

硕 士 研 究 生 工 作 报 告



题目 [TDB-verify](https://github.com/Zhejiang-University-GKC/TDB-verify)的工作报告

作者姓名 王超超，王聪，徐欣，刘曌，陈龙

指导教师 袁 昕

课程名称 数字媒体

所在学院 计算机学院

提交日期 二○一四年十二月

本小组的研究课题为识别12306验证码。

分工安排如下：

王超超负责样本的采集、分类、分析等工作。

王聪负责进行图片的二值化和去噪（由于12306没有噪点，只做二值化，以及学习去噪的方法），将图片变为0，1的黑白图片

陈龙负责进行图片的非粘连切割，将黑白的图片切割为多张图片，每个图片都由全部有连接的像素点组成。

徐欣负责根据切割出来的字符进行大量的训练，找出每个字母的特征，以便后续做对比，找出最相似的字母。

刘曌负责根据训练出来的特征做对比，将目标字母与样本特征库的特征进行对比，根据一定的算法计算出权值最大，即最相似的图片来。

最后陈龙做整合，将整个流程串起来，做成一个完整的验证码对比程序。

其中对于字母的特征是采用上下左右的趋向性的，计算出每个字母上下左右的趋向性，然后根据大量的样本，计算出每种趋向的概率，最后用于计算对比的权值，按照最大比例的结果作为最相似结果。

本次课题的不足有如下几点：

1. 训练的方法还不太优，对于部分趋向性很相似的子母分辨率很低
2. 还无法切割出粘连的字母，这样就无法做到识别所有验证码