

# 上海交通大学

SHANGHAI JIAO TONG UNIVERSITY

学士学位论文

THESIS OF BACHELOR



论文题目 基于视频的疲劳驾驶检测系统研究

学生姓名 杜若飞

学生学号 5090109228

指导教师 吕宝粮 教授

专 业 计算机科学与技术

学院 (系) 电子信息与电气工程

Submitted in total fulfilment of the requirements for the degree of  
Bachelor  
in Computer Science and Engineering

# Research on Fatigue Driving Detection System Based on Video Signals

RUOFEI DU

Supervisor

Prof. BAO-LIANG LU

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, SCHOOL OF ELECTRONIC  
INFORMATION AND ELECTRICAL ENGINEERING  
SHANGHAI JIAO TONG UNIVERSITY  
SHANGHAI, P.R.CHINA

Jan. 17th, 2013

# 上海交通大学

## 毕业设计（论文）学术诚信声明

本人郑重声明：所呈交的毕业设计（论文），是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

作者签名：\_\_\_\_\_

日期：\_\_\_\_\_年\_\_\_\_月\_\_\_\_日

# 上海交通大学

## 毕业设计（论文）版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权上海交通大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保 密 ，在 \_\_\_\_ 年解密后适用本授权书。

本学位论文属于

不保密 。

(请在以上方框内打“√”)

作者签名：\_\_\_\_\_

指导教师签名：\_\_\_\_\_

日 期：\_\_\_\_\_年 \_\_\_\_月 \_\_\_\_日

日 期：\_\_\_\_\_年 \_\_\_\_月 \_\_\_\_日

# 基于视频的疲劳驾驶检测系统研究

## 摘 要

疲劳驾驶是交通事故的主要原因之一，疲劳状态的预测对于飞行员、汽车和高速铁路驾驶员具有重要意义。然而，现有的疲劳驾驶检测系统往往基于单一信号源来提取单一特征，受光照、肤色的影响较大，预测不精确且成本高昂。对此，本文首创了一种经济、实用、准确的基于视频的疲劳驾驶预警新系统：该系统综合了几何与纹理特征，可以较好地结合从眼电、握力信号提取的特征进行预警，可以部署在便携式智能手机设备。

笔者第一个设计并完成了融合眼电与视频信号的疲劳驾驶模拟实验，于 ICONIP 2012 上首次提出：眼电信号具有表征准确的特性，视频信号具有表义丰富的特性；通过二者结合，可以大幅提升疲劳驾驶检测系统的准确率。该系统同时与基于眼电、脑电信号的疲劳驾驶检测系统参展 2011 年中国国际工业博览会获二等奖。此外，作者参与并完成了融合握力与视频信号的疲劳驾驶模拟实验，再次证明了视频系统的鲁棒性和准确性。

本系统通过基于 Cascade 结构和 Adaboost 算法的人脸检测模型和主动外观模型 (ASM) 对视频实时处理，采用二值化及图像梯度算法提取 PERCLOS 特征，并计算出慢速眼动 (SEM)、快速眼动 (REM)、打哈欠频率、头部朝向等特征，通过 Gabor 变换提取了纹理特征并采用 Adaboost 进行降维，最后用线性动力系统 (LDS) 进行去噪，支持向量机 (SVM) 回归拟合。此外，通过融合视频与眼电信号中的多种特征，我们显著提升了系统的准确率。实验表明，本系统在驾驶员出现疲劳驾驶时能够及时报警，可作为预防驾驶员疲劳驾驶的一种简单、经济、实用的辅助提醒工具，发挥较大的社会效益和经济效益。

**关键词：**疲劳驾驶 疲劳检测 主动外观模型 情绪识别  
PERCLOS

# Research on Fatigue Driving Detection System Based on Video Signals

## ABSTRACT

Fatigue driving is one of the major causes of traffic accidents. Additionally, the prediction of fatigue state is of crucial significance for pilots, drivers and chief conductors. Nevertheless, most of the existing Fatigue Driving Detection Systems(FDDS) are based on single signal source or extract single feature, which suffer from dim lighting, varying skin color, inaccuracy and high cost. Hence, this thesis is the first to propose a cheap, easy-to-use, accurate FDDS based on video signals, which combines geometry and texture features and predicts fatigue state by combining video with Electroencephalography(EOG) signals and gripping power.

The author is the first to propose fatigue driving experiments by combining video and EOG signals on ICONIP 2012: EOG is more accurate while video includes more information; by combining both signals, we greatly improve the prediction accuracy. This system won the 2<sup>nd</sup> prize on ICCF 2011. Meanwhile, the author conducted similar experiments with gripping power, which again proves the robustness and accuracy.

This system adopts Haar Cascade Adaboost for face detection and Active Shape Model for alignment, calculates PERCLOS by binarization and image gradient algorithms, calculates SEM, REM, yawn frequency and extracts boosted Gabor features from texture. Finally, we apply Linear Dynamic System to denoise and Support Vector Machine for regression. As a result, we improve the accuracy greatly by combining the features from EOG and video. The experiments indicate that our system could serve as an easy-to-use, economical tool for drivers thus benefiting the entire society.

**KEY WORDS:** Drowsy Driving, Fatigue Detection, Active Shape Model, Emotion Recognition, PERCLOS

# 目 录

<b>摘要</b>	<b>i</b>
<b>ABSTRACT</b>	<b>ii</b>
<b>目录</b>	<b>iii</b>
<b>主要符号对照表</b>	<b>viii</b>
<b>第一章 绪论</b>	<b>1</b>
1.1 课题研究的目的与意义 . . . . .	1
1.2 国内外研究现状综述 . . . . .	3
1.2.1 疲劳度的度量 . . . . .	3
1.2.2 基于视频的疲劳驾驶检测算法综述 . . . . .	3
1.2.3 基于视频的疲劳驾驶系统综述 . . . . .	4
1.2.4 基于其他传感器的疲劳驾驶系统综述 . . . . .	5
1.3 主要研究方法与创新点 . . . . .	6
1.3.1 技术路线 . . . . .	6
1.3.2 创新点 . . . . .	6
1.4 章节内容概述 . . . . .	7
<b>第二章 疲劳驾驶检测系统概述</b>	<b>8</b>
2.1 基于真实轿车与虚拟现实技术的模拟驾驶系统 . . . . .	8
2.2 基于 NeroScan 与红外摄像的疲劳驾驶训练系统 . . . . .	9
2.3 本章小结 . . . . .	10

