### 系统功能性需求分析

本系统是一个基于生活场景帮助人们快速识别文字（OCR）并将识别结果播报（语音合成）出来，有助于不方便使用手机等手持设备时识别内容，并且能根据人们的需求实现语音识别功能。

本系统是Android 客户端的应用程序，方便用户使用手机等安卓手持设备实现识别文字（OCR）等功能。

### 文字识别需求核心技术

1. 文字识别过程主要有：预处理、特征提取、特征匹配、识别后处理。
2. 由于普通拍摄图片会存在干扰，所以要对图片进行预处理，增强图片，降低干扰。
   1. 图片灰度化技术

要对采集到的图片进行灰度化处理，主要是因为灰度图片相对彩色的图片抛出了一部分的数据信息，数据量更小，更容易处理，并且更能显示文字的特点。

RGB图片灰度化技术，通俗点说就是对图像的RGB三个分量进行加权平均得到最终的灰度值。最常见的加权方法如下：

1）Gray=B；Gray=G；Gray=R

2）Gray=max(B+G+R)

3）Gray=(B+G+R)/3

4）Gray= 0.072169B+ 0.715160G+ 0.212671R

5）Gray= 0.11B+ 0.59G+ 0.3R

这三种方法中，第一种为分量法，即用RGB三个分量的某一个分量作为该点的灰度值；第二种方法为最大值法，将彩色图像中的三分量亮度的最大值作为灰度 图的灰度值。第三种方法将彩色图像中的三分量亮度求平均得到一个灰度图；后两种都是属于加权平均法，其中第四种是OpenCV开放库所采用的灰度权值，第 五种为从人体生理学角度所提出的一种权值（人眼对绿色的敏感最高，对蓝色敏感最低）。

* + 1. 二值技术
    2. 图像的二值化处理就是将图像上的点的灰度值为0或255，也就是将整个图像呈现出明显的黑白效果。即将256个亮度等级的灰度图像通过适当的阈值选取而获得仍然可以反映图像整体和局部特征的二值化图像。在数字图像处理中，二值图像占有非常重要的地位，特别是在实用的图像处理中，以二值图像处理实现而构成的系统是很多的，要进行二值图像的处理与分析，首先要把灰度图像二值化，得到二值化图像，这样子有利于在对图像做进一步处理时，图像的集合性质只与像素值为0或255的点的位置有关，不再涉及像素的多级值，使处理变得简单，而且数据的处理和压缩量小。

为了得到理想的二值图像，一般采用封闭、连通的边界定义不交叠的区域。所有灰度大于或等于阈值的像素被判定为属于特定物体，其灰度值为255表示，否则这些像素点被排除在物体区域以外，灰度值为0，表示背景或者例外的物体区域。 如果某特定物体在内部有均匀一致的灰度值，并且其处在一个具有其他等级灰度值的均匀背景下，使用阈值法就可以得到比较的分割效果。

如果物体同背景的差别表现不在灰度值上（比如纹理不同），可以将这个差别特征转换为灰度的差别，然后利用阈值选取技术来分割该图像。动态调节阈值实现图像的二值化可动态观察其分割图像的具体结果。

OpenCV中有两个函数可以实现图片的二值化：

cvThreshold( dst, dst,230 , 255, CV\_THRESH\_BINARY\_INV);

cvAdaptiveThreshold( dst, dst, 255, CV\_ADAPTIVE\_THRESH\_MEAN\_C,CV\_THRESH\_BINARY, 9, -10);

方法（1）是手动指定一个阈值，以此阈值来进行二值化处理。其中的第四个参数决定了该方法的结果：

threshold\_type=CV\_THRESH\_BINARY:

dst(x,y) = max\_value, if src(x,y)>threshold 0, otherwise.

threshold\_type=CV\_THRESH\_BINARY\_INV:

dst(x,y) = 0, if src(x,y)>threshold; dst(x,y) = max\_value, otherwise.

threshold\_type=CV\_THRESH\_TRUNC:

dst(x,y) = threshold, if src(x,y)>threshold; dst(x,y) = src(x,y), otherwise.

threshold\_type=CV\_THRESH\_TOZERO:

dst(x,y) = src(x,y), if (x,y)>threshold ; dst(x,y) = 0, otherwise.

threshold\_type=CV\_THRESH\_TOZERO\_INV:

dst(x,y) = 0, if src(x,y)>threshold ; dst(x,y) = src(x,y), otherwise.

