步的，在一个基于多种算法的分类识别的步骤中，采用基于相对距离的 KNN

算法，KNN 算法的步骤如下：

[0161] 首先为避免由于属性值域不同而影响样本距离的计算，特征值数据库的每个样本应该对第 i 维属性值为 X[i]，计算最大值 Max[i]、最小值 Min[i]，再利用公式 X[i] = (X[i]-Mini[i])/(Max[i] - Min[i]) 进行归一化操作，样品各属性归一化后其值域为 [0,1]，然后根据特征值数据库构建数据集 D = {X1, …, XL}，其中 $X_i \in R^n$，i = 1…L；设样本共有 ClassNum 个类；设 C_i 表示第 i 类中的所有样本的集合，且 $C_i \cap C_j = \Phi$(i, j = 1, …, ClassNum)，样本集也可表示为：$D = C_1 \cup C_2 \cup \cdots \cup C_r$；

[0162] 设两个虫卵样本间的距离为 Dist，数据集 D 有 m 个属性，其数据集构成为 $R(A_1, A_2, \cdots, A_m)$，X 和 Y 分别为数据集 D 中的两个样本，则 X 与 Y 的距离度量公式为：

[0163] $$Dist(X,Y)=\sqrt{\sum_{i=1}^{m}(X.x_i-Y.y_i)^2}$$

[0164] 测试样本中第 i 类的 K- 最近邻距离均值为：

[0165] $$Avgdis(i)=\frac{\sum_{j=1}^{k_i} Dist(X_j,Y)}{k_i} \quad X_j \in C_i \quad i=1,\cdots,ClassNum$$

[0166] K_i 为 C_i 中的样本个数，Y 为 Xj 的最近邻，测试样本 X 和训练样本 Y 之间的相对距离即为：D = Dist(X,Y)/Avgdis(i)，$Y \in C_i$；

[0167] 在 N = 3 时，只要计算数据集各样本到测算样本的距离，比较选取测试样本的 3 个最近邻，即可判别它的类别，分类结果由 score 来体现，设输入图片的特征为 (f1,f2,...,fn)，数据库中某样本的特征是 (x11,x12,...,x1n)，则 score = s(f1,x11)*s(f2,x12)*.....*s(fn,x1n)；

[0168] 此处 s 函数构造为：令 maxV = max(f1,x11)；minV = min(f1,x11)，则 diff = (maxV-minV)/maxV；s = X*(pow(e,-diff)-1/e)+B，其中 e 是自然数，X = (A-B)*e/(e-1)，即可令 A 取到最大值，如果 diff = 0，则 s 是最大值 A；如果 diff = 1，则 s 是最小值 B。

[0169] 进一步的，如因过度归一化等问题造成图像出现断点，导致遗失边界点，则用区域生长算法加以弥补，所述的区域生长算法的具体步骤是：先对每个需要分割的区域找一个种子像素作为生长的起点，然后将种子像素周围邻域中与种子像素具有相同或相似性质的像素合并到这一区域中，将这些新像素当作新的种子像素继续进行上面的过程，直到再没有满足条件的像素可被包括进来。

[0170] 本发明中多个特征通过相乘的方式组合。由于两个虫卵可能其他方面相似，但是在某一特征方面差异非常大（实际镜检中往往根据一个特征就否定了同类的可能性），相乘可以把这个极大差异更好的表现出来，去影响总 score 的值。