<b>第三章 基于计算机视觉的预处理算法</b>	<b>11</b>
3.1 基于主动外观模型的技术路线	11
3.1.1 人脸检测算法	11
3.1.2 主动外观模型算法	12
3.1.3 基于 Paint Selection 的图像实时分割算法	13
3.2 基于深度学习的技术路线	15
3.2.1 深度信念网	15
3.2.2 用于人脸检测的深度学习网络	16
3.3 本章小结	17
<b>第四章 疲劳驾驶检测系统的特征提取算法</b>	<b>18</b>
4.1 基于视频信号的特征提取	18
4.1.1 PERCLOS 特征	18
4.1.2 眨眼特征	19
4.1.3 打哈欠特征	20
4.1.4 头部朝向特征	20
4.2 基于眼电信号的特征提取	20
4.2.1 传统眼电与前额眼电	20
4.2.2 独立成分分析	21
4.2.3 垂直眼电特征提取	21
4.2.4 水平眼电特征提取	22
4.3 基于握力信号的特征提取	22
4.3.1 握力信号的采集与转换	22
4.3.2 握力信号预处理	23
4.3.3 握力信号中的特征提取	23
4.4 本章小结	23
<b>第五章 基于视频与握力信号疲劳驾驶模拟实验</b>	<b>25</b>
5.1 实验概述	25



5.1.1	实验准备 . . . . .	25
5.1.2	实验过程 . . . . .	25
5.2	实验分析 . . . . .	26
5.3	本章小结 . . . . .	28
<b>第六章</b>	<b>基于视频与眼电信号的疲劳驾驶模拟实验</b>	<b>32</b>
6.1	实验概述 . . . . .	32
6.1.1	实验准备 . . . . .	32
6.1.2	实验过程 . . . . .	32
6.2	实验分析 . . . . .	33
6.2.1	实验结果 . . . . .	33
6.3	本章小结 . . . . .	35
<b>第七章</b>	<b>全文总结</b>	<b>36</b>
7.1	主要结论 . . . . .	36
7.2	创新点 . . . . .	37
7.3	前景展望 . . . . .	37
<b>附录 A</b>	<b>基于视频的疲劳驾驶检测系统开发大事记</b>	<b>39</b>
	<b>参考文献</b>	<b>40</b>
	<b>攻读学位期间发表的学术论文目录</b>	<b>45</b>
	<b>攻读学位期间参与的项目</b>	<b>46</b>
	<b>致谢</b>	<b>47</b>

## 表格索引

1-1 疲劳驾驶检测系统中视频、眼电、脑电信号应用优劣对比 . . . . .	6
3-1 CK+ 数据库中表情种类的分布数量 . . . . .	13
3-2 模糊积分算法合并两种分类器的实验结果 . . . . .	13
5-1 基于视频、握力的视频握力的疲劳度预测相关系数与均方误差 .	28
6-1 基于视频、眼电、视频眼电的疲劳度预测相关系数与均方误差 .	34

## 插图索引

1-1	生活中普遍存在的疲劳驾驶现象 . . . . .	1
2-1	基于真实轿车与虚拟现实技术的模拟驾驶系统实验图 . . . . .	8
2-2	基于握力与视频的疲劳驾驶检测系统监视图 . . . . .	9
2-3	基于 NeroScan 与红外摄像的疲劳驾驶训练系统实验图 . . . . .	10
2-4	按键实验所用四种交通标志示意图 . . . . .	10
3-1	预处理流程图 . . . . .	11
3-2	基于 Paint Selection 的图像实时分割算法实验图 . . . . .	14
3-3	基于深度信念网的手写体数字识别训练权值 . . . . .	16
4-1	直方图二值化和图像梯度算法流程图 . . . . .	19
4-2	前额眼电与水平眼电电极放置与比较示意图 . . . . .	21
4-3	利用差分算法在垂直眼电信号提取特征 . . . . .	22
4-4	握力信号的采集与转换原理示意图 . . . . .	23
5-1	基于视频与握力信号驾驶员的疲劳驾驶实验环境示意图 . . . . .	26
5-2	训练后的主动外观模型实验效果图 . . . . .	27
5-3	基于视频特征疲劳驾驶检测结果分析图 . . . . .	29
5-4	基于握力特征疲劳驾驶检测结果分析图 . . . . .	30
5-5	结合视频特征与握力特征疲劳驾驶检测结果分析图 . . . . .	31
6-1	系统流程图 . . . . .	33
6-2	线性动力系统特征平滑示意图 . . . . .	34
6-3	基于视频、眼电、视频眼电的疲劳度预测曲线 . . . . .	35

## 主要符号对照表

<i>AAM</i>	主动外观模型 (Active Appearance Model)
<i>ASM</i>	主动形状模型 (Active Shape Model)
<i>ANN</i>	人工神经网络 (Artificial Neural Network)
<i>BP</i>	反向传播算法 (Back Propagation)
<i>CNN</i>	卷积神经网络 (Convolutional Neural Network)
<i>DBN</i>	深度信念网络 (Deep Belief Network)
<i>EEG</i>	脑电信号 (Electroencephalography)
<i>EOG</i>	眼电信号 (Electrooculography)
<i>FACS</i>	面部动作编码系统 (Facial Action Coding System)
<i>ICA</i>	独立成分分析 (Independent Component Analysis)
<i>LGBP</i>	小波局部二值模式 (Local Gabor Binary Pattern)
<i>JSD</i>	约翰斯困乏尺度 (Jensen-Shannon Divergence)
<i>PCA</i>	主成分分析 (Principle Component Analysis)
<i>PERCLOS</i>	双眼平均闭合程度 (PERcentage of CLOSure)
<i>RBF</i>	径向基核函数 (Radial Basis Function)
<i>REM</i>	快速眼动 (Rapid Eye Movement)
<i>SEM</i>	慢速眼动 (Slow Eye Movement)
<i>SVM</i>	支持向量机 (Support Vector Machine)

