

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200810105224.3

[51] Int. Cl.

G06F 19/00 (2006.01)

G06K 9/20 (2006.01)

G09B 7/00 (2006.01)

[43] 公开日 2008年12月31日

[11] 公开号 CN 101334814A

[22] 申请日 2008.4.28

[21] 申请号 200810105224.3

[71] 申请人 华北电力大学

地址 102206 北京市昌平区朱辛庄华北电力大学

[72] 发明人 程文刚 吴克河 陈圣俭 柳长安
陆俊 廖斌 周登文 仲莉恩

[74] 专利代理机构 北京华谊知识产权代理有限公司

代理人 刘月娥

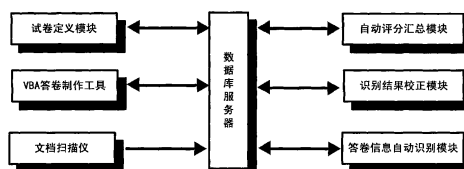
权利要求书 3 页 说明书 8 页 附图 5 页

[54] 发明名称

一种自动化的扫描阅卷系统及阅卷方法

[57] 摘要

一种自动化的扫描阅卷系统及方法，属于图像处理和模式识别技术领域。该系统由试卷定义模块、基于 VBA 的答卷制作工具、文档扫描仪、数据库服务器、答卷信息自动识别模块、答卷识别结果校正模块和自动评分汇总模块组成，整个自动阅卷系统是一个星形结构，数据库服务器是系统的中心，其他各个模块及设备都单独以网络的形式与其相连接，用于信息传递和网络共享；一种自动的扫描阅卷方法包含以下步骤：试卷定义、基于 VBA 的试卷制作、考试、主观题评阅、答卷扫描、答卷信息识别、识别结果校正和自动评分汇总。优点在于，本发明能判阅同时包含主观题和客观题的试卷，操作技术和设备要求低，对答卷的纸张质量、用笔和印刷要求低，校正方便。



1、一种自动化的扫描阅卷系统，其特征在于，由试卷定义模块、基于 VBA 的答卷制作工具、文档扫描仪、数据库服务器、答卷信息自动识别模块、答卷识别结果校正模块、自动评分汇总模块组成；整个自动阅卷系统是一个星形结构，数据库服务器是系统的中心，其他各个模块及设备都单独以网络的形式与其相连接，用于信息传递和网络共享。

2、按照权利要求 1 所述的系统，其特征在于，所述的试卷定义模块安装在用户的计算机上，用于试卷基本信息、结构信息以及答案与评分标准的定义，并把设置结果存储到数据库服务器中，为答卷制作提供信息。

3、按照权利要求 1 所述的系统，其特征在于，所述的基于 VBA 的答卷制作工具安装在用户的计算机上，用于生成答卷的电子模板，其输出结果转化为标准答卷图像和考试实际使用的答卷。

4、按照权利要求 1 所述的系统，其特征在于，所述的文档扫描仪安装在扫描人员的设备间，用来接收大量的经过主观题评阅的考生答卷，对其电子化并自动转存到数据库服务器中。

5、按照权利要求 1 所述的系统，其特征在于，所述的答卷信息自动识别模块安装在用户的计算机上，它从数据库服务器中获取考生答卷，之后进行自动识别，最后将识别结果返回给数据库服务器。

6、按照权利要求 1 所述的系统，其特征在于，所述的答卷信息校正模块，安装在用户的计算机上，从数据库服务器获取自动阅卷的结果信息并显示，供考务人员校正，并将校正结果返回存储到数据库服务器上。

7、按照权利要求 1 所述的系统，其特征在于，所述的自动评分汇总模块安装在用户的计算机上，从数据库获取识别校正的结果数据，并根据试卷定义存储的信息，进行自动的评分与汇总；打印机是系统的可选组件，与自动评分汇总模块所在的计算机相连接，用于打印输出成绩等统计数据。

8、一种采用权利要求 1 所述的自动化扫描阅卷系统阅卷的方法，其特征在于，包含以下步骤：

a. 试卷定义，定义试卷的结构，并进行答案与评分标准的设置，

试卷结构定义，是出题教师利用本系统的试卷定义模块，定义试卷的结构；首先，需要创建一份试卷的基本信息，包括满分值、考试年级和考试科目；其次，创建题目的数目、各个题目的主客观性质和各个题目的满分值，并把相应的信息存入数据库服

务器中;