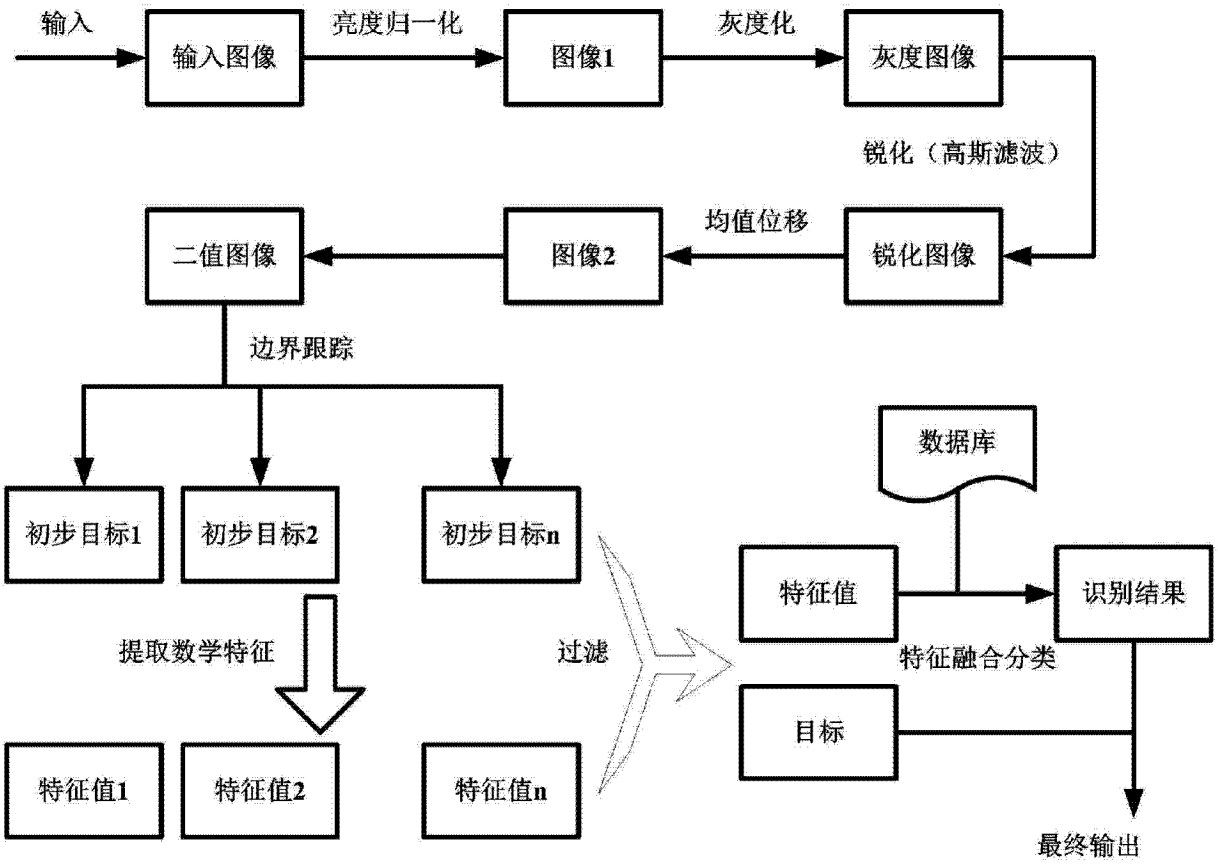

输入
输入图像
亮度归一化
图像1
灰度化
灰度图像
锐化（高斯滤波）
锐化图像
均值位移
图像2
二值图像
边界跟踪
初步目标1
初步目标2
初步目标n
提取数学特征
特征值1
特征值2
特征值n
过滤
数据库
特征值
识别结果
特征融合分类
目标
最终输出