# 第一章 绪论

本章主要阐述研发基于视频的疲劳驾驶检测系统的目的、意义和综述，以及本文的主要研究方法和创新点。首先，通过列举国际公路协会以及我国交通局、统计局发布的统计数据，论证研发疲劳驾驶检测系统的时代背景和重要意义。其次，本章会涵盖现有的主要疲劳驾驶检测系统、疲劳驾驶检测方法，并讨论它们的优缺点和亟待解决的问题。然后，笔者简述一下本文阐述的基于视频的疲劳驾驶检测系统的主要技术路线和创新点。最后，笔者介绍一下后面的各章节安排，起到提纲挈领的作用。

## 1.1 课题研究的目的是与意义



图 1-1 生活中普遍存在的疲劳驾驶现象

毋庸置疑，安全，尤其是交通安全，是关系到每一位公民切身利益的重要课题。而疲劳驾驶，正逐渐成为极大威胁交通安全的罪魁祸首，被称为交通事故的“隐形杀手”：相比于酒后驾车，疲劳驾驶有着受众人群更加广泛，高速公路上潜在威胁更加严重，定量标准更加难以统一等特点。具体请看下面一组数字：

国际公路协会 2011 年发布的《世界公路统计数据 2004-2009》<sup>[1]</sup> 表明：仅在 2009 年，全世界 41 个主要国家和地区共发生交通事故 494.8 万起，因交通事故而受伤的人数为 718.8 万，死亡人数则高达 41.6 万人。据美国联邦高速公路安全管理局的一份报告<sup>[2]</sup> 阐述，约三分之一的高速公路交通及人员伤亡事故是人为因素造成的，其中疲劳驾驶是事故发生的主要原因。

根据《中华人民共和国道路交通事故统计年报》<sup>[3]</sup>，仅在 2003 年，全国因疲劳驾驶造成的交通事故高达 7767 起，占总交通事故的 1.16%，死亡人数高达 2231 人，占总死亡人数的 2.14%，造成直接经济损失 1.1 亿人民币，占总损失的 3.41%。其中，在高速公路上，超速行驶和疲劳驾驶是肇事致人死亡人数最多的交通违法行为，导致事故死亡人数占高速公路死亡人数的 18.9%；在恶性事故中所占比率更是高达 45.7%。此外，机动车驾驶员因疲劳驾驶肇事有着逐年递增的趋势，2010 年统计年报表明，因疲劳驾驶导致交通事故的死亡人数同比上升 1%，造成经济损失高达 4.2 亿人民币。在高速公路上，疲劳驾驶是仅次于超速驾驶的交通事故主要原因。在部分地区（如浙江）高速公路上，疲劳驾驶占交通事故的比重更是达到了惊人的 45.7%。鉴于目前对于疲劳驾驶并没有一个统一的判定标准，实际上及驾驶员因疲劳驾驶导致的交通事故比例要比交通局统计数据高很多。文献 [4] 表明，高速列车司机作业后，脑力疲劳和心理疲劳比眼睛疲劳及体力疲劳更为明显，即司机的疲劳从外表不易被观察到。

就在 2013 年伊始，中华人民共和国公安部发布第 123 号公安部令—《机动车驾驶证申领和使用规定》<sup>[5]</sup>，将“连续驾驶中型以上载客汽车、危险物品运输车辆超过 4 小时未停车休息或者停车休息时间少于 20 分钟”判定为疲劳驾驶，记 12 分。然而，不幸的是，目前对于疲劳驾驶尚无可靠的检测手段和量化的评价体系，从驾驶时间上很难准确判定驾驶员的疲劳程度，不但执法人员无法对疲劳驾驶采取有效的监管措施，而且科研人员也一直没有准确、便捷地检测疲劳驾驶的方案。

因此，对于驾驶疲劳的产生机理、定量描述和检测方法的研究是神经科学、认知科学和智能交通系统研究的一个重要课题。研发司机疲劳状态监测及预警技术，把“马路杀手”遏制在萌芽状态，对于改善我国、乃至全球的交通状况都有着重要意义。它不但具有极高的理论价值，同时具有重要的社会和经济效益。对于挽救那些不时处于疲劳驾驶而危及生命的飞行员、动车和高速铁路驾驶员来说，更是一项刻不容缓的工作。

上海交通大学智能计算与智能系统微软-教育重点实验室致力于研发基于红外视频、Kinect、眼电、脑电、握力、心率等多模态疲劳驾驶检测系统，其中，基于视频的疲劳驾驶检测系统因其成本低廉，与驾驶者无接触，成为了其中一个重要组成部分。

## 1.2 国内外研究现状综述

自 1956 年起，心理学家 Adams 就已经开展了对疲劳问题的研究 [6]。这个时期的研究主要着眼于人体在清醒和睡眠两种状态的区别，哪些生理特征发生了改变。后来的研究者 Baker 通过实验，逐渐深入研究人体从清醒到睡眠的变化过程 [7]。然而，直到今天，虽然疲劳相关研究已经取得了一些进展，但尚未取得疲劳度量的一个统一标准。近些年的研究主要围绕视频信号、眼电信号、脑电信号，以及脉搏、呼吸频率、体温等生理信号进行疲劳驾驶检测。随着计算机视觉和机器学习技术的发展，人的面部特征被越来越多地用于疲劳驾驶检测。从视频信号中提取眼动信息、表情特征、头部朝向等，具有价格低廉，使用方便，与被试者无接触等优点。

### 1.2.1 疲劳度的度量

现有的疲劳度度量主要有生物学专家标注和通过模拟实验中被试者错误率标注这两个方法。

早期的实验往往采用生物学专家标注的方法进行。从生物学角度看，疲劳度一般被定义为从完全清醒到完全睡眠的中间状态。M.T.R. Peiris, et.al 将疲劳度定义为完全清醒、轻度困倦、重度困倦、完全睡眠四个状态 [8]，请四位专家对二十八名航空交通管理员在执行任务中的 EEG (Electroencephalography) 和 EOG (Electrooculography) 信号进行标注。