答案与评分标准设置,是指利用本系统的试卷定义模块,在完成试卷结构的定义后,在系统中录入各个客观题的标准答案以及多选题的评分标准,而后试卷定义模块会将这些信息存入数据库服务器中;

b. 基于 VBA 的答卷制作,是指利用本系统提供的基于 VBA 的试卷制作工具,根据已定义的试卷结构制作出答卷的电子模板,并把答卷模版存入数据库服务器中;已经制作好的电子模板主要用于以下两个方面:一是生成标准答卷图像,二是作为考生的答卷;

c. 考试,是指考生根据考试安排参与考试的过程;考试之前,考务人员需要将考生的考号收集并存储到数据库服务器中;考试过程中,考生需要书写并填涂答卷中的考生信息区域、将客观题的答题结果填涂在答卷的客观题区域以及将主观题的答题结果书写在答卷的主观题答题区域;

d. 主观题评阅,是指阅卷人员对考生答卷中主观题评阅的过程,在评阅完成之后,阅卷人员将主观题的得分填涂在考生答卷的主观题分数涂卡区中;

e. 答卷扫描,是指利用文档扫描仪,将主观题评阅过的考生答卷成批扫描成一个 TIF 格式的图像文件,通常同一试室或班级的考生的答卷会被扫描成一个 TIF,所得的图像存入数据库服务器内,用于后续的考生答卷信息识别之用;

f. 答卷信息识别,是指利用本系统的答卷信息识别模块,识别出标准答卷和考生答卷中标记位、考号、客观题和主观题等区域的信息,并将识别信息存入到数据库服务器中;还包括:

标准答卷信息识别,是定位并记录标准答卷图像的各部分包括标记位、考号、客观题区域、各个主观题答题区域和各个主观题涂卡区域的坐标信息,之后答卷信息识别模块将这些坐标信息存储到数据库服务器中;

考生答卷信息识别,是识别出考生答卷的填涂信息;这是在标准答卷图像信息识别完成的基础上进行的,首先将倾斜的考生答卷旋转并校正到与标准答卷图像统一的坐标系上,然后根据位置标准答卷图像中各部分的坐标获取各个识别项的位置信息,以识别出考生答卷的填涂信息;

g. 答卷识别结果校正,是利用本系统的答卷识别结果校正模块,将错误的识别信息进行校正;对考生答卷图像分割成为考号区域图像、考生信息区域图像、客观题区域图像以及各个主观题区域图像;根据数据库服务器中的考号信息,自动判断已识别考号的有效性;包括考号是否识别完全、考号是否存在、考号是否重复出现等情况;

对于主观题，根据当前题目的满分值判断教师所判分数是否合理；针对各种不合理的
情况，给出友好明确的提示；答卷识别结果校正模块提供了各个切割图像以及相应的
客观题答案识别或主观题评分结果，阅卷人员据此实现校正，校正后的信息将存储到
数据库服务器中；

h. 自动评分汇总，是利用本系统的自动评分汇总模块，根据自动识别和校正后的
客观题答案与主观题评分识别结果，依据评分规则，自动评分汇总每份答卷的分数。

一种自动化的扫描阅卷系统及阅卷方法

技术领域

本发明属于图像处理和模式识别技术领域，特别是提供了一种自动化的扫描阅卷系统，利用数字图像处理和模式识别技术对答卷信息进行自动识别的系统和阅卷方法，尤其涉及一种用于包含客观题和主观题试卷的自动阅卷方法。

背景技术

考试是评估学生知识掌握程度的一种最常用、最普遍的手段，通过各种规模的考试有利于教师教学工作的推进、学生学习成绩的提高和整个教育水平的提升。客观题可以检测学生的判断能力，而主观题则可以反应出学生的逻辑推理、总结概括的能力。因此，一份既含有客观题又含有主观题的试卷具有相当的测试价值。

实现自动化扫描试卷的评阅是一项涉及多方面技术的问题，目前国内外在这方面都进行了一些研究和实践。应用最多的是光学标记阅读机(OMR, Optical Mark Reader)，它基于光电转换的功能原理，以快速、准确的性能使得其在国内外标准化考试的阅卷、评分等信息处理系统中得到普遍应用；但与此同时，它还存在着仅能评阅客观题、对答题卡纸张质量要求高、对信息卡印刷精度要求高、对填涂区灰度的浓度、颜色要求高以及需要专门的光标阅读机器等问题。另外一种在大型选拔考试（例如，高考）中广泛使用的是网上阅卷系统，它采用试卷和答卷分离的方式，主观题由不同阅卷教师通过网络在计算机上对考生答卷的电子图像分别进行评分，最终由计算机系统自动进行核分和成绩校验的一种阅卷方式。网上阅卷有着以下优点：它采用一卷多评制（主观题有多个老师进行评阅）和误差控制机制，有助于阅卷教师更好地把握评分标准、控制误差、提高阅卷质量；但同时存在着对阅卷环境、扫描设备、答题纸质等要求高，此外庞大的阅卷成本也使其难以在普通考试中推广应用。

我国基础教育阶段的考试有着频繁性的特点。为了能及时地检测学生对知识的掌握程度，促进学生知识水平的提高，除了常规的期中、期末考试外，还增加了月考、小测试等各种科目、各种规模的考试。与频繁的考试相对应的是，基础教育阶段的考试迫切需要快速、准确、小成本地阅卷。教师要在较短的时间内评阅试卷，时间之紧、工作量之大容易使得各个环节不可避免地出现人工误差，而且试卷登分、核分、保管等各项工作需要耗费人力、物力以及场地等资源。因此，为了能够满足我国基础教育阶段考试的需要，一种自动的扫描阅卷系统的出现就显得尤为重要了。