图 1

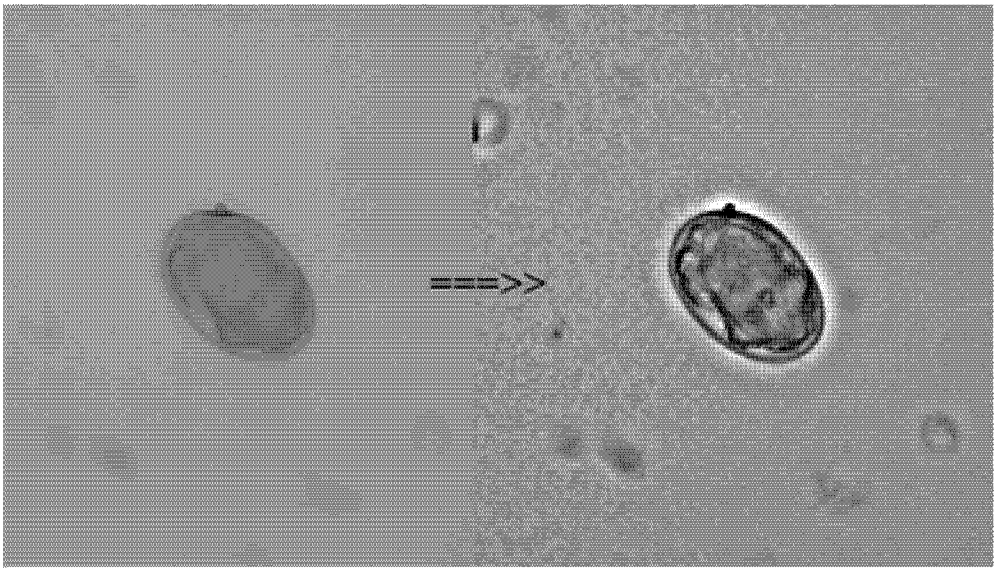

===>>

图 2

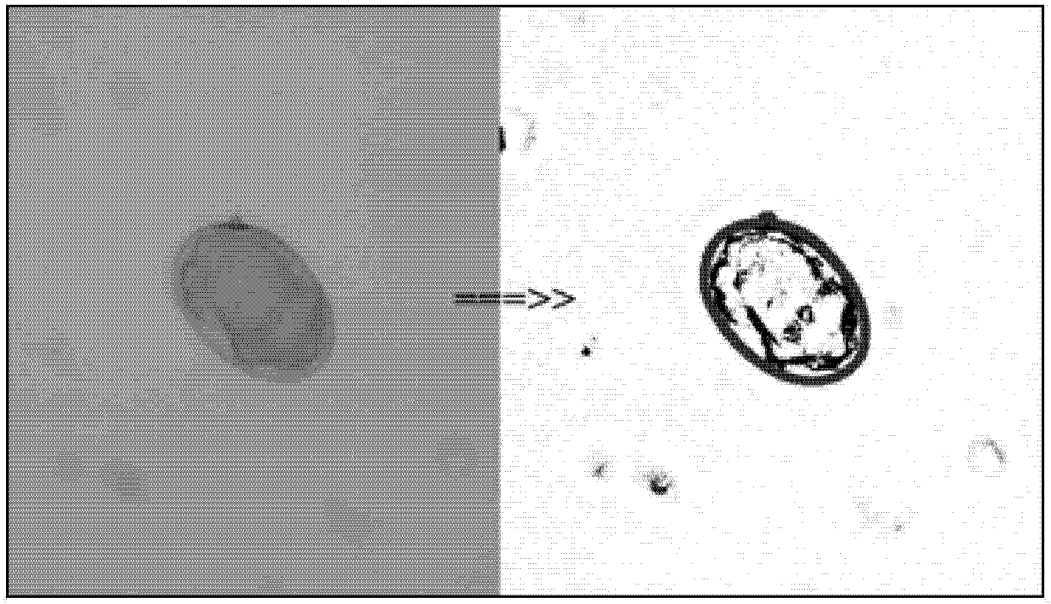


图 3

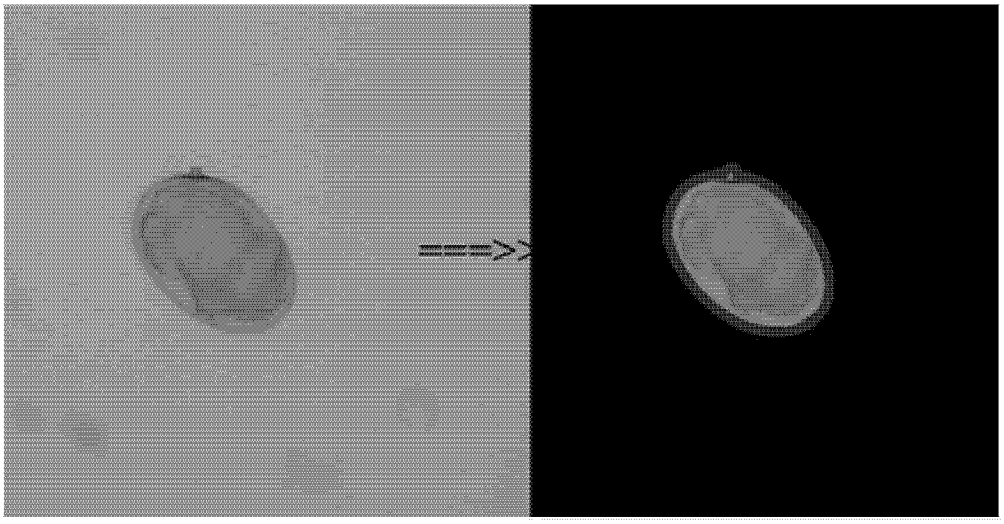


图 4

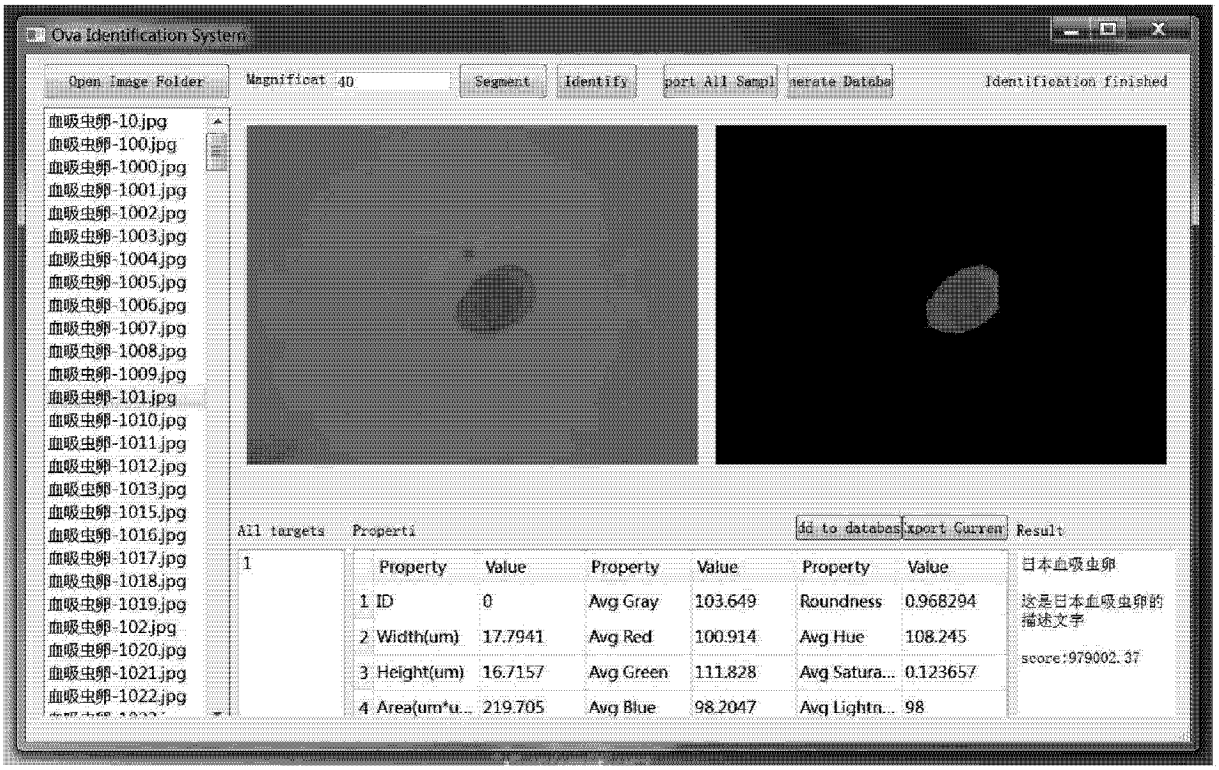

Ova Identification System
Open Image Folder
Magnificat 40
Segment
Identify
port All Sampl
nerate Databa
Identification finished
血吸虫卵-10.jpg
血吸虫卵-100.jpg
血吸虫卵-1000.jpg
血吸虫卵-1001.jpg
血吸虫卵-1002.jpg
血吸虫卵-1003.jpg
血吸虫卵-1004.jpg
血吸虫卵-1005.jpg
血吸虫卵-1006.jpg
血吸虫卵-1007.jpg
血吸虫卵-1008.jpg
血吸虫卵-1009.jpg
血吸虫卵-101.jpg
血吸虫卵-1010.jpg
血吸虫卵-1011.jpg
血吸虫卵-1012.jpg
血吸虫卵-1013.jpg
血吸虫卵-1015.jpg
血吸虫卵-1016.jpg
血吸虫卵-1017.jpg
血吸虫卵-1018.jpg
血吸虫卵-1019.jpg
血吸虫卵-102.jpg
血吸虫卵-1020.jpg
血吸虫卵-1021.jpg
血吸虫卵-1022.jpg
All targets
Properti
dd to databas
xport Curren
Result
1
Property Value Property Value Property Value
1 ID 0 Avg Gray 103.649 Roundness 0.968294
2 Width(um) 17.7941 Avg Red 100.914 Avg Hue 108.245
3 Height(um) 16.7157 Avg Green 111.828 Avg Satura... 0.123657
4 Area(um*u... 219.705 Avg Blue 98.2047 Avg Lightn... 98
日本血吸虫卵
这是日本血吸虫卵的描述文字
score:979002.37