值得一说的是threshold\_type可以使用CV\_THRESH\_OTSU类型，这样该函数就会使用大律法OTSU得到的全局自适应阈值来进行二值化图片，而参数中的threshold不再起作用。比如：cvThreshold( dst, dst,300 , 255, CV\_THRESH\_OTSU | CV\_THRESH\_BINARY\_INV);这种方法对于灰度直方图呈现二峰特征的图片处理起来效果很好。

方法（2）是一个自适应阈值二值化方法，通过设定最后两个参数来调整效果。

* + 1. 基于统计模型（隐马尔可夫模型）语义纠错

隐马尔可夫模型是由两种机理构成的随机过程。一种是内在的有限状态马尔可夫链,另一种是一系列随机函数所组成的集合。其中每一个函数都与一个状态相联系,马尔可夫链按照转移概率矩阵改变状态。因为观察者只能看到与每一状态相关联的随机函数的输出值,而不能观察到马尔可夫链的状态故称为隐马尔可夫模型[3]。

* + 1. OCR技术难点是什么？

复杂背景、艺术字体、低分辨率、非均匀光照、图像退化、字符形变、多语言混合、文本行复杂版式、检测框字符残缺，等等。

如克服这些难点的？从几个方面入手。一是使用场景，另一方面是从技术上的改进文本检测技术方进行了深度优化，通过设计合理的网络结构来提升各尺度的文字检测／提取能力。引入RNN多层自适应网络和Refinement结构来提升检测完整性和准确性。

##### Android 客户端需求分析

Android客户端应用程序具备多个文字识别的内容，帮助用户在不同场景下快速实现完成文字识别的功能（OCR），并且不方便查看识别结果时将结果语音合成播报，也需要满足人们不方便打字时通过语音识别功能识别人们话语。

图1 详细的分析安卓应用所需要的需求：



图 1

* 1. 系统功能
     1. 通用文字识别

具备调用手机摄像头拍摄任意带有文字的内容，比如一张图片中的文字，纸巾包装上的文字等。

* + 1. 身份证识别

可以识别身份证正反面的内容，识别出姓名，出生年月日，性别等内容。

* + 1. 银行卡识别

可以通过手机摄像头拍摄银行卡识别出银行卡的卡号与发卡行。

* + 1. 车牌识别

可以通过手机摄像头拍摄车牌号（可以是图片），识别出车牌号码。

* + 1. 车型识别

可以通过手机摄像头拍摄常见的车型（可以是图片），系统根据拍摄的图片识别出车型。

* + 1. 语音识别

可以识别出用户的语音，并且显示到手机屏幕上

* + 1. 语音合成

可以将文字识别的结果合成语音，通过手机扬声器播放出来，能设置语速，播讲人等参数。

* 1. 系统非功能性需求分析

整体应用对性能有较高的要求，调取摄像头时要注意防止内存泄漏，也不能出现程序无响应异常，需要站在用户的角度上体验整个应用，关键时刻出现错误，非常影响用户体验。

除此之外，进行应用开发时，心里对整个应用要保持着简洁、美观的要求，一个美观的界面与功能完整但是界面简陋，用户更加倾向于美观简洁的界面，同时，作为开发人人员要具备产品思维，一个好的应用，就就必须注重细节。另外还需要注意的是，Android 手机发展多年，最新版本也发展到 Android 9.0（2019年），我们做出来的应用不仅要注意适配市面上常见的手机设备，还要注意适配Android的版本，这里是非常难的。

## 系统设计

本系统的文字识别引擎与语音合成引擎皆采用百度公司的识别引擎，可以应用于多门语言，比如Java语言，PHP语言，Python语言，C#语言，C++语言，更包括可以应用于Android 与 iOS 平台上。

### 百度文字识别支持系统与硬件

支持的系统和硬件版本

系统：支持 Android 4.0（API Level 15）到Android7.0（API Level 25）系统。需要开发者通过minSdkVersion来保证支持系统的检测。