但是，专家标注得到的参照标准往往不够客观。因此，近些年的疲劳驾驶检测实验往往通过设计实验，研究被试者在实验中的错误率，以此为基准衡量被试者的疲劳度。如 2005 年，Johns M 等人在文献 [9] 中设计了一系列操作任务，通过操作的准确率以及反应延迟时间，衡量疲劳驾驶检测系统的正确性和有效性。

### 1.2.2 基于视频的疲劳驾驶检测算法综述

目前，国际上已经研发出一些疲劳驾驶检测和预警系统，主要采用基于 PERCLOS (Percentage of Closure) 特征 [10] 以及红外线瞳孔定位的方法 [11] 或者基于更加复杂的视频采集技术。它们的共同之处，都是通过计算机视觉的手段，测量驾驶员眼睑的运动特征，以发现驾驶员是否处于疲劳状态。

1994年,日本 Nissan 公司开始利用视频进行疲劳驾驶分析的研究 [12],但是限于当时的硬件条件,他们对驾驶员的动作和表情分析很难达到很高的精度要求。他们定义了三种疲劳状态,并研究了这三种状态与眨眼频率、面部表情之间的关系。

1996年,卡耐基梅隆 (Carnegie Mellon) 大学的驾驶研究中心的 Dinges 等研究员首次提出了 PERCLOS 特征与相关疲劳驾驶检测算法 [10],其核心思想是通过测定眼睛闭合时间占单位窗口时间的百分比,估测被试者的疲劳度。他们的系统采用了双摄像头,利用瞳孔在红外线照射下明显发亮的原理来分析双眼的位置和大小,判断眼睛的闭合程度。如果单位时间窗口内,眼睛闭合度超过 80% 的时间与眨眼总时间的比值超过了一定阈值,则向驾驶员报警。目前,美国公路管理局认定 PERCLOS 方法是一种可靠有效的疲劳检测特征,并在车载疲劳检测系统中得到了一定的商业运用。

### 1.2.3 基于视频的疲劳驾驶系统综述

世界上已研发的疲劳驾驶检测系统并不多,已发明的疲劳驾驶提醒装置主要由红外摄像头、信息处理单元以及报警装置组成,信号来源单一,难以提供本文所述的综合多种特征的疲劳驾驶检测系统的准确率和鲁棒性。

美国 Attention Technologies 公司推出的 Driver Fatigue Monitor (DD850) 是一款基于驾驶员眼部特征的疲劳驾驶预警产品 [13]。该产品通过红外摄像头采集驾驶人眼部信息,采用 PERCLOS 作为疲劳报警指标,可直接安装在仪表盘上,报警的敏感度和报警音量均可调节,目前已推广应用,但价格昂贵,售价约 10,000 人民币,而且面临着判断标准单一的问题。

与此同时, gogo850B 是国内已经商业化的疲劳驾驶预警系统,售价约 8,000 人民币。该系统和 DD850 系统相近,同样基于红外摄像头,主要检测驾驶员的眼睛开合情况,有一定实用性,但同样面临着判断标准单一的问题。

2003年,澳大利亚的 Seeing Machines 公司研发了一套 FaceLAB 系统 [14]。该系统采用了一对立体红外摄像机来测量数据,可以检测到使用者面部的各种特征,包括驾驶员头部朝向、视线方向、眼睑运动、瞳孔直径等。该系统不仅支持 PERCLOS 方法来检测疲劳,还可以分析其他眼睑运动特征,以作为疲劳检测标准。该系统适用于各种脸型,眼睛和多数眼镜,并可以在阳光下或暗房中工作,受试者可以距离摄像机几十公分到几米远。目前这套系统已发展至第



五代，但依旧面临着成本昂贵，室外环境鲁棒性不够强的诟病。

此外，国际市场上还有 SafeTRAC(美国 AssistWare Technology 公司开发，售价约 20,000 人民币) [15]、“径卫视觉”疲劳预警器等，但是普遍价格高、安装不便，特征来源单一。

#### 1.2.4 基于其他传感器的疲劳驾驶系统综述

现有的驾驶员疲劳生理特征的检测方法较多，可分为接触式和非接触式检测两大类，主要检测方法有：EEG（脑电信号）检测、EOG（眼电信号）检测、视频信号检测、方向盘的转动幅度及方向盘的握力检测、道路追踪器检测等。

基于脑电的疲劳检测方面, 1993 年, 美国加州大学圣迭戈分校 Swartz 计算神经科学中心就开展了相关实验 [16], 他们侧重于分析疲劳与脑电频谱特征间的对应关系, 以及大脑皮层与疲劳相关的脑区放电方式。2005 年, 台湾新竹交通大学脑科学研究中心发表论文称 [17]: 他们利用脑电频谱特征, 结合模糊神经网络等模式识别算法, 进行了一系列的模拟驾驶环境下的疲劳检测系统研究, 同时与 Swartz 中心就干电极脑电采集技术以及脑电信号处理等问题进行了合作研究, 他们致力于研制便携式的脑电采集设备。近年来, 上海交通大学智能计算与智能系统实验室的 Ruo-Nan Duan 等人基于脑电开展了一系列情绪识别实验 [18], 通过视频材料刺激被试, 基于被试者的脑电特征对情绪进行分类, 取得了一定进展。

基于眼电的疲劳检测方面, 早起研究主要是从基于眼电信号的睡眠阶段分类上衍生出的, 主要研究方向在于眼电特征的自动检测与提取, 以及单一眼电特征与疲劳度的关系探索上, 并没有发挥眼电技术的专长, 往更实用的角度进化。2011 年, Andreas Bulling 等人在国际顶级刊物《模式识别与机器智能 (PAMI)》上发表论文 {bulling2011eye, 通过眼电信号对各类活动中的眼动行为进行了大量实验, 切实证明了眼电技术对于眼动识别的准确率和实用性。2012 年, Zheng-Ping Wei 等从眼电信号中的抽取了眨眼、慢速眼动、快速眼动等 48 种特征 [19], 显著提升了眼电分析警觉度的准确率。此外, 传统眼电有遮挡视线, 使用不便等诟病。而上海交通大学智能计算与智能系统实验室设计的前额眼电则具有价格低廉, 使用便捷的特点。目前的研究已经证明: 眼电不但可以提取到视频中的眼皮开合情况, 此外还可以提取到眼球的运动特征, 如快速眼动、慢速眼动, 以及眼睑水平和垂直运动, 这些特征都与人的睡眠或是疲劳存