图 5

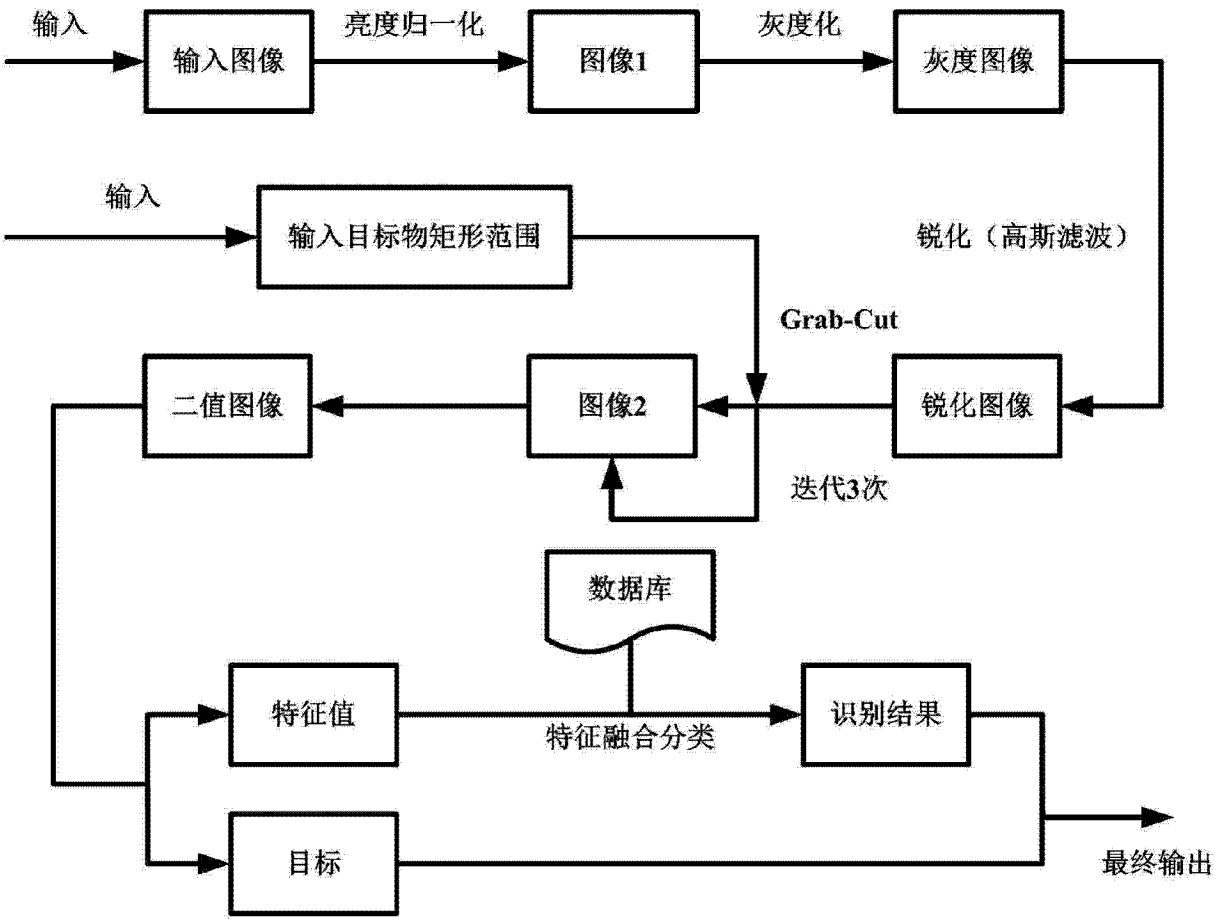

输入
输入图像
亮度归一化
图像1
灰度化
灰度图像
输入
输入目标物矩形范围
锐化（高斯滤波）
Grab-Cut
二值图像
图像2
锐化图像
迭代3次
数据库
特征值
特征融合分类
识别结果
目标
最终输出