发明内容

本发明的目的在于提供一种自动化的扫描阅卷系统及阅卷方法，有效地解决了传统人工阅卷存在的时间紧、任务重、查找纸质试卷工作量大、登分和核分保存困难以及耗时耗力等问题。提高了教师的阅卷质量，减轻了教师的劳动负担，有效地减少人工处理及答卷周转环节中的差错，省时省力，而且试卷图像可存档，为提高阅卷质量增添了一层保障，最大限度地保障评卷公平、公正，并且对电子化教学管理有重大的意义和影响。

本发明采用的技术方案是：采用试卷和答卷分离的方式，设计出一种包含主观题

和客观题的答卷的自动识别信息系统和方法，不仅使得教师可以方便快捷地制作答卷，而且能够高效地实现答卷分数等相关信息的识别。

本发明的系统由以下部分组成：试卷定义模块、基于VBA的答卷制作工具、文档扫描仪、数据库服务器、答卷信息自动识别模块、答卷识别结果校正模块、自动评分汇总模块。整个自动阅卷系统是一个星形结构，数据库服务器是系统的中心，其他各个模块及设备都单独以网络的形式与其相连接，用于信息传递和网络共享。试卷定义模块是整个系统的第一个组件，安装在用户的计算机上，用于试卷基本信息、结构信息以及答案与评分标准的定义，并把设置结果存储到数据库服务器中，为答卷制作提供信息；基于VBA的答卷制作工具是系统的第二个组件，安装在用户的计算机上，用于生成答卷的电子模板，其输出结果转化为标准答卷图像和考试实际使用的答卷；系统的第三个组件是文档扫描仪，安装在扫描人员的设备间，用来接收大量的考生答卷，对其电子化并自动转存到数据库服务器中；答卷信息自动识别模块是系统的第四个组件，安装在用户的计算机上，它从数据库服务器中获取考生答卷，之后进行自动识别，最后将识别结果返回给数据库服务器；系统的第五个组件是答卷信息校正模块，安装在用户的计算机上，从数据库服务器获取自动阅卷的结果信息并显示，供考务人员校正，并将校正结果返回存储到数据库服务器上；自动评分汇总模块是系统的第六个组件，安装在用户的计算机上，从数据库获取识别校正的结果数据，并根据试卷定义存储的信息，进行自动的评分与汇总。打印机是系统的可选组件，与自动评分汇总模块所在的计算机相连接，用于打印输出成绩等统计数据。

本发明的自动扫描阅卷方法包括以下几个步骤：

a. 试卷定义，是指利用本系统的试卷定义模块，完成试卷结构的定义和答案与评分标准的设置。步骤a还包括以下步骤：

a1. 试卷结构定义，是命题教师利用本系统的试卷定义模块，定义试卷的结构。首先，需要创建一份试卷的基本信息，包括满分值、考试年级、考试科目等信息；其次，创建大题，设置大题的主客观属性、大题数目和大题满分值；然后针对每道大题设置所包含的小题数目及各个小题的满分值，并把相应的信息存入数据库服务器中。

a2. 答案与评分标准设置，是指利用本系统的试卷定义模块，在完成试卷结构的定义后，在系统中录入各个客观题的标准答案以及多选题的评分标准，而后试卷定义模块会将这些信息存入数据库服务器中。

b. 基于VBA的答卷制作，是指利用本系统提供的基于VBA的答卷制作工具，根据已定义的试卷结构制作出答卷的电子模板，并把答卷模版存入数据库服务器中。已经制作好的电子模板主要用于以下两个方面：一是生成标准答卷图像，二是作为考生实际考试使用的答卷。在本发明中，标准答卷图像是指根据已制作好的答卷电子模板，利用Microsoft Office Document Image Writer的功能而直接生成的TIFF图像，从标准答卷图像识别出的坐标信息用于考生答卷图像的信息获取。

c. 考试，是指考生根据考试安排参与考试的过程。考试之前，考务人员需要将考生的考号收集并存储到数据库服务器中。考试过程中，考生需要书写并填涂答卷中的考生信息区域、将客观题的答题结果填涂在答卷的客观题区域以及将主观题的答题结