在一定的联系，可以作为疲劳检测指标。

笔者对比了一下疲劳驾驶检测系统中视频、眼电、脑电 [20]、握力信号应用优劣，如表1-1所示。

表 1-1 疲劳驾驶检测系统中视频、眼电、脑电信号应用优劣对比

信号	视频	前额眼电	脑电	握力
接触性	无接触	眼电帽接触	脑电帽接触	方向盘接触
准确性	适中	最准确	较低	适中
特征	PERCLOS, 快慢眼动、哈欠、头部朝向	眨眼、快速眼动、慢速眼动、能量等	$\delta$ 波和 $\theta$ 波	统计特征、时序特征、频率特征
信噪比	中	高	低	较高
影响	易受震荡、光照影响	影响较小	易受电磁场影响	易受震荡影响

### 1.3 主要研究方法与创新点

#### 1.3.1 技术路线

本文提出了基于主动外观模型与支持向量机算法，基于深度学习算法的两条技术路线，并针对第一条技术路线完成了产品原型的研发。该产品原型与基于眼电、脑电的疲劳驾驶检测系统一同成功参展 2011 年中国国际工业博览会并获得二等奖；针对第二条技术路线提出了算法思想和实验方案。

本系统通过基于 Cascade 结构和 Adaboost 算法的人脸检测模型和主动外观模型 (ASM) 对实时视频作预处理，通过二值化及图像梯度方法获得 PERCLOS 特征，通过主动外观模型 (ASM) 在一定时间窗口内计算出慢速眼动 (SEM)、快速眼动 (REM)、打哈欠频率、头部朝向等特征，通过 Gabor 变换提取纹理特征并利用 Adaboost 进行降维，最后用线性动力系统 (LDS) 进行去噪，支持向量机 (SVM) 进行回归拟合。

#### 1.3.2 创新点

本文的创新点如下所述：

1. 完成第一个综合 PERCLOS 特征，主动外观模型几何特征，打哈欠特征的基于视频的在线疲劳驾驶检测系统，并开发了智能手机端系统。
2. 第一个设计并完成了综合眼电与视频信号进行疲劳驾驶检测的实验，证明了眼电信号的可靠性与视频信息的丰富性，第一个提出综合二者特征可显著提高疲劳驾驶检测准确率的实验结论，该文章已被 ICONIP 2012 收录。
3. 参与并完成综合几何特征与纹理特征的表情识别实验，并通过特征融合极大提高了预测准确率。
4. 参与并完成综合握力与视频信号进行疲劳驾驶检测的实验。

## 1.4 章节内容概述

全文共分为八章，具体安排如下：

第二章主要介绍本文研发的基于视频的疲劳驾驶检测系统概述，其中，2011 至 2012 年度研发的基于视频和眼电的实验主要基于 NeuroScan 的系统开展按键测试实验，2012 至 2013 年度研究的视频与握力，主要基于真实桑塔纳轿车与虚拟驾驶环境开展实验。

第三章主要介绍本文采用的计算机视觉中用到的预处理算法。主要包括人脸检测、使用主动外观模型进行人脸对齐、实时图像分割、深度学习在人脸识别的前瞻等内容；同时在主动外观模型算法阐述部分，简要介绍了一下基于视频开展的情绪分类实验及结果分析。

第四章主要介绍基于视频的疲劳驾驶检测系统中用到的特征提取算法。主要包括 PERCLOS、慢速眼动 (SEM)、快速眼动 (REM)、打哈欠频率、头部朝向等特征提取算法、眼电中独立成分分离、握力中的时序特征提取算法。

第五章和第六章分别介绍基于视频与握力、视频与眼电信号开展的疲劳驾驶检测实验及结果分析，主要包括实验准备（被试、实验时长、实验背景、实验过程），预测曲线的相关系数与均方误差等结果分析。

第七章进行全文总结，同时对未来基于智能便携设备的疲劳驾驶检测系统进行前瞻性展望，再一次确认本文算法的可行性。

## 第二章 疲劳驾驶检测系统概述

### 2.1 基于真实轿车与虚拟现实技术的模拟驾驶系统

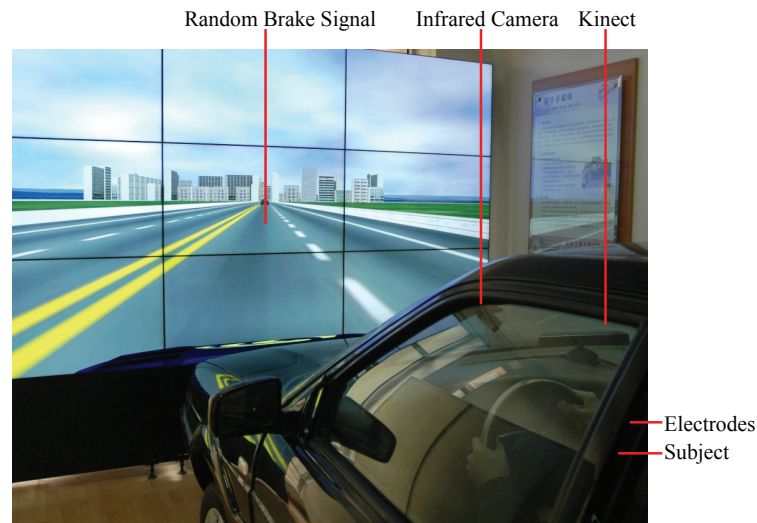


图 2-1 基于真实轿车与虚拟现实技术的模拟驾驶系统实验图

首先，为了探索视频图像与握力方法的区别与联系，第一组实验采用了智能计算与智能系统实验室部署的基于真实轿车与虚拟现实技术共同搭建的模拟驾驶系统。这组实验每次时长 90 分钟，旨在以被试者对踏板信号的反应速度与失误率作为基准，通过同时提取视频中的人脸图像特征与驾驶员握方向盘的握力大小与位置信号，估测被试者疲劳程度。