图 6

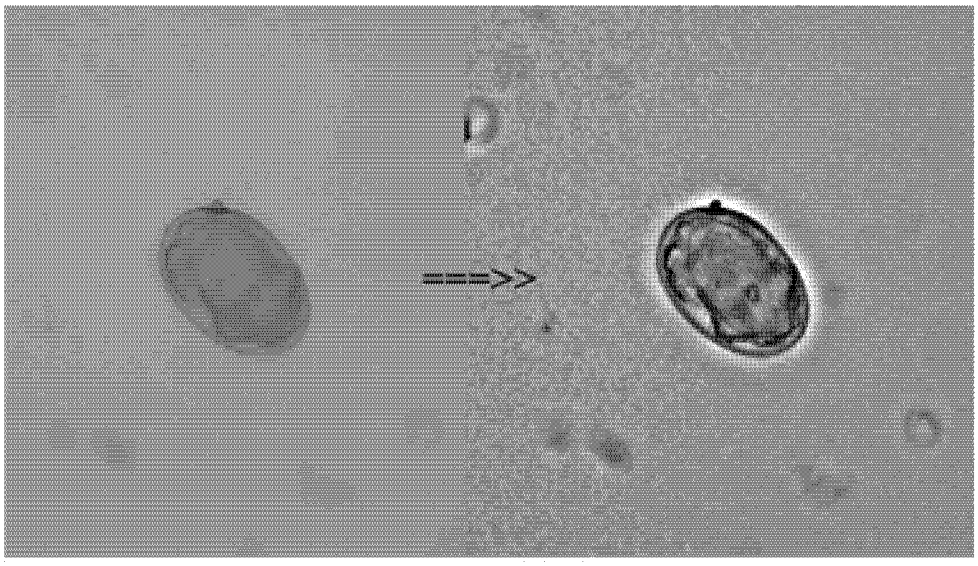

===>>

图 7

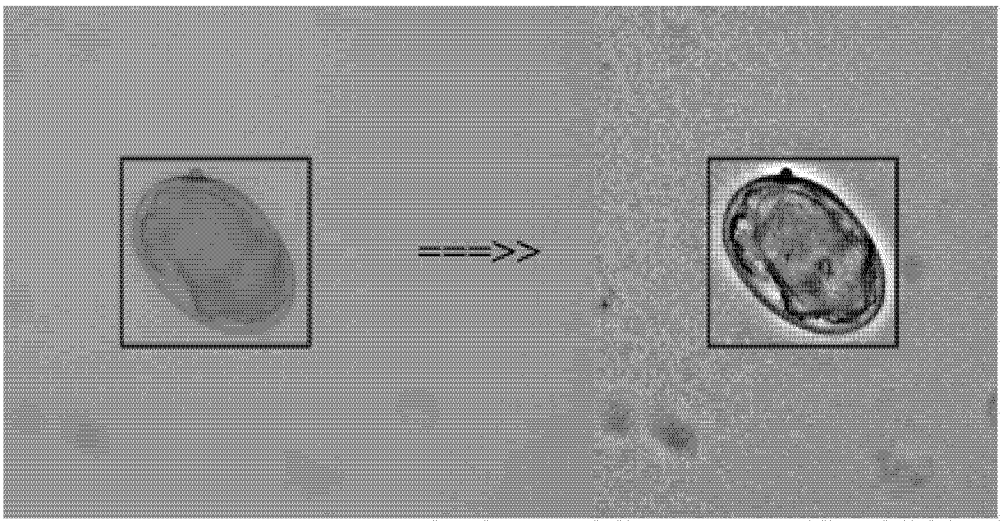


图 8

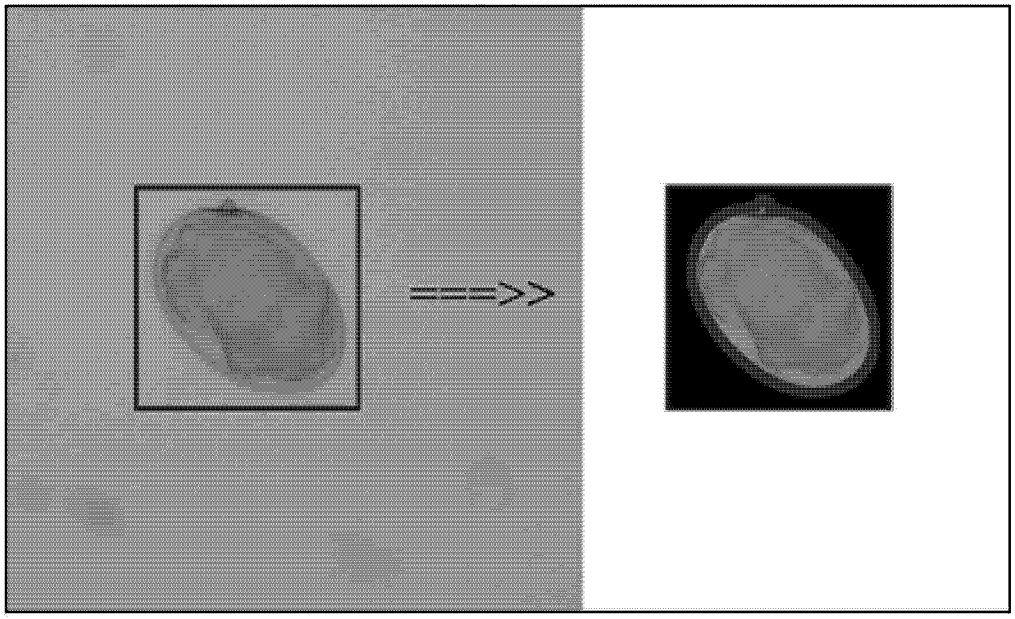


图 9

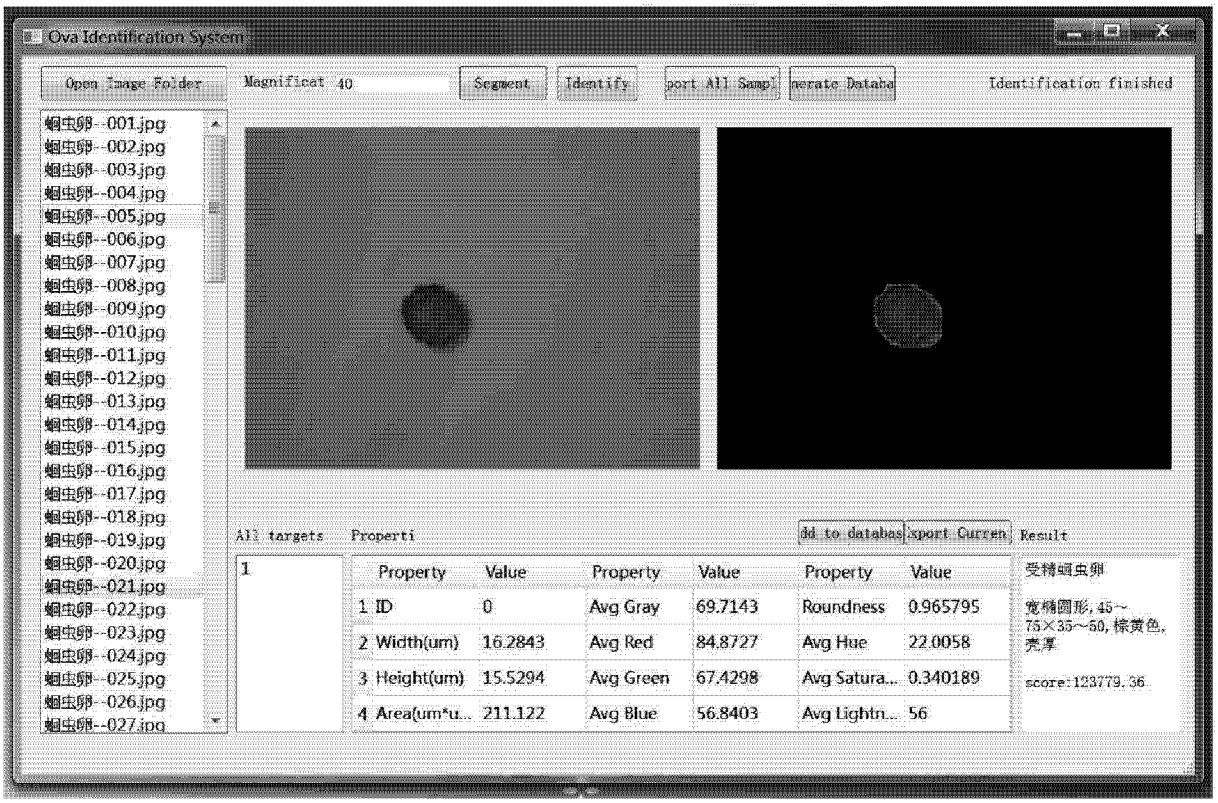

Ova Identification System
Open Image Folder
Magnificat 40
Segment
Identify
port All Sampl
nerate Databa
Identification finished
蛔虫卵--001.jpg
蛔虫卵--002.jpg