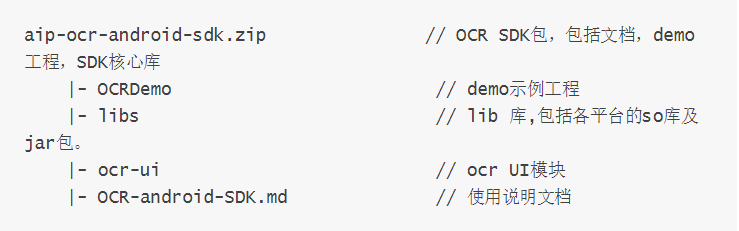
CPU架构：armeabi，arm64-v8a，armeabi-v7a，x86

机型：手机和平板皆可

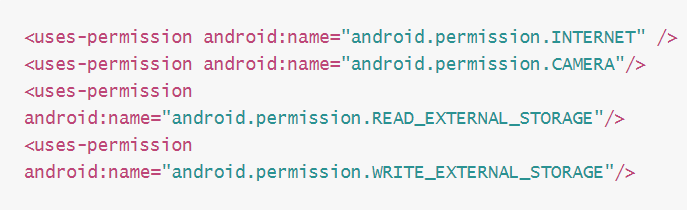
硬件要求：要求设备上有相机模块。

网络：支持WIFI及移动网络，移动网络支持使用NET网关及WAP网关（CMWAP、CTWAP、UNIWAP、3GWAP）。

* 1. 开发包说明



### 程序开始前需要添加的权限



各个权限的用途说明见下表：



### 身份验证与安全

### 系统界面设计

Android 界面布局设计，通常两种，一是静态布局文件设计，二是Java代码动态设置。

本应用两者皆有采用，主体不需要变动内容，采用静态布局文件设计，而需要根据不同结果采用不同的内容则采用Java代码动态设置。

### 首页设计

首页作为主要设计页面，可以让用户通过点击Android控件完成数据采集功能。其效果图如图：



图 2

用户打开本应用，在主页面上选择想要实现的功能，系统会调用相应的引擎进行处理，最后返回结果，可以通过语音合成播报识别结果。

文字识别时序图：



系统功能时序图 1

详细流程如下：

1. 启动本应用，系统将进行初始化操作
2. 用户通过点击屏幕相应的功能，开启识别功能。
3. 系统启动手机的摄像头，用户拍摄图片，系统系统已拍摄图片。用户点击确定，调用文字识别引擎，处理图片。
4. 识别完成后将数据返回给首页，调用对话框弹出返回的识别结果，
5. 系统调用语音合成，进行语音播报
   1. 系统相机获取图片

调用系统的相机模块共两种使用方式，用户可以拍摄图片或者选取手机中存储的图片，下图为调用系统相机的流程图：



图 3 调用系统相机获取图片流程图

* 1. Android XML布局部分代码展示：

<**Button**

android:id="@+id/general\_basic\_button"

android:layout\_width="match\_parent"

android:layout\_height="wrap\_content"

android:text="通用文字识别"

tools:ignore="HardcodedText" />

<**Button**

android:id="@+id/idcard\_button"

android:layout\_width="match\_parent"

android:layout\_height="wrap\_content"

android:text="身份证识别"

tools:ignore="HardcodedText" />

<**Button**

android:id="@+id/bankcard\_button"

android:layout\_width="match\_parent"

android:layout\_height="wrap\_content"

android:text="银行卡识别"

tools:ignore="HardcodedText" />

<**Button**

android:id="@+id/license\_plate\_button"

android:layout\_width="match\_parent"

android:layout\_height="wrap\_content"

android:text="车牌识别"

tools:ignore="HardcodedText" />

<**LinearLayout**

android:layout\_width="match\_parent"

android:layout\_height="match\_parent"

android:orientation="vertical">

<**Button**

android:id="@+id/car\_model\_button"

android:layout\_width="match\_parent"

android:layout\_height="wrap\_content"

android:text="车型识别"

tools:ignore="HardcodedText" />

<**Button**

android:id="@+id/speechrecognition\_button\_start"

android:layout\_width="match\_parent"

android:layout\_height="wrap\_content"

android:text="语音识别开始"