如图2-1所示，实验设备主要由一辆桑塔纳型轿车（内含罗技人机交互方向盘、压敏传感器、踏板、加速杆等）、九块显示器搭建的显示屏、一只红外摄像头、一只 Kinect 深度摄像头等组成。通过三维虚拟现实技术，在 9 块组合显示器上模拟出被试者的驾驶环境；每隔 20 秒，屏幕正前方会以一定概率随机出现刹车信号。当刹车信号出现，我们要求被试者迅速踩下安有压力传感器的离合器踏板。通过计算用户的反应时间和失误率，适当结合路况信息，可以计算出被试者的疲劳程度。方向盘上安置有压敏传感器，用于接受被试者

的握力信号；被试者正前方安置有红外摄像头和 Kinect，用于录制视频画面。红外摄像头保证了夜间的正常运作，Kinect 丰富了视频的深度信息。这样，如图2-2所示，握力信号、视频信号可以同时采集，便于后续分析。



图 2-2 基于握力与视频的疲劳驾驶检测系统监视图

## 2.2 基于 NeroScan 与红外摄像的疲劳驾驶训练系统

此外，为了探索视频图像与眼电方法的区别与联系，第二组实验采用了仿脑计算与人工智能中心的 NeuroScan 系统。这组实验每次时长 76 分钟，旨在以被试者按键测试中的错误率为基准，通过同时提取视频中的人脸图像特征与驾驶员的眼电特征，估测被试者的疲劳程度。

如图2-3所示，实验设备主要由显示器、红外摄像头、眼电电极、放大器、装有四个颜色按键的控制面板等组成。在实验过程中，显示器上会随机呈现出图2-4中所示的红、黄、蓝、绿四种颜色的交通标志，每种标志闪现 500 ms，之后黑屏 5 至 7 秒，如此循环往复。当标志闪现时，我们要求被试者迅速反应按下正确的按键，当被试者因困倦而导致精力分散时，会产生一定错误率。通过同时采集眼电信号与红外视频信号，我们可以进一步对图像、眼电特征与被试者疲劳程度之间的关联进行分析。

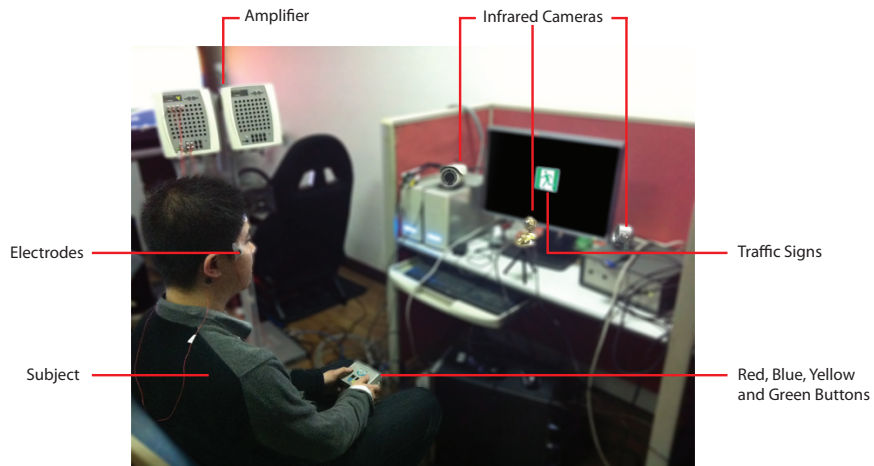


图 2-3 基于 NeroScan 与红外摄像的疲劳驾驶训练系统实验图



图 2-4 按键实验所用四种交通标志示意图

## 2.3 本章小结

本章主要简述了论文中基于真实桑塔纳轿车与虚拟现实技术的模拟系统，以及基于 NeuroScan 软件的按键测试实验系统，为后文实验章节做背景介绍。

## 第三章 基于计算机视觉的预处理算法

由于实际环境中，摄像头捕捉的画面可能存在不止一个人脸，加之光照、明暗、对比度等因素影响，我们需要将摄像头捕捉的画面进行人脸识别、主动外观模型、局部正则化等处理，以便于特征提取阶段的进一步分析和检测。

### 3.1 基于主动外观模型的技术路线

#### 3.1.1 人脸检测算法

本实验初步采用 Viola 2004 年提出的采用 Haar 特征和 cascade [21] 的 Adaboost 分类器融合的人脸检测算法 [22] 进行快速人脸检测，对于每一帧图像，检测出人脸位置，通过前后帧图像对比，对人脸位置在邻近区域进行跟踪，当跟踪人脸丢失，可以再起启动人脸检测模块，这样的策略保证了系统的高效运行。

图 3.1.1 中展示了整个预处理的流程图，在第一步我们使用基于类 Haar 特征的 Adaboost 人脸检测算法定位出图片中的人脸区域，然后在这个范围内我们使用主动外形模型进行搜索，定位特征点的位置。在最后一步中，我们需要对人脸进行标准化，并缩放到一个统一的大小，以便于提取 Gabor 系数。在情绪识别的实验中，我们将人脸区域缩放到  $48 \times 48$  的大小，在疲劳驾驶检测系统中，我们的系统使用原始图像大小，可以



图 3-1 预处理流程图

### 3.1.2 主动外观模型算法

此外，我们使用主动外观模型实现人脸区域重要位置定位。1995年，Cootes等人提出了主动形状模型（Active Shape Model，简称ASM）算法[23]。该算法采用参数化的采样形状来构成对象形状模型，并利用主成分分析（Principle Component Analysis，简称PCA）算法对形状控制点的概率分布进行分析，最后建立描述形状的控制点的运动模型。1998年，Cootes等人继续利用了对应的纹理信息，提出了主动外观模型（Active Appearance Model，简称AAM）来学习物体形状和内部构造[24]。

，即关键位置，然后训练人脸形态模型，之后就可以通过训练好的模型在检测时使用形状、纹理等信息将人脸模型与图片中的人脸匹配，最终得到一个最贴合图片中人脸的模型。这样就可以得到比较精确的关键点定位，从而获得眼睛的位置。

主动外观模型一般通过诸如梯度下降的优化算法来拟合图像中的几何特征。应用主动外观模型检测人脸重要位置，首先需要在大量数据集中标注需要训练的形状的控制点的位置。主动外观模型中定义的形状  $\mathbf{S}$  通常由  $n$  个控制点确定，记为  $\mathbf{S} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$ ，其中  $\mathbf{a}_n = (x_i, y_i)$  是第  $i_{th}$  个控制点。在我们的疲劳驾驶检测系统中，训练了有 68 个标定点的主动外观模型，包括了如眼睛，嘴角，鼻子，眉毛等器官的关键位置。此外，我们有：

- 平均值： $\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$
- 欧氏距离： $D(i, j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$
- 协方差矩阵： $= \frac{1}{N} \sum_{i=1}^N d\mathbf{x}_i d\mathbf{x}_i^T$
- 特征值和特征向量： $S_k = \lambda_k \alpha_k, k = 1, \dots, 2n$

这样，任意的形状可以用如下形式表示：

$$\mathbf{S} = \bar{\mathbf{S}} + \mathbf{A} \cdot \mathbf{b} \quad (3-1)$$

其中， $\mathbf{A} = (\alpha_1, \alpha_2, \dots, \alpha_{2n})$  是特征向量的集合，而  $\mathbf{b} = (b_1, b_2, \dots, b_{2n})^T$  是每个特征向量的权值，也即任意形状均可表示为平均形状和特征向量的加权和。



为了测试主动外观模型算法效果，为接下来的疲劳驾驶检测实验做铺垫，我们在 CK+ 人脸表情数据库 [25] 中做了表情识别实验，该数据库中表情种类分布如表3-1所示。这个数据库中包含了 123 名被试者的 593 组视频序列，每个视频序列都是从自然表情开始，直到被试者按照要求做出了相应的表情或者脸部运动单元 (Facial Action Unit, 简称 FAU)。每个视频均为 25 帧每秒，持续时间长短不一，为 1 到 3 秒。也就是说，每个视频序列可以提取出 30 至 70 帧左右的图像。

表 3-1 CK+ 数据库中表情种类的分布数量

表情	高兴	害怕	生气	轻蔑	厌恶	悲伤	吃惊
数量	69	25	45	18	59	28	83

最后，我们使用模糊积分将基于主动外观模型的几何分类器，以及基于 Gabor 小波和流行学习降维后的纹理特征合并起来，实验结果如表 3-2 所示，证明了在大多数分类结果上我们的算法都比 2010 年文献 [25] 中提供的分类算法要优。

### 3.1.3 基于 Paint Selection 的图像实时分割算法

实验之外，本文作者提出使用 Paint Selection 来进行对人体器官的精确实时分割。首先，基于主动外观模型，我们已经可以得到器官的控制点。但控制

表 3-2 模糊积分算法合并两种分类器的实验结果

	生气	轻蔑	厌恶	害怕	高兴	悲伤	吃惊
生气	<b>100.00</b>	0.00	0.00	0.00	0.00	0.00	0.00
轻蔑	16.67	<b>50.00</b>	0.00	0.00	0.00	33.33	0.00
厌恶	0.00	0.00	<b>100.00</b>	0.00	0.00	0.00	0.00
害怕	0.00	0.00	0.00	<b>87.50</b>	12.50	0.00	0.00
高兴	0.00	0.00	0.00	0.00	<b>100.00</b>	0.00	0.00
悲伤	0.00	0.00	0.00	0.00	0.00	<b>100.00</b>	0.00
吃惊	3.57	0.00	0.00	0.00	0.00	0.00	<b>96.43</b>



图 3-2 基于 Paint Selection 的图像实时分割算法实验图

点并不能精确到每个像素。因此，我们采用了 Jian Sun 于 2010 年提出的 Paint Selection 算法来实时分割图片，效果如图3-2所示：

该算法无论从实用性还是从准确率来讲，都要比 Carsten Rother 等人 2004 年于 SIGGRAPH 上发表的著名的 GrabCut 算法 [26] 要优。GraphCut 在使用时，需要标定一个区域面积。而 Paint Selection 仅靠待分割物体的特征点即可获得较优的分割效果。首先，它通过对背景随机采样（通常为 1200 个点），建立一个符合高斯混合分布的 8 变量颜色模型  $p^b(\cdot)$ 。在特征点上取样，同样建立一个颜色模型。这样，特征点附近的像素的标签预测问题  $X = x_p$  可以等价为一个能量最小化问题（Boykov 和 Jolly 在 2011 年提出）：

$$E(X) = \sum_p E_d(x_p) + \lambda \sum_{p,q} E_c(x_p, x_q) \quad (3-2)$$

其中  $\lambda$  为权值（默认为 60）， $E_d(x_p)$  为数据项，表示像素  $p$  的标签为  $x_p$  (1-foreground), 0-background) 的代价； $E_c(x_p, x_q)$  为对比项，代表两个邻近像素  $p$  和  $q$  的标签代价，在本算法中，我们采用如下的对比项：

$$E_c(x_p, x_q) = |x_p - x_q| \cdot (\beta \cdot \|I_p - I_q\| + \epsilon)^{-1} \quad (3-3)$$

其中,  $\epsilon = 0.05$ ,  $\beta = (\langle \|I_p - I_q\|^2 \rangle)^{-1}$ , 这里  $\langle \cdot \rangle$  是整幅图片的期望算子 [27]。

## 3.2 基于深度学习的技术路线

### 3.2.1 深度信念网

人工神经网络 (Artificial Neural Network, 简称 ANN) 的早期研究工作起源于 1943 年, 著名心理学家 W. McCulloch 和数理逻辑学家 W. Pitts 在文献 [28] 中通过分析神经元的基本特性, 率先提出了神经元的数学模型, 被称为神经网络研究的先驱。

1986 年, 科学家 Rumelhart 和 McClelland 提出了误差反向传播算法 (Back Propagation, 简称 BP) [29], 该算法通过误差逆传播算法来训练多层前馈神经网络, 是目前应用最广泛的神经网络模型之一。[30]

1988 年, LeCun 提出了用卷积神经网络 (Convolutional Neural Network, 简称 CNN) 训练手写体数字识别 [31], 该算法是深度学习的先驱, 是历史上第一个真正多层结构学习的算法, 它利用了增加神经网络的深度, 减少了参数数目, 显著提高了神经网络的准确率。