蛔虫卵--003.jpg
蛔虫卵--004.jpg
蛔虫卵--005.jpg
蛔虫卵--006.jpg
蛔虫卵--007.jpg
蛔虫卵--008.jpg
蛔虫卵--009.jpg
蛔虫卵--010.jpg
蛔虫卵--011.jpg
蛔虫卵--012.jpg
蛔虫卵--013.jpg
蛔虫卵--014.jpg
蛔虫卵--015.jpg
蛔虫卵--016.jpg
蛔虫卵--017.jpg
蛔虫卵--018.jpg
蛔虫卵--019.jpg
蛔虫卵--020.jpg
蛔虫卵--021.jpg
蛔虫卵--022.jpg
蛔虫卵--023.jpg
蛔虫卵--024.jpg
蛔虫卵--025.jpg
蛔虫卵--026.jpg
蛔虫卵--027.jpg
All targets
Properti
dd to databas
xport Curren
Result
1
Property Value Property Value Property Value
1 ID 0 Avg Gray 69.7143 Roundness 0.965795
2 Width(um) 16.2843 Avg Red 84.8727 Avg Hue 22.0058
3 Height(um) 15.5294 Avg Green 67.4298 Avg Satura... 0.340189
4 Area(um*u... 211.122 Avg Blue 56.8403 Avg Lightn... 56
受精蛔虫卵
宽椭圆形，45～75×35～50，棕黄色，壳厚
score:123779.36

图 10