tools:ignore="HardcodedText" />

<**Button**

android:id="@+id/speechrecognition\_button\_stop"

android:layout\_width="match\_parent"

android:layout\_height="wrap\_content"

android:text="语音识别停止"

tools:ignore="HardcodedText" />

<**RelativeLayout**

android:layout\_width="match\_parent"

android:layout\_height="match\_parent">

<**ImageView**

android:id="@+id/car\_model\_iv"

android:layout\_width="match\_parent"

android:layout\_height="wrap\_content"

android:scaleType="centerCrop" />

<**TextView**

android:id="@+id/scan\_value\_tv"

android:layout\_width="match\_parent"

android:layout\_height="wrap\_content"

android:layout\_below="@id/car\_model\_iv"

android:background="@color/colorAccent"

android:gravity="center"

android:text="结果扫描中"

android:textColor="@color/colorWhite"

android:textSize="20sp"

android:textStyle="bold"

android:visibility="gone" /

</RelativeLayout>

<**TextView**

android:id="@+id/re\_result\_tv"

android:layout\_width="match\_parent"

android:layout\_height="wrap\_content" />

</LinearLayout>

* 1. 关键代码部分展示：

*// 通用文字识别，跳转调用手机摄像头*

findViewById(R.id.general\_basic\_button).setOnClickListener(new View.OnClickListener() {

@**Override**

**public** **void** onClick(**View** v) {

if (!checkTokenStatus()) {

return;

}

**Intent** intent = new Intent(MainActivity.this, CameraActivity.class);

intent.putExtra(CameraActivity.KEY\_OUTPUT\_FILE\_PATH,

FileUtil.getSaveFile(getApplication()).getAbsolutePath());

intent.putExtra(CameraActivity.KEY\_CONTENT\_TYPE,

CameraActivity.CONTENT\_TYPE\_GENERAL);

startActivityForResult(intent, REQUEST\_CODE\_GENERAL\_BASIC);

}

});

*// 识别成功回调，通用文字识别*

if (requestCode == REQUEST\_CODE\_GENERAL\_BASIC && resultCode == Activity.RESULT\_OK) {

RecognizeService.recGeneralBasic(this, FileUtil.getSaveFile(getApplicationContext()).getAbsolutePath(),

new RecognizeService.ServiceListener() {

@**Override**

**public** **void** onResult(**String** result) {

Log.i(TAG, "onResult: == " + getUniverstalTextJsonBean(result));

infoPopText("识别结果,查看以下内容", getUniverstalTextJsonBean(result));

speak(getUniverstalTextJsonBean(result));

}

});

}

*//弹出对话框*

**private** **void** alertText(**final** **String** title, **final** **String** message) {

this.runOnUiThread(new Runnable() {

@**Override**

**public** **void** run() {

alertDialog.setTitle(title)

.setMessage(message)

.setPositiveButton("确定", new DialogInterface.OnClickListener() {

@**Override**

**public** **void** onClick(**DialogInterface** dialog, **int** which) {

mSpeechSynthesizer.stop();

}

}) .show();

}

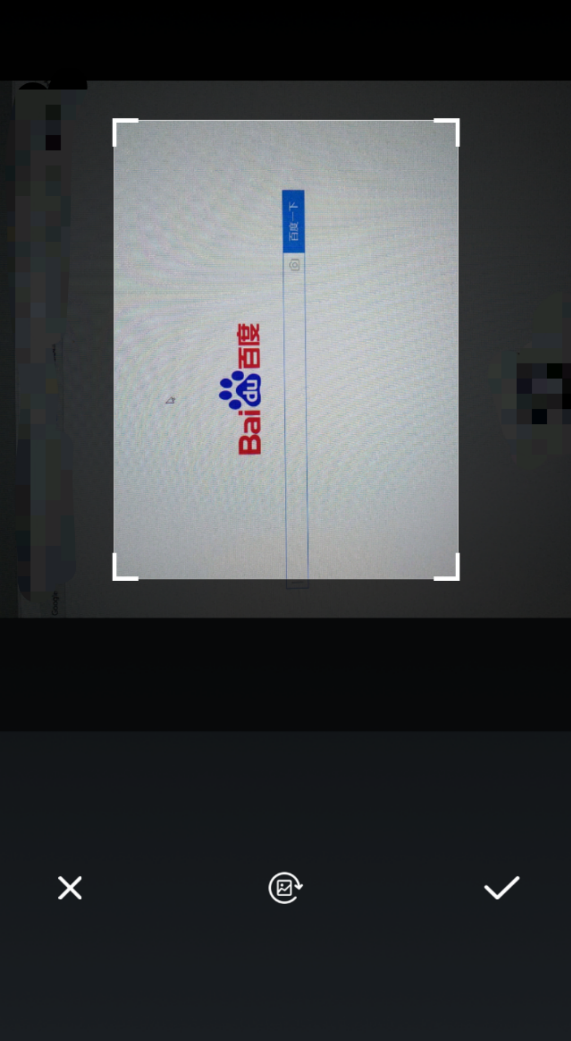
});