2006 年, 多伦多大学的 Hinton 革命性地提出了深度信念网 (Deep Belief Networks, 简称 DBN) [32]。传统人工神经网络、支持向量机 (SVM) 算法往往只有 2 的深度; 但这样带来的后果是, 网络中需要的节点数 (参数数量) 可能会非常大。反观人类的大脑, 其深度和复杂性则远远超过了传统神经网络, 而人类的认知过程往往是一个有深度的, 循序渐进的过程。处于这样的动机, Hinton 设计了深度信念网, 其主要原则如下:

1. 用无监督学习预训练每一层。
2. 每次迭代只训练同一层, 该层的输出作为下一层的输入。
3. 最后用反向传播算法调整每一层的参数。

笔者基于 MNIST 手写体数字数据库 [33]，使用深度信念网训练了一个有 100 个隐藏结点，其准确率可以达到 97.51%。经过无监督学习和反向传播算法，100 个隐藏结点最后得到的权值如图3-3所示。

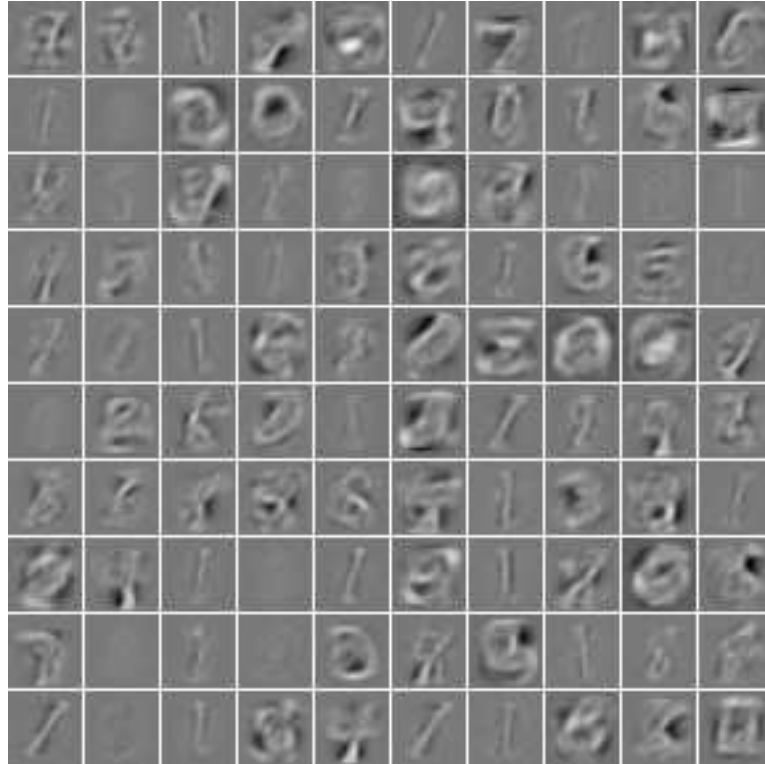


图 3-3 基于深度信念网的手写体数字识别训练权值

### 3.2.2 用于人脸检测的深度学习网络

2012 年，Ping Luo 提出了用深度学习的方法来进行人脸检测 [34]，其核心思想是：采用深度信念网（Deep Belief Network），边检测器作为根节点，连通块检测器、器官检测器作为上层结点，人脸检测器作为顶层结点，通过无监督学习的训练，逐步调整每一层检测器的参数。

该算法首先需要利用基础的边检测器和分块检测器，将图片分割成若干分块，设每个分块为  $I_1, I_2, \dots, I_n$ ，该算法通过深度为  $K + 1$  的深度训练网刻画每个分块  $I_b$  与  $K$  个隐藏层  $\mathbf{h}^1, \dots, \mathbf{h}^k$  之间的关系：

$$P(\mathbf{I}_b, \mathbf{h}^1, \dots, \mathbf{h}^k; \rho_{dbn}) = \prod_{k=0}^{K-2} (P\rho(\mathbf{h}^k | \mathbf{h}^{k+1} | \rho_{dbn}) P(\mathbf{h}^{K-1}, \mathbf{h}^K; \rho_{dbn})) \quad (3-4)$$

其中,  $\mathbf{I}_b = \mathbf{h}^0$ , 而  $P(\mathbf{h}^k | \mathbf{h}^{k+1}, \rho_{dbn})$  代表第  $k$  层 RBM 上的隐藏独立分布,  $P(\mathbf{h}^{K-1} | \mathbf{h}^K, \rho_{dbn})$  为最顶层的联合分布。而算法的核心就是通过深度信念网, 无监督地学习每一层的参数  $\rho = (\rho_{dbn}, \rho_{reg})$ 。首先, 预训练  $\rho_{dbn} = \{\mathbf{W}^i, \mathbf{u}^i, \mathbf{z}^i\}$ , 随机初始化  $\rho_{reg} = (\mathbf{W}^r, \mathbf{u}^r)$ , 通过对数回归来调整参数, 使之更优, 最终达到收敛。

由于深度学习训练时间过长, 迫于时间关系, 笔者没有进行该组实验。本小节意在抛砖引玉, 激励后来人通过深度学习的方式, 力争进一步接近生理上疲劳驾驶特征的本质, 主动无监督学习此类特征, 达到更高的准确率和鲁棒性。

### 3.3 本章小结

在本章中, 我们详细阐述了基于计算机视觉的预处理算法, 重点阐述了主动外观模型在疲劳驾驶分析中的原理和重要作用。同时, 创新地提出了使用 `PaintSelection` 进行人脸图像的分割, 以及使用深度学习的方法在未来对大规模数据进行无监督学习, 通过大量的边检测器构筑深度信念网, 从而主动学习特征, 达到更高的准确率。

## 第四章 疲劳驾驶检测系统的特征提取算法

本章主要阐述我们在疲劳驾驶检测系统使用的特征提取算法。首先，重点介绍本文作者提出的基于视频信号的特征提取算法，诸如 PERCLOS、眨眼、打哈欠、头部朝向等特征。其次，本章会涵盖现有的主要疲劳驾驶检测