果书写在答卷的主观题答题区域。

d. 主观题评阅,是指阅卷人员对考生答卷中主观题评阅的过程,在评阅完成之后,阅卷人员将主观题的得分填涂在考生答卷的主观题分数涂卡区中。

e. 答卷扫描,是指利用文档扫描仪,将主观题评阅过的考生答卷成批扫描成一个一个TIFF格式的图像文件(通常,同一试室或班级的考生的答卷会被扫描成一个TIFF文件),所得的图像存入数据库服务器内,用于后续的考生答卷信息识别之用。

f. 答卷信息识别,是指利用本系统的答卷信息识别模块,识别出标准答卷和考生答卷中标记位、考号、客观题和主观题等区域的信息,并将识别信息存入到数据库服务器中。步骤f还包括以下两个步骤:

f1. 标准答卷信息识别,是识别并记录标准答卷图像的各部分(包括标记位、考号、客观题区域、各个主观题答题区域、各个主观题涂卡区域)的坐标信息,之后答卷信息识别模块将这些坐标信息存储到数据库服务器中。

f2. 考生答卷信息识别,是识别出考生答卷的填涂信息。这是在标准答卷图像信息识别完成的基础上进行的,首先将倾斜的考生答卷旋转并校正到与标准答卷图像统一的坐标系上,然后根据位置标准答卷图像中各部分的坐标获取各个识别项的位置信息,以识别出考生答卷的填涂信息。

g. 答卷识别结果校正,是利用本系统的答卷识别结果校正模块,将错误的识别信息进行校正。为了提高系统的准确性,对考生答卷进行图像分割,分割成为考号区域图像、考生信息区域图像(如姓名、学校和班级等)、客观题区域图像以及各个主观题区域图像。根据数据库服务器中的考号信息,自动判断已识别考号的有效性(包括考号是否识别完全、考号是否存在、考号是否重复出现等情况);对于主观题,根据当前题目的满分值判断教师所判分数是否合理等。针对各种不合理的情况,给出友好明确的提示。答卷识别结果校正模块提供了各个切割图像以及相应的答案识别(针对客观题)或评分(针对主观题)结果,阅卷人员可以据此实现校正,校正后的信息将存储到数据库服务器中。

h. 自动评分汇总,是利用本系统的自动评分汇总模块,根据自动识别和校正后的客观题答案与主观题评分识别结果,依据评分规则,自动评分汇总每份答卷的分数。本发明的有益效果是:

本发明的答卷信息自动识别方法能够高效地识别出答卷的各项信息,准确率达到99%以上,识别结果基本不需要校正,极大地提高了阅卷的效率,降低了阅卷过程的复杂度,减少了人工误差,提高了透明度,减轻了考务人员工作量,最大限度地实现了考试的公开公平性。

自动化扫描阅卷系统有良好的经济效益和社会效益:

1. 自动阅卷系统,性能稳定、识别准确和实时性好,自动阅卷的阅卷速度在1分钟30页以上,准确率在99%以上。

2. 硬件设备较少,技术要求低,只要会使用办公软件Word、高速扫描仪以及本发明中的软件工具即可。

3. 对答卷纸张质量要求低,使用普通的复印纸即可,靠图像的标记位定位而不需

卡纸定位。

4. 对答题用笔要求低, 学生答题过程和阅卷人员主观题判分过程中可用铅笔、钢笔、圆珠笔等各种书写笔填涂, 而不特定要求专用的2B铅笔, 尊重了学生的答题习惯。

5. 对答卷印刷要求低, 系统能够处理倾斜、倒置以及小变形等情况。

6. 操作简单, 阅卷人员只需批阅主观题, 其余工作都用系统的各个模块实现, 大大减轻相关人员工作量。

7. 阅卷之后的答卷能够返还给学生, 有利于学生了解自己的答题情况, 有针对性地加强自己的学习。

8. 应用本系统, 实现试卷文档的电子化, 对于构建试题库、科学的分析试卷的信度和效度以及构建区域常模等有积极的作用。

9. 从试卷定义到答卷制作, 出题教师可以利用本系统方便快捷地实现。

10. 答卷图像存储在数据库中, 校正时查阅方便, 利于用户调整识别数据时有据可寻, 提高阅卷的准确度和透明度。

11. 有效地降低了人工误差, 免去了保存纸质试卷的劳累, 大大节约了人力、物力, 节省了各种费用的开销。

附图说明

图1 是本发明的自动化扫描阅卷系统结构示意图。

图2 是本发明的阅卷方法的步骤示意图。

图3 是本发明的自动化扫描阅卷系统的试卷定义模块之功能模块图。

图4 是本发明的阅卷方法所采用的答卷各要素说明图。

图5 是本发明的阅卷方法之标准答卷信息识别流程图。

图6 是本发明的阅卷方法之考生答卷信息识别流程图。

具体实施方式

本发明用于含有主观题和客观题答卷的自动化扫描阅卷, 主要原理是: 采用试卷和答卷分离的方式, 命题教师定义并制作考试答卷, 待考试完毕后, 教师评阅主观题, 而后考生的答卷通过扫描仪扫描成图像文件, 利用图像处理和模式识别技术在计算机上对电子化了了的考生答卷进行评阅。其包括以下步骤:

a. 试卷定义, 是指利用本系统的试卷定义模块, 完成试卷结构的定义和答案与评分的设置, 并将相关设置信息存储到数据库服务器中, 为答卷制作及自动评分汇总提供信息。

a1. 试卷结构定义, 是命题教师利用本系统的试卷定义模块, 定义试卷的结构。首先, 需要创建一份试卷的基本信息, 包括满分值、考试年级、考试科目等信息; 其次, 创建大题, 设置大题的主客观属性、大题数目和大题满分值; 然后针对每道大题设置所包含的小题数目及各个小题的满分值, 并把相应的信息存入数据库服务器中。

a2. 答案与评分标准设置, 是指利用本系统的试卷定义模块, 在完成试卷结构的定义后, 在系统中录入各个客观题的标准答案以及多选题的评分标准, 而后试卷定义模块会将这些信息存入数据库服务器中。

b. 基于VBA的答卷制作, 是指利用本系统提供的基于VBA的试卷制作工具, 根据已

定义的试卷结构制作出答卷的电子模板，并把答卷模版存入数据库服务器中。已经制作好的电子模板主要用于以下两个方面：一是生成标准答卷（没有填图痕迹的答卷）图像，二是作为考生的答卷。步骤b具体还包括以下步骤：

b1. 标记位的生成，命题教师利用答卷制作工具在答卷的左上、左下、右上自动生成三个圆形黑点，它们就是所谓的标记位。其中，左上与左下竖直对齐，左上与右上水平对齐。标记位一旦确定，就不再变动，它用来校正扫描后发生了倾斜的考生答卷图像。

b2. 考生信息的生成，考生信息包括考生的姓名、班级、学校、考号，根据需要时可增加A/B卷类型及缺考标记，这些信息已被预先制作为固定图像，当打开试卷模板后，无需出题教师干涉，试卷会自动加载该图像到固定位置。

b3. 考试名称的生成，考试名称包括考试时间地点和考试科目两部分，答卷制作工具会根据命题教师的输入自动加载到答卷文档的固定位置。

b4. 客观题的生成，命题教师根据试卷结构，确定客观题的数目，答卷制作工具根据客观题的数量自动选择合适的客观题图片模板(0~100题)并自动加载到固定位置。

b5. 主观题的生成，命题教师根据本次考试的试卷情况，设置答卷所需的页数、每页的栏数、每栏的题目数、每个题目的高度及分数，答卷制作工具根据这些设置自动生成相应的页、栏、主观题答题区。答卷制作工具自动判断每个主观题的分数，当分数小于等于9.5分时，增加一行涂卡分数的图片，当分数大于9.5分时，增加两行涂卡分数的图片。

b6. 标准答卷图像的生成，是指答卷制作工具调用Microsoft Office Document Image Writer的功能，将Word中设计完成的电子模板输出为TIFF格式的图像文件，并将该图像保存在数据库服务器上，用于获得标准答卷图像的位置信息。

b7. 考生答卷的生成，以答卷电子模板为样板印刷后作为考生答卷。

c. 考试，是指考生根据考试安排参与考试的过程。考试过程中，考生需要书写并填涂答卷中的考生信息区域，并根据试卷题目将客观题的答题结果填涂在答卷的客观题区域，将主观题的答题结果书写在答卷的主观题答题区域。

d. 主观题评阅，是指阅卷人员对考生答卷中主观题评阅的过程，在评阅完成之后，阅卷人员将主观题的得分填涂在考生答卷的主观题分数涂卡区中。

e. 答卷扫描，是指利用文档扫描仪，以一定的分辨率（150dpi左右），将主观题评阅过的考生答卷成批扫描成一个个TIFF格式的图像文件（通常，同一试室或班级的考生的答卷会被扫描成一个TIFF文件），所得的图像存入数据库服务器内，用于后续的考生答卷信息识别之用。

f. 答卷信息识别，是利用本系统的答卷信息识别模块，识别出标准答卷和考生答卷图像中的相关信息，并把识别信息存入到数据库服务器中。步骤f含有以下步骤：

f1. 标准答卷信息识别，是识别并记录标准答卷图像的各部分（包括标记位、考号、客观题区域、各个主观题答题区域、各个主观题涂卡区域）的坐标信息，之后答卷信息识别模块将这些坐标信息存储到数据库服务器中。步骤g1的还包括如下步骤：

f1.1. 标准答卷图像读取与预处理,是指利用系统的答卷信息识别模块将每个答卷图像从标准答卷图像包TIFF文件包中分离出来,读入内存并转化成256个灰度级别BMP格式的图像,之后对BMP图像进行二值化。

f1.2. 标记位识别与定位,是指识别并记录标记位的坐标值。基于模板匹配方法,识别答卷中的左上、左下和右上等三个位置的标记位坐标,并记录到数据库服务器中,这些坐标信息用于旋转校正扫描过程中发生了倾斜的考生答卷,以使其与标准图像在同一个坐标系中。

f1.3. 考号信息坐标定位,是指识别并记录考号区域及每个具体考号的坐标。根据标记位坐标与考生区域相对位置的信息,分别向X轴、Y轴投影,投影值与阈值对比,从而识别出考号区域的边框坐标,之后在每个考号区域内中向Y轴投影获得10个数字的坐标,并按一定规则记录到数据库服务器中。

f1.4. 客观题坐标定位,是识别出客观题区域及每道选择题的A、B、C、D具体选项的坐标值。根据标记位与客观题区域相对位置的信息,分别向X轴、Y轴投影,并将投影值与阈值对比来识别客观题区域的边框坐标,之后在其区域内分别向X轴、Y轴投影以识别各选项的横纵坐标信息,并按一定规则记录到数据库服务器中。

f1.5. 主观题坐标定位,是确定每个主观题的区域坐标。先将BMP格式的答卷图像整体膨胀后投影,然后根据主观题答题区横竖栏线黑色像素较多的特点,分别向X轴、Y轴投影,投影值与阈值对比从而获取边框线的坐标,并根据宽度高度等信息进行分组判断,从而得出属于同一道主观题答题区的边框线,由此获得主观题答题区坐标信息。以每个主观题答题区的下边线坐标为基准,继续使用投影的方法,得到涂卡分数区域的相对坐标信息,并按一定规则将主观题答题区绝对坐标和涂卡分数区的相对坐标信息记录到数据库服务器中。

f2. 考生答卷信息识别,是识别出考生答卷的填涂信息。这是在标准答卷图像信息识别完成的基础上进行的,首先将倾斜的考生答卷旋转并校正到与标准答卷图像统一的坐标系上,然后根据位置标准答卷图像中各部分的坐标获取各个识别项的位置信息,以识别出考生答卷的填涂信息。步骤g2的具体步骤如下:

f2.1. 考生答卷图像读取与预处理,是将TIFF格式的考生答卷分解成BMP格式的答卷。考生答卷是以TIFF格式的图像包读入的,包含了多张答卷图像。在识别考生信息之前,需先将某个考生的答卷的图像数据从TIFF图像包中分离出来,读入内存并转化成256灰度的BMP格式图像,之后对BMP格式的答卷图像进行去噪、二值化、增强等预处理。

f2.2. 考生答卷图像倾斜校正。扫描得到的图像通常会有一些倾斜,本系统的答卷信息识别模块首先识别出考生答卷中的标记位坐标,并以标准答卷图像的标记位坐标为标准,对考生答卷进行旋转和平移,使得考生答卷图像与标准答卷图像位于同一个坐标系中。

主要步骤为:首先,使用模板匹配算法获取考生答卷的标记位坐标;其次,根据三个标记位的相对位置判断是否为倒置答卷(若左上、左下、右上有标记位而右下无标记位,则说明答卷是正置的;若左下、右上、右下有标记位而左上无标记位则说明

答卷是倒置的。后种情况需要对倒置答卷旋转180度让其正置)；然后，根据左上和右上的标记位是水平的先验值来计算倾斜角度，并以此角度进行旋转；最后，将标准答卷标记位坐标与旋转后考生答卷标记位坐标相比较，得到水平和垂直距离差，而后进行相应的平移，使得考生答卷与标准答卷各部分坐标重合。

f2.3. 考号识别，是识别出考生填涂的考号值。考号识别是考生答卷信息识别的第一步，每张答题卡切割的各部分图像是以考号命名区分的。首先根据标准答卷的考号区域坐标，设定识别范围；向Y轴投影并将投影值与阈值对比识别填涂区域，再根据标准答卷已知各数字的坐标判断填涂数字；之后将识别出的所有数字格式化为考号并返回；最后将考号区域图像切割、索引并保存，以备结果展示和分数核查时使用。

f2.4. 客观题选项识别，是识别出客观题的具体选项。首先，根据标准答卷的客观区域坐标，设定每个题的识别范围；然后，向X轴投影并将投影值与阈值对比从而识别填涂区域；继而，根据标准答卷已知各选项的坐标判断具体填涂选项；之后，将每个题的选项格式化为答案并返回；最后，将整个客观题区域图像切割、索引并保存，以备结果展示和分数核查时使用。

f2.5. 主观题分数识别，是识别出每个主观题的得分。主观题分数是以涂卡方式记录的，根据标准答卷可知各个主观题答题区的坐标值，再根据涂卡区的相对坐标信息设定识别范围，分别向X轴、Y轴投影并将投影值与阈值对比从而判断填涂位置，根据标准答卷已知的各数字坐标可以识别出所填涂的具体数字；之后将识别的每个数字格式化为分数返回；最后将该主观题及其分数区域的图像切割、索引并保存，以备结果展示和分数核查时使用。

f2.6. 识别结果返回，是返回一份完整答卷的填涂信息。当以上所有信息识别结束后，把所有的信息记录到数据库服务器中，以备结果展示和查询使用。

g. 答卷识别结果校正：是指利用系统的答卷识别校正模块，将识别错误的信息进行校正。为了提高系统的准确性，系统在批量评卷结束后，阅卷教师可以选择任意一个TIFF考生考卷包，察看学生的评分情况，进行误差修正。利用图像切割算法对考生答卷图像进行了分割，分割成为考号区域图像、考生信息区域图像（如姓名、学校和班级等）、客观题区域图像以及各个主观题区域图像。答卷识别校正模块提供了各个切割图像以及相应的答案识别（针对客观题）或评分（针对主观题）结果，用户可以据此实现校正，以获得准确的数据。

g1. 修正考号，当识别出的考号出现识别不全、考号重复、考号不存在等异常问题时，问题考号将自动被列出，并根据数据库中的记录将相应的考号图像显示出来，阅卷教师可方便地根据考号图像手工修正考号。

g2. 修正客观题分数，客观题也会将其整个题目图像显示出来，以供阅卷教师参考修正。当客观题选项异常，如本应单选但考生多选或未选，该题答案会显示异常颜色，以提示阅卷教师注意。阅卷教师可根据图像中的实际填涂选项修正答卷答案。

g3. 修正主观题分数，当主观题分数出现未识别、分数越界这些异常时。系统也会将主观题相应的题目图像显示出来，以供阅卷教师参考修正。

h. 自动评分汇总模块：是根据数据库中的考号信息，自动判断考号的有效性，包

括考号是否识别完全，考号是否存在，考号是否重复出现等情况；对于主观题，根据当前题目的满分值判断教师所判分数是否合理等。之后将各部分得到的分数进行汇总，从而得出整份答卷的最终分数。

本发明的图像处理所采用的关键技术：

基于 VBA 的试卷制作工具是在 MS Word 基础上做的二次开发，利用了 VBA 的宏编程技术，能够自动生成标记位，并可根据用户与系统的交互自动地确定答卷的页数、每页的栏数、客观题的数目及主观题的数目与评分区域，使得试卷制作操作简单，所产生的标准答卷图像准确有效。

标记位识别是系统阅卷的关键部分和信息正确识别的基本前提，本系统采用模板匹配技术来实现标记位的识别，并利用标记位实现考生答卷的旋转、平移校正。

对边框、填涂等信息采用投影方法与阈值比较方法，向 X 轴或 Y 轴投影，并将此投影值与事先设定的阈值对比，从而判断是否有边框或者是否填涂值。该方法简单、快速、有效，保证了系统的性能。

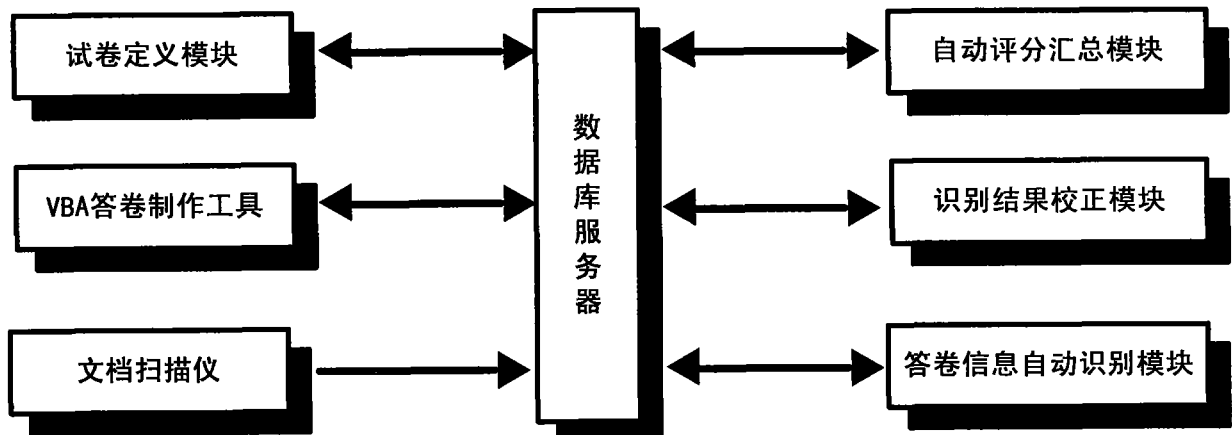


图 1

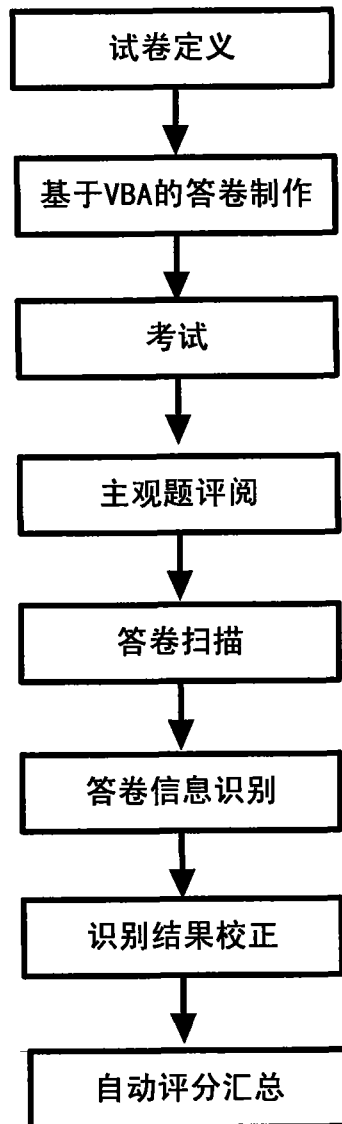


图 2

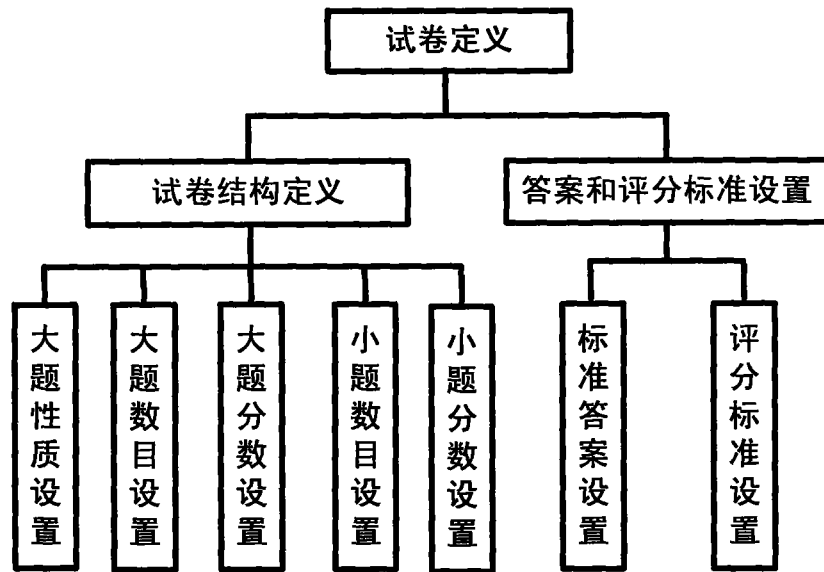


图 3

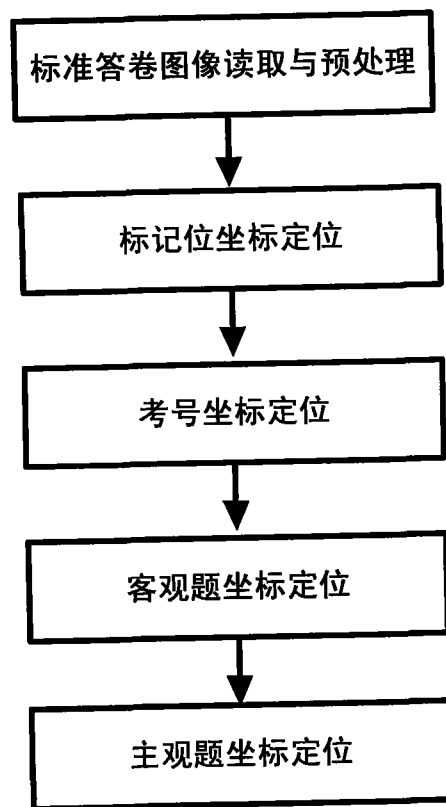


图 5

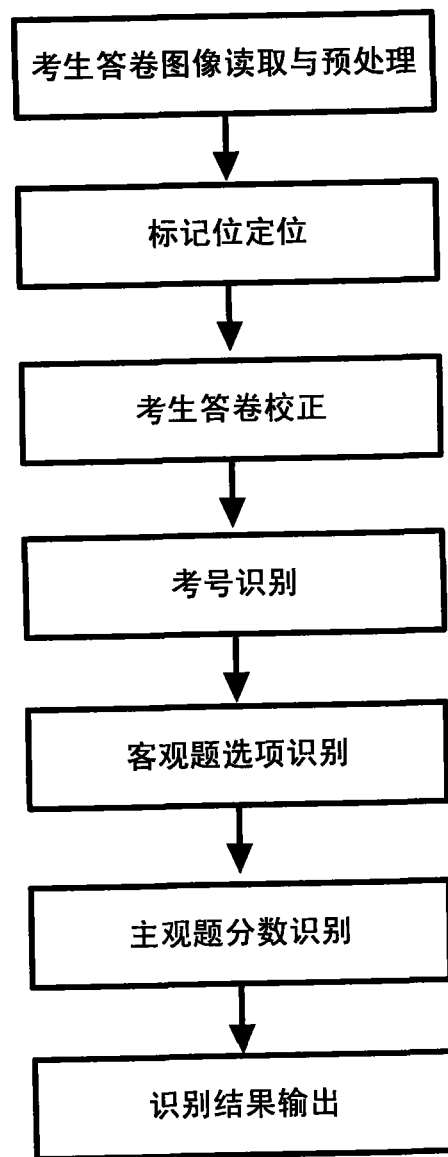


图 6