}

应用运行截图



图 4



* 1. 语音合成与语音识别核心技術

语音合成和[语音识别](https://baike.baidu.com/item/%E8%AF%AD%E9%9F%B3%E8%AF%86%E5%88%AB)技术是实现人机语音通信，建立一个有听和讲能力的口语系统所必需的两项关键技术。使电脑具有类似于人一样的说话能力，是当今时代信息产业的重要竞争市场。和语音识别相比，语音合成的技术相对说来要成熟一些，并已开始向产业化方向成功迈进，大规模应用指日可待。

语音合成，又称文语转换（Text to Speech）技术，能将任意文字信息实时转化为标准流畅的语音朗读出来，相当于给机器装上了人工嘴巴。它涉及声学、语言学、[数字信号处理](https://baike.baidu.com/item/%E6%95%B0%E5%AD%97%E4%BF%A1%E5%8F%B7%E5%A4%84%E7%90%86)、计算机科学等多个学科技术，是[中文信息处理](https://baike.baidu.com/item/%E4%B8%AD%E6%96%87%E4%BF%A1%E6%81%AF%E5%A4%84%E7%90%86)领域的一项前沿技术，解决的主要问题就是如何将文字信息转化为可听的声音信息，也即让机器像人一样开口说话。我们所说的“让机器像人一样开口说话”与传统的声音回放设备（系统）有着本质的区别。传统的声音回放设备（系统），如磁带[录音机](https://baike.baidu.com/item/%E5%BD%95%E9%9F%B3%E6%9C%BA)，是通过预先录制声音然后回放来实现“让机器说话”的。这种方式无论是在内容、存储、传输或者方便性、及时性等方面都存在很大的限制。而通过计算机语音合成则可以在任何时候将任意文本转换成具有高自然度的语音，从而真正实现让机器“像人一样开口说话”。

文语转换系统实际上可以看作是一个人工智能系统。为了合成出高质量的语言，除了依赖于各种规则，包括语义学规则、词汇规则、语音学规则外，还必须对文字的内容有很好的理解，这也涉及到自然语言理解的问题。下图显示了一个完整的文语转换系统示意图。

文语转换过程是先将文字序列转换成音韵序列，再由系统根据音韵序列生成语音波形。其中第一步涉及语言学处理，例如分词、字音转换等，以及一整套有效的韵律控制规则；第二步需要先进的语音合成技术，能按要求实时合成出高质量的语音流。因此一般说来，文语转换系统都需要一套复杂的文字序列到音素序列的转换程序，也就是说，文语转换系统不仅要应用数字信号处理技术，而且必须有大量的语言学知识的支持。[4]

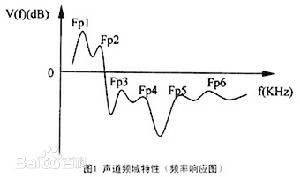


图 5

参考文献

1. 吕俊哲，图像二值化算法研究及其实现[J]，科技情报开发与经济，2004，（3）：125-129
2. 苏金明、王永利，MATALAB图形图像[M]，河北：华北电力大学，2008
3. 陈友仁、赵正校，基于隐马尔可夫模型的车牌自动识别技术，上海，2001
4. 张斌,全昌勤,任福继. 语音合成方法和发展综述[J]. 小型微型计算机系统,2016,(01):186-192.