

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

H04L 9/00 (2006.01)

G06T 7/00 (2006.01)

G06T 5/00 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200710130809.6

[43] 公开日 2008 年 3 月 19 日

[11] 公开号 CN 101145902A

[22] 申请日 2007.8.17

[21] 申请号 200710130809.6

[71] 申请人 东南大学

地址 210096 江苏省南京市四牌楼 2 号

[72] 发明人 曹玖新 罗军舟 毛 波

[74] 专利代理机构 南京经纬专利商标代理有限公司

代理人 陆志斌

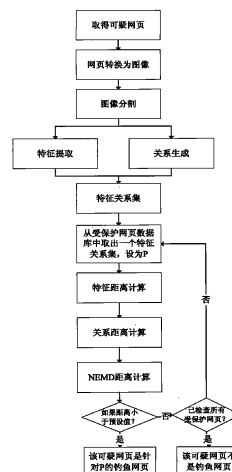
权利要求书 3 页 说明书 11 页 附图 5 页

[54] 发明名称

基于图像处理的钓鱼网页检测方法

[57] 摘要

本发明涉及图像信息处理以及网络安全领域，尤其涉及一种基于图像分割的钓鱼网页检测方法。本发明采用逐一计算可疑钓鱼网页与各个受保护网页之间的相似度，将可疑网页与受保护网页之间的相似度小于预先设定的值的可疑网页判为针对该受保护网页的钓鱼网页。本发明的特点是直接分析用户所访问的网页图像，可靠性高，不易被网页设计技巧所欺骗；通过对网页图片进行分割可以进行部分匹配，防止由于部分相似而引起的网络钓鱼；使用子图像特征和关系并重的匹配方法，更加符合视觉心理学，使匹配结果更加准确。根据试验测得的数据显示，本发明具有网页图像检测准确性高的优点。



1、一种基于图像处理的钓鱼网页检测方法，其特征在于方法具体如下：
逐一计算可疑钓鱼网页与各个受保护网页之间的相似度，将可疑网页与受保护网页之间的相似度小于预先设定的值的可疑网页判为针对该受保护网页的钓鱼网页；

上述网络钓鱼为：通过仿造网上金融或其他网络服务提供商的相关页面，诱导用户输入敏感信息，并非法取得该敏感信息的手段；

上述可疑钓鱼网页为：从可疑邮件中检测出的链接地址所指向的网页，或由用户上报的钓鱼网页；

上述受保护网页为：网络钓鱼所伪造的对象网页；

上述计算网页相似度的方法为：

步骤 1，将网页转换为图像；

步骤 2，对所得图像进行分割；

步骤 3，提取其分割后子图像的图像特征；

步骤 4，生成子图像之间的相对位置关系；

步骤 5，将子图像的图像特征和子图像的相对位置关系组成该页面的特征关系集合；

步骤 6，计算两个网页的特征关系集合之间的距离并以此作为这两个网页的相似度，两个网页之间的距离小于 0.02 的两个网页为相似网页；

其中步骤 4 中子图像之间的相对位置关系由一个 9 维向量表示，该 9 维向量的每个分量表示由任一子图像所确定的 9 个象限中的每个象限与另一子图像之间的位置关系，设用于确定 9 个象限的任一子图像为矩形，延长该矩形的四条边将把图像平面分为 9 个部分，从左上角开始顺时针编号 1~8，中心编号为 9，由此形成第 1~9 象限，该子图像与另一子图像之间相对位置关系的求取方法为：只需判断另一子图像在该子图像所分割的 9 个部分中的分

布状况，将出现全部或部分另一子图像的象限记为 1，则与该象限对应的第 i 维分量就为 1，否则为 0，根据上述方法，求出由关系矩阵表示的原网页图像所有子图像之间的相对位置关系。

2、根据权利要求 1 所述的基于图像处理的钓鱼网页检测方法，其特征是：图像分割方法由以下步骤组成：

步骤 1，预处理：预处理是采用通用的坎尼算子边界检测方法将网页图像转换为初始二值图像；

步骤 2，将预处理所得的二值图像设为初始子图像，置于待分割子图像集合中（初始该集合中只有初始子图像）；

步骤 3，判断待分割子图像集合是否为空，如果为空则结束，否则进入步骤 4；

步骤 4，从待分割子图像集合中任取某个子图像对其进行收缩处理，得到有效子图像；

步骤 5，判断该有效子图像是否可分割，如果可分割，检测出该子图像中面积最大的分割带，并以此将该子图像一分为二，同时把分割后的两个子图像存入待分割子图像集合中，如果不能分割则将其存入以分割子图像集合中；

步骤 6，转步骤 3。

3、根据权利要求 1 所述的这种基于图像处理的钓鱼网页检测方法，其特征是：步骤 3 所述的图像特征包括彩色信息，灰度信息以及长度和宽度，彩色信息和灰度信息采用直方图统计的方法；子图像长度和宽度直接由分割结果得到，

上述彩色信息采用直方图获取方法：将原图转换到色度纯度亮度（HSV）颜色空间并将其非均匀量化为 32 种颜色，在此基础上提取彩色直方图，并对其进行归一化处理；灰度信息采用直方图获取方法：将原图转换为黑白图像，

并对其进行拉伸，增强对比度，再将整个灰度空间量化为 32 个灰度，在此基础上提取灰度直方图，并对其进行归一化处理。

4、根据权利要求 3 所述的这种基于图像处理的钓鱼网页检测方法，其特征是：计算两个网页的特征关系集合之间相似度的方法采用嵌套运土者距离方法，该嵌套运土者距离方法为：确定步骤 5 所述的特征关系集合中的子图像的图像特征的相似度，所述子图像的图像特征相似度的确定方法采用综合加权距离，该综合加权距离包括归一化彩色直方图相似度、归一化灰度直方图相似度以及长宽相似度，其权值各为 1/3；确定步骤 5 所述的特征关系集合中的子图像的相对位置关系的相似度，所述子图像的相对位置关系的相似度的确定方法采用归一化的运土者距离；最后根据子图像的图像特征相似度和子图像的相对位置关系相似度，采用嵌套运土者距离计算两个特征关系集合之间的相似度，所述嵌套运土者距离的计算方法为：首先求出所有子图像两两之间的内部距离矩阵，该内部距离矩阵的大小为 $M \times N$ ，其中 M 、 N 为分别为两个待检测网页中的子图像的个数，内部距离矩阵的元素由前述子图像的图像特征相似度和子图像的相对位置关系相似度给出，根据该内部距离矩阵，采用运土者距离算法求出子图像之间的外部距离，所有子图像之间的外部距离构成一个外部距离矩阵，大小为 $M \times N$ ，其中 M 、 N 为分别为两个待检测网页中的子图像的个数，基于该外部距离矩阵，并将两个待检测网页中所有子图像的权值设为相等，再次使用运土者距离算法求出两个待检测网页的距离，该距离即为网页的嵌套运土者距离。

基于图像处理的钓鱼网页检测方法

技术领域

本发明涉及图像信息处理以及网络安全领域，涉及一种反网络钓鱼的方法，尤其涉及一种基于图像处理的钓鱼网页检测方法。

背景技术

网络钓鱼（Web Phishing）是一种在线身份冒充，它综合使用社会工程学（即骗术）和技术手段（包括恶意代码、钓鱼邮件、监听软件等），以窃取网络用户的个人身份信息或帐号、密码、信用卡号等。网络钓鱼目前已经成为用户信息安全的极大威胁，并给用户造成了巨大的经济损失（据权威机构预测美国 2007 年因网络钓鱼引起的损失达到 28 亿美元）；同时由于网络钓鱼的肆虐，恶化了网络生态环境，严重阻碍了我国网上银行、在线支付等电子金融的发展。

目前常用对网络钓鱼的防范主要包括以下三种措施。

一、电子邮件检测。该方法通过在邮件服务器上设置邮件过滤策略，通过检测邮件中的超链接，内嵌脚本中域名的有效性和一致性来发现钓鱼网页。

二、个人信息保护。该方法通过检测用户发出的所有信息的目的地址，以及所发信息的机密程度来判断此次发送的安全性，如果发现不安全的发送便会阻断该发送或向用户询问是否有威胁的情况下继续发送。

三、网络行为检测。该方法通过模拟用户的提交行为，向网页发送预先构造的用户名和密码，如果服务器回应提交成功，则证明该网页为钓鱼网页。

但以上三种方法都存在很大的不足，方法一无法检测包含图像的邮件，包括使用图片和映射的方法提供的链接或以图片显示的地址等；方法二则会

大幅降低网络速度，同时规则的维护也极为复杂，而且涉及用户的隐私，因此很难部署；方法三无法应对桥接攻击，同时简单的自动提交防范机制（验证码）就能使该方失效。

发明内容

本发明针对现有技术的不足和缺陷，提供一种准确性高、可靠性强且能够减小被欺骗可能性的基于图像处理的钓鱼网页检测方法。

本发明采用如下技术方案：

逐一计算可疑钓鱼网页与各个受保护网页之间的相似度，将可疑网页与受保护网页之间的相似度小于预先设定的值的可疑网页判为针对该受保护网页的钓鱼网页；

上述网络钓鱼为：通过仿造网上金融或其他网络服务提供商的相关页面，诱导用户输入敏感信息，并非法取得该敏感信息的手段；

上述可疑钓鱼网页为：从可疑邮件中检测出的链接地址所指向的网页，或由用户上报的钓鱼网页；

上述受保护网页为：网络钓鱼所伪造的对象网页；

上述计算网页相似度的方法为：

步骤 1，将网页转换为图像；

步骤 2，对所得图像进行分割；

步骤 3，提取其分割后子图像的图像特征；

步骤 4，生成子图像之间的相对位置关系；

步骤 5，将子图像的图像特征和子图像的相对位置关系组成该页面的特征关系集合；

步骤 6，计算两个网页的特征关系集合之间的距离并以此作为这两个网页的相似度，两个网页之间的距离小于 0.02 的两个网页为相似网页；

其中步骤 4 中子图像之间的相对位置关系由一个 9 维向量表示，该 9 维向

量的每个分量表示由任一子图像所确定的 9 个象限中的每个象限与另一子图像之间的位置关系，设用于确定 9 个象限的任一子图像为矩形，延长该矩形的四条边将把图像平面分为 9 个部分，从左上角开始顺时针编号 1~8，中心编号为 9，由此形成第 1~9 象限，该子图像与另一子图像之间相对位置关系的求取方法为：只需判断另一子图像在该子图像所分割的 9 个部分中的分布状况，将出现全部或部分另一子图像的象限记为 1，则与该象限对应的第 i 维分量就为 1，否则为 0，根据上述方法，求出由关系矩阵表示的原网页图像所有子图像之间的相对位置关系。本发明通过将网页转化为图像，并对所图像进行分割、特征提取、子图像关系检测得到特征关系集合，再求出可疑网页与受保护网页之间特征关系集合的 NEMD 距离，从而发现可疑网页与受保护网页之间的相似度，并根据该相似度判断可疑网页是否是钓鱼网页。如果可疑网页与受保护网页及其相似，则该可疑网页已经侵犯了受保护网页的“肖像权”，即使该可疑网页并非钓鱼网页也会误导用户并引发不必要的危害。

本发明的特点是直接分析用户所访问的网页图像，可靠性高，不易被网页设计技巧所欺骗；通过对网页图片进行分割可以进行部分匹配，防止由于部分相似而引起的网络钓鱼；使用子图像特征和关系并重的匹配方法，更加符合视觉心理学，使匹配结果更加准确。根据试验测得的数据显示，本发明具有网页图像检测准确性高的优点。

附图说明

图 1 钓鱼网页检测流程图。

图 2 为本发明所述分割方法的流程图。

图 3 为本发明所述系统的结构示意图。

图 4 为本发明所述的子区域关系示意图。

图 5 为本发明与国际现有先进方法的比较。

图 6 为本发明与国际现有先进方法的平均比较。

具体实施方式

基于图像处理的钓鱼网页检测方法，具体如下：逐一计算可疑钓鱼网页与各个受保护网页之间的相似度，将可疑网页与受保护网页之间的相似度小于预先设定的值的可疑网页判为针对该受保护网页的钓鱼网页；

上述网络钓鱼为：通过仿造网上金融或其他网络服务提供商的相关页面，诱导用户输入敏感信息，并非法取得该敏感信息的手段；

上述可疑钓鱼网页为：从可疑邮件中检测出的链接地址所指向的网页，或由用户上报的钓鱼网页；

上述受保护网页为：网络钓鱼所伪造的对象网页；

上述计算网页相似度的方法为：

步骤 1，将网页转换为图像；

步骤 2，对所得图像进行分割；

步骤 3，提取其分割后子图像的图像特征；

步骤 4，生成子图像之间的相对位置关系；

步骤 5，将子图像的图像特征和子图像的相对位置关系组成该页面的特征关系集合；

步骤 6，计算两个网页的特征关系集合之间的距离并以此作为这两个网页的相似度，两个网页之间的距离小于 0.02 的两个网页为相似网页；

其中步骤 4 中子图像之间的相对位置关系由一个 9 维向量表示，该 9 维向量的每个分量表示由任一子图像所确定的 9 个象限中的每个象限与另一子图像之间的位置关系，设用于确定 9 个象限的任一子图像为矩形，延长该矩形的四条边将把图像平面分为 9 个部分，从左上角开始顺时针编号 1~8，中心编号为 9，由此形成第 1~9 象限，该子图像与另一子图像之间相对位置关系的求取方法为：只需判断另一子图像在该子图像所分割的 9 个部分中的分布状况，将出现全部或部分另一子图像的象限记为 1，则与该象限对应的第 i 维分量就为 1，否则为 0，根据上述方法，求出由关系矩阵表示的原网页图像

所有子图像之间的相对位置关系。

上述图像分割方法由以下步骤组成：

步骤 1，预处理：预处理是采用通用的坎尼算子边界检测方法将网页图像转换为初始二值图像；

步骤 2，将预处理所得的二值图像设为初始子图像，置于待分割子图像集合中（初始该集合中只有初始子图像）；

步骤 3，判断待分割子图像集合是否为空，如果为空则结束，否则进入步骤 4；

步骤 4，从待分割子图像集合中任取某个子图像对其进行收缩处理，得到有效子图像；

步骤 5，判断该有效子图像是否可分割，如果可分割，检测出该子图像中面积最大的分割带，并以此将该子图像一分为二，同时把分割后的两个子图像存入待分割子图像集合中，如果不能分割则将其存入以分割子图像集合中；

步骤 6，转步骤 3。

步骤 3 所述的图像特征包括彩色信息，灰度信息以及长度和宽度，彩色信息和灰度信息采用直方图统计的方法；子图像长度和宽度直接由分割结果得到，

上述彩色信息采用直方图获取方法：将原图转换到色度纯度亮度（HSV）颜色空间并将其非均匀量化为 32 种颜色，在此基础上提取彩色直方图，并对其进行归一化处理；灰度信息采用直方图获取方法：将原图转换为黑白图像，并对其进行拉伸，增强对比度，再将整个灰度空间量化为 32 个灰度，在此基础上提取灰度直方图，并对其进行归一化处理。

上述计算两个网页的特征关系集合之间相似度的方法采用嵌套运土者距离方法，该嵌套运土者距离方法为：确定步骤 5 所述的特征关系集合中的子图像的图像特征的相似度，所述子图像的图像特征相似度的确定方法采用综

合加权距离，该综合加权距离包括归一化彩色直方图相似度、归一化灰度直方图相似度以及长宽相似度，其权值各为 1/3；确定步骤 5 所述的特征关系集合中的子图像的相对位置关系的相似度，所述子图像的相对位置关系的相似度的确定方法采用归一化的运土者距离；最后根据子图像的图像特征相似度和子图像的相对位置关系相似度，采用嵌套运土者距离计算两个特征关系集合之间的相似度，所述嵌套运土者距离的计算方法为：首先求出所有子图像两两之间的内部距离矩阵，该内部距离矩阵的大小为 $M \times N$ ，其中 M 、 N 为分别为两个待检测网页中的子图像的个数，内部距离矩阵的元素由前述子图像的图像特征相似度和子图像的相对位置关系相似度给出，根据该内部距离矩阵，采用运土者距离算法求出子图像之间的外部距离，所有子图像之间的外部距离构成一个外部距离矩阵，大小为 $M \times N$ ，其中 M 、 N 为分别为两个待检测网页中的子图像的个数，基于该外部距离矩阵，并将两个待检测网页中所有子图像的权值设为相等，再次使用运土者距离算法求出两个待检测网页的距离，该距离即为网页的嵌套运土者距离。

下面参照附图，对本发明的具体实施方式加以详细描述：

根据网页图像的特征关系集合，逐一计算可疑钓鱼网页与各个受保护网页之间的相似度，将可疑网页与受保护网页之间的相似度小于预先设定的值的可疑网页判为针对该受保护网页的钓鱼网页；受保护网页的特征关系集合可以预先计算出并存储于数据库中，这样可以提高检索效率；

上述网络钓鱼为：通过仿造网上金融或其他网络服务提供商的相关页面，诱导用户输入敏感信息，并非法取得该敏感信息的手段；

上述可疑钓鱼网页为：从可疑邮件中检测出的链接地址所指向的网页，或由用户上报的钓鱼网页；

上述受保护网页为：网络钓鱼所伪造的对象网页；

根据国际反网络钓鱼工作组（Anti-Phishing Work Group APWG）的统计，80%的网络钓鱼针对 11 家网络金融服务提供商，因此受保护网页数据库不会

十分庞大。同时通过电子邮件检测、网络流量捕获，用户报告等方式各个网络钓鱼分析节点获取疑似钓鱼网站地址。

如图 1 所示，上述计算网页相似度的方法为：

步骤 1，将网页转换为图像；根据需求不同可以使用 Java 语言中的容器，或商业软件如 HTML2JPG 完成。该转换以网页地址（URL）为输入，取得网页内容后将其自动转换为设定格式的图片（如 JPEG 或 BMP 等）并保存在预先制定的位置；

步骤 2，对所得图像进行分割；对步骤 1 所得的待检测网页的图像进行分割，将其分成包含网页中各个特征块的子图像。如图 2 所示，具体步骤如下：

步骤 2.1，预处理。预处理是指采用通用的坎尼算子边界检测方法将网页图像转换为初始二值图像；

步骤 2.2，将预处理所得的二值图像设为初始子图像，置于待分割子图像集合中（初始该集合中只有初始子图像）；

步骤 2.3，判断待分割子图像集合是否为空，如果为空则结束分割并转步骤 3，否则进入步骤 2.4；

步骤 2.4，从待分割子图像集合中任取某个子图像对其进行收缩处理，得到有效子图像；

步骤 2.5，判断该有效子图像是否可分割，如果可分割，检测出该子图像中面积最大的分割带，并以此将该子图像一分为二，同时把分割后的两个子图像存入待分割子图像集合中，如果不能分割则将其存入以分割子图像集合中；

步骤 2.6，转步骤 2.3。

步骤 3，提取分割后子图像的图像特征；在子图像特征提取方面，特征包括彩色信息，灰度信息以及度长和宽度，彩色信息和灰度信息采用直方图统计的方法；子图像长度和宽度直接由分割结果得到。具体而言，彩色信息直方图的获取方法为，将原图转换到 HSV 颜色空间并将其非均匀量化为 32 种颜

色[见参考文献 1]，在此基础上提取彩色直方图，并对其进行归一化处理。灰度信息直方图获取方法为，将原图转换为黑白图像，并对其进行拉伸，增强对比度，再将整个灰度空间量化为 32 个灰度，在此基础上提取灰度直方图，并对其进行归一化处理。步骤 4，生成子图像之间的相对位置关系；

步骤 4，生成子图像之间的相对位置关系；子图像之间的相对位置关系由一个 9 维向量表示，该 9 维向量的每个分量表示由一个子图像所确定的 9 个象限中另一子图像的分布状况。根据分割的结果，每个子图像都为矩形，延长该矩形的四条边将把平面分为 9 个部分，从左上角开始顺时针编号 1~8，中心编号为 9，如图 4 所示。求该子图像与另一子图像之间的关系时，只需判断另一子图像在该子图像所分割的 9 个部分中的分布状况，如果另一子图像出现在该子图像所确定的第 i 个象限中，则该 9 维向量的第 i 维为 1，否则为 0。例如图 4 中，子图像 A 与 B 的关系 $R(A, B) = \{0, 0, 0, 0, 1, 1, 0, 0, 0\}$ ，其中第五和第六个分量为 1，表示 B 在由 A 所分割的平面上的第五和第六个区域。根据以上方法，可以求出原网页图像所有子图之间的相对位置关系（由关系矩阵表示）。

步骤 5，子图的特征和关系组成该页面的特征关系集合；

步骤 6，使用嵌套运土者距离方法计算两个网页的特征关系集合之间的距离并以此作为这两个网页的相似度；在特征关系集合之间的匹配方面，使用综合加权相似度描述特征之间的相似程度，该综合加权距离包括归一化彩色直方图相似度，归一化灰度直方图相似度以及长宽相似度，其权值各为 1/3；使用归一化的运土者距离（EMD Earth Mover's Distance）来计算关系相似度；最后根据特征和关系距离并使用嵌套运土者距离（NEMD）来计算两个特征关系集之间的相似度。

具体而言，归一化彩色直方图和灰度直方图的相似度为对应的相同颜色或灰度的最小值相加，这样 1 表示完全相似，0 表示完全不同。长宽相似度为较小的长和宽的乘积比较大的长和宽的乘积，同样，0 为完全不同，1 为完全相

似。这样再对以上三个相似度进行加权求和，将所得的相似度作为两个子图像的特征相似度，再用 1 减去该特征相似度得到子图像特征的距离（0 为完全相似，1 为完全不同）。

关系距离使用 EMD 计算，其中关系由 9 维向量表示，并以此作为 EMD 算法中的特征向量，再根据预设的每个分量之间的距离生成距离矩阵，该预设距离可以根据具体应用进行设定，例如可以采用曼哈顿距离，或欧拉距离等。

NEMD 首先根据输入的两个子图像特征关系集合及其对应的特征和关系之间的距离，求出内部 EMD，再根据内部 EMD 求出两个子图像之间的距离；求出所有子图像之间的距离，组成外部 EMD，并由该外部 EMD 求出两个子图像特征关系集合的距离，并以此判断这两个集合的相似度，从而得到其对应网页的相似度。

实施实例。如图 3 所示，将本发明方法部署于独立的钓鱼分析节点，并使用网络钓鱼控制中心保存和维护受保护网页信息，其数据分布与域名服务（DNS）相似，由中心节点——网络钓鱼控制中心负责受保护数据库的更新维护，各个分析节点从控制中心取得受保护网页，并建立本地缓存机制，定期更新。为了进行大规模的钓鱼网页检测，在各个邮件服务器端的垃圾邮件检测规则中增加对钓鱼邮件检测的相关配置，对钓鱼邮件进行初步处理，并将可疑钓鱼邮件交付网络钓鱼分析结点，各个分析结点可以位于邮件服务其或独立的分析服务器上，并通过网络与钓鱼控制中心相连，分析服务器主要对可疑邮件中的网页进行处理并提取出可疑网页的特征关系集合，在与从钓鱼控制中心得到的受保护网页的特征关系集合进行匹配，以发现钓鱼网页，如果发现钓鱼网页则通知钓鱼控制中心，中心汇总各个网络钓鱼分析结点所提供的数据建立钓鱼网站黑名单，并通知所有分析结点和受害网站，受害网站可以警告其用户，或向公安机关报案，公安机关则可据此取缔钓鱼网站或钓鱼邮件发送主机并追究相关责任人的法律责任。

发明特点。基于本发明方法开发了原型系统，包括网页转换为图像，图像

预处理，网页图像分割，特征关系提取，和特征关系集匹配等。从实施效果来看，本发明能在合理的时间内，准确的判断两个网页的相似程度，抗干扰性好，分析结果可靠准确，超过目前已有的钓鱼网页分析方法。此外由于引入了图像的分割使得本发明可以匹配部分网页，大大增加了本方法的有效性。

在对7个钓鱼网页的检测中，本发明方法体现出明显的优势，具体数据如下：

表1 Yu 的钓鱼网页检测方法

	t-eBay	t-Earth Link	t-ICBC	t-Wells Fargo	t-US Bank	t-Wash ington
f-eBay1	0.0041	0.0292	0.065	0.0432	0.0196	0.0256
f-eBay2	0.0048	0.0294	0.0643	0.0434	0.0203	0.0249
f-EarthLink	0.0187	0.0293	0.0609	0.0561	0.0248	0.0143
f-ICBC	0.0591	0.0633	0.003	0.0664	0.0566	0.0589
f-WellsFargo	0.0424	0.0571	0.0672	0.0121	0.0419	0.0559
f-US Bank	0.0172	0.0240	0.0596	0.0413	0.0017	0.0228
f-Washington	0.0293	0.0231	0.0597	0.0614	0.0299	0.0095

表2 本发明钓鱼网页检测方法

	t-eBay	t-Earth Link	t-ICBC	t-Wells Fargo	t-US Bank	t-Wash ington
f-eBay1	0.0151	0.2044	0.3483	0.1472	0.3458	0.2383
f-eBay2	0.0032	0.2051	0.3232	0.1452	0.3395	0.2405
f-EarthLink	0.1985	0.1989	0.4257	0.0820	0.3490	0.2449
f-ICBC	0.3219	0.4168	0.0010	0.4599	0.2155	0.4210
f-WellsFargo	0.1414	0.1343	0.4516	0.0135	0.2706	0.1685
f-US Bank	0.3370	0.3393	0.2153	0.2720	0.0052	0.3354
f-Washington	0.2470	0.2642	0.4280	0.1777	0.3387	0.0125

表 1, 2 中分别显示了 Yu 检测方法与本发明方法对钓鱼网页（以 f-开头）及其所针对的真实网页（以 t-开头）之间相似度的计算结果。其中针对 EarthLink 的钓鱼网页与真实网页有很大的区别，所以两种方法都无法检测。根据表 1 和 2 的结果，求出了最坏情况对比度（图 5 所示）和平均对比度（图 6 所示），从图 5 和图 6 的对比度来看本发明方法明显优于 Yu 的方法。

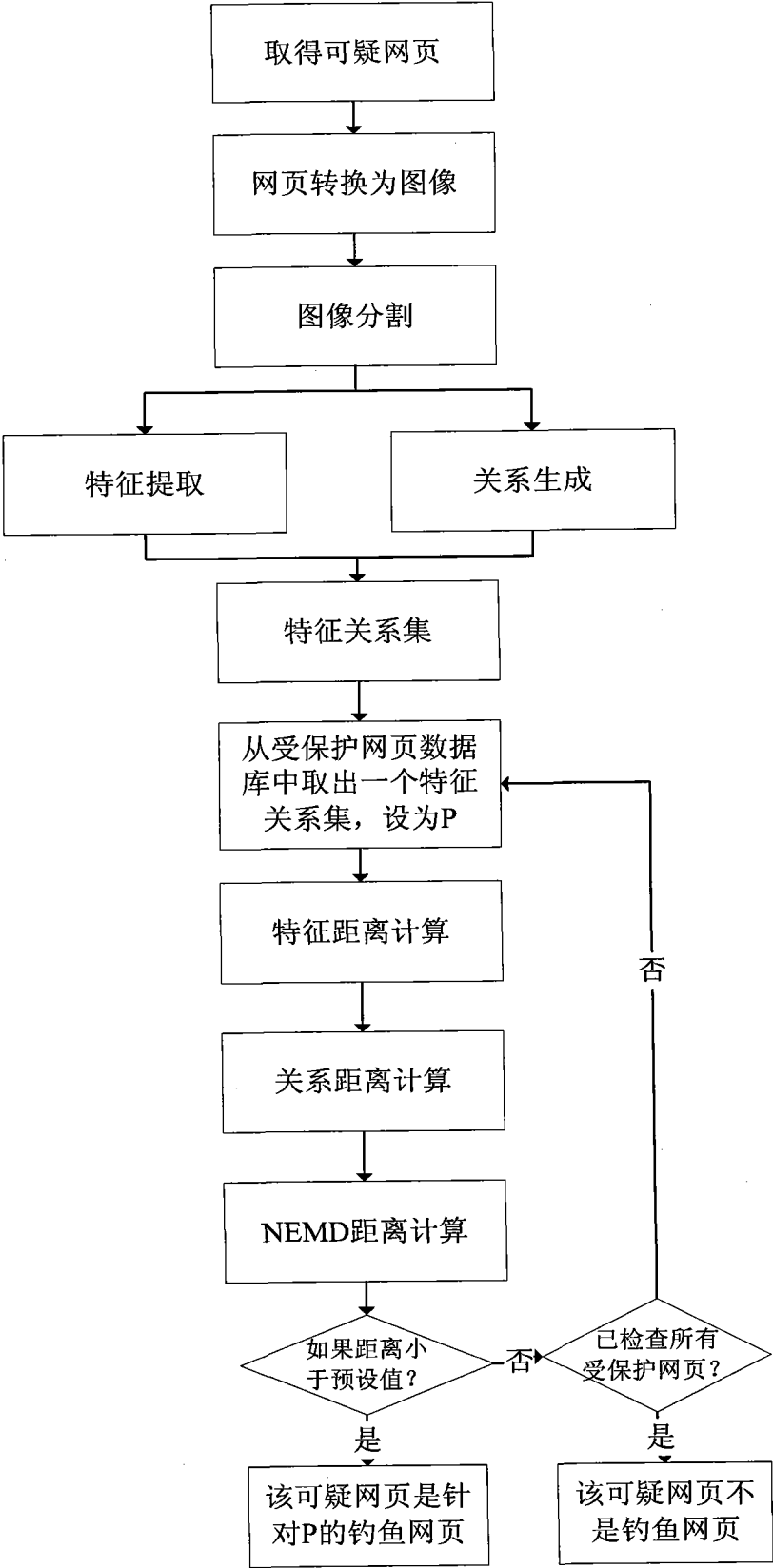


图 1

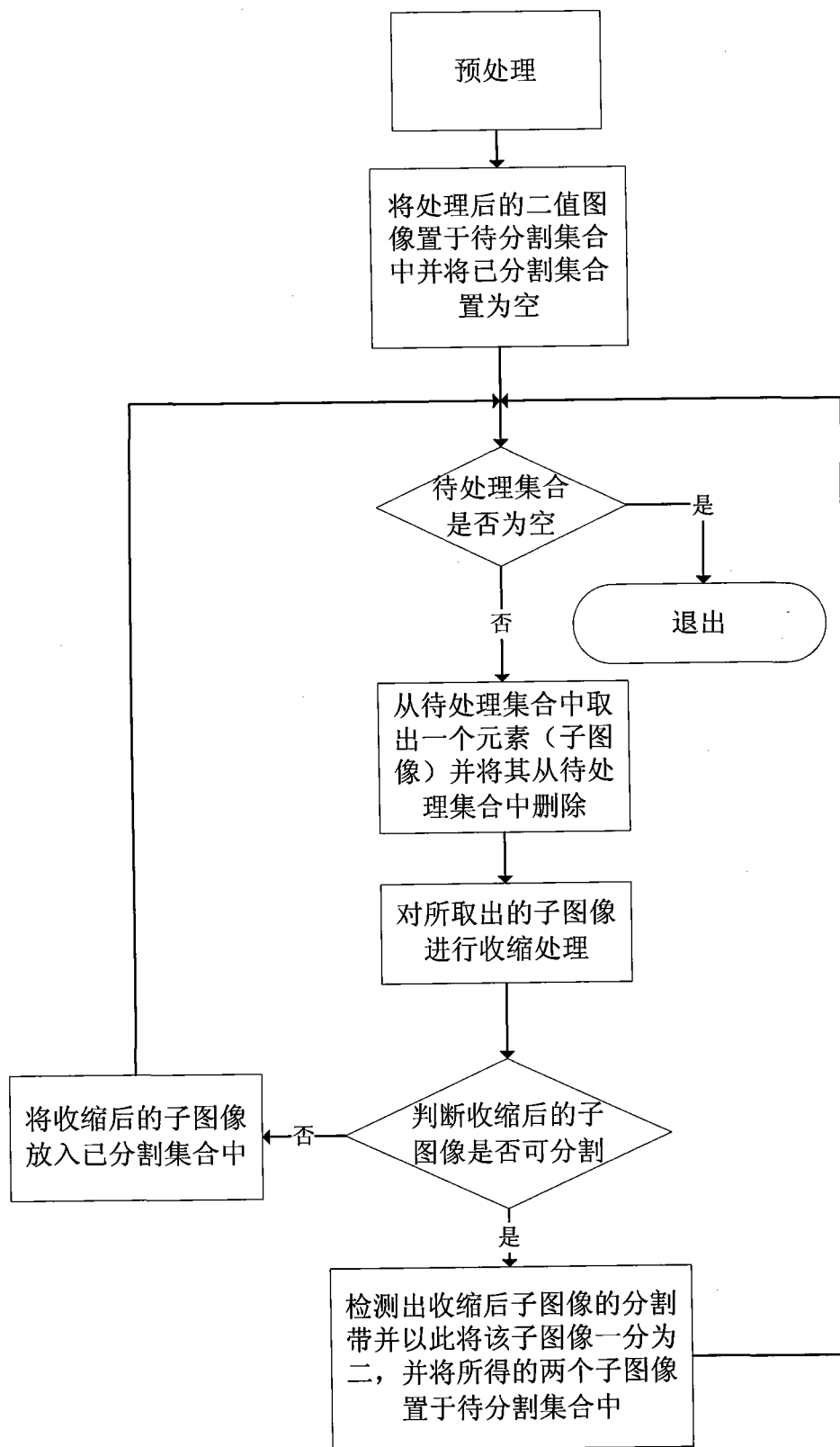
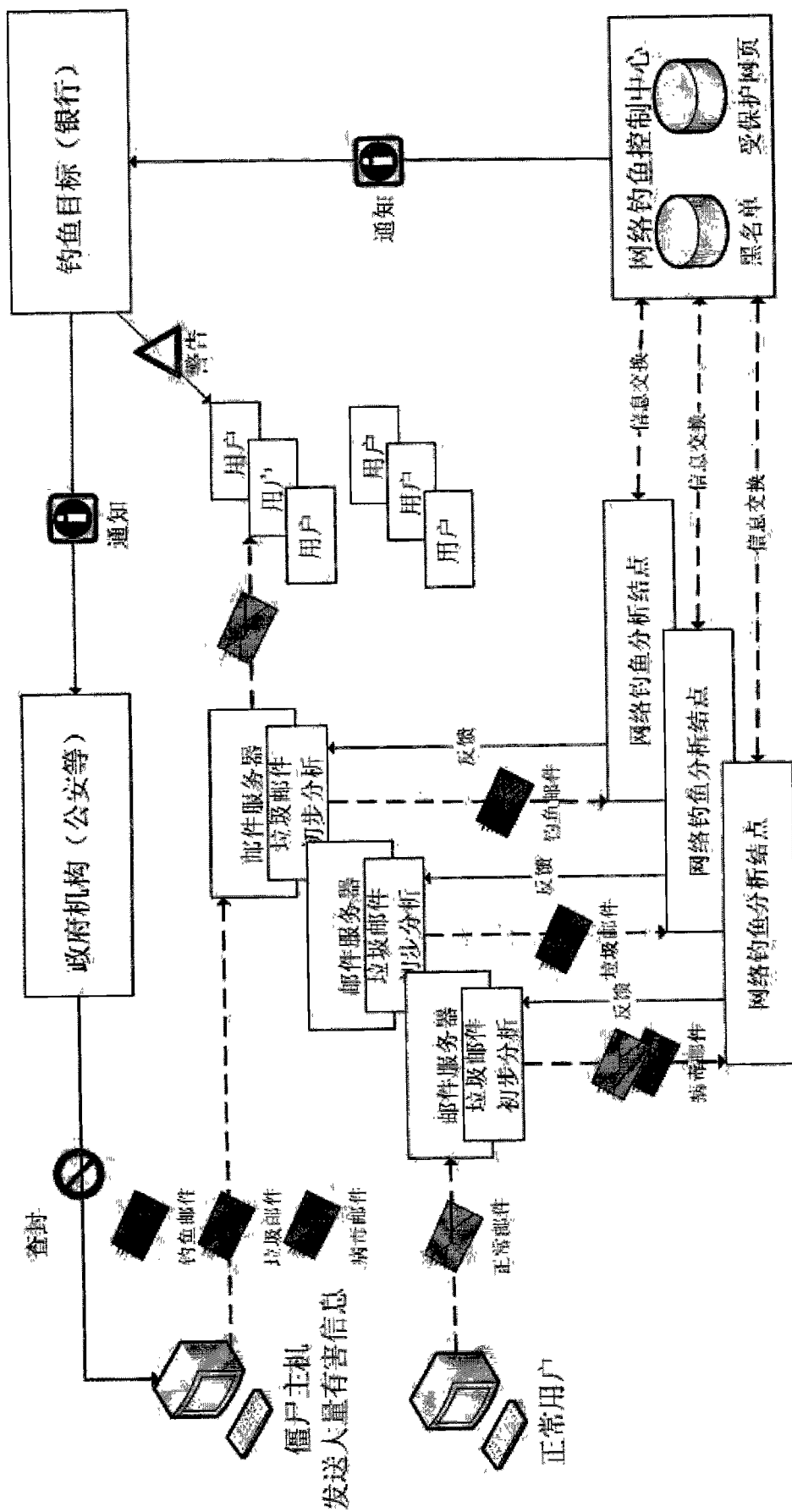


图 2

3

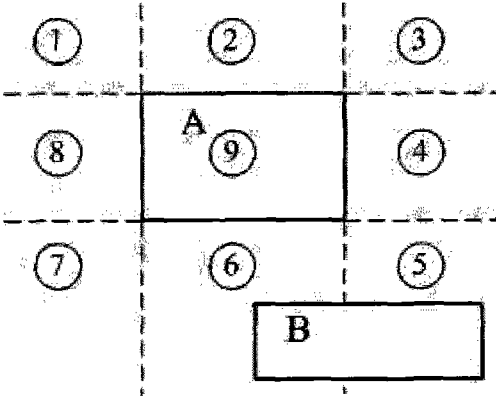



图 4

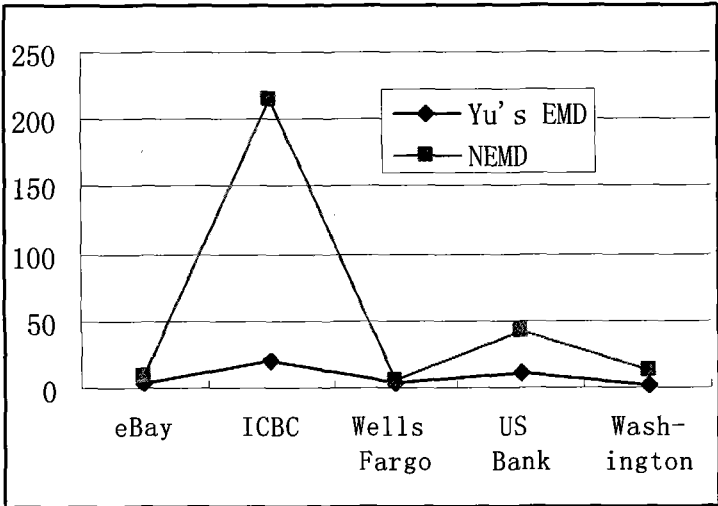


图 5

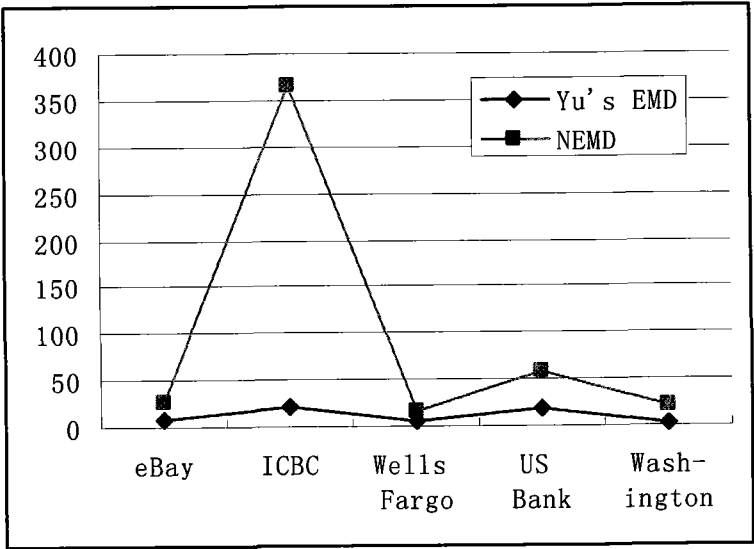


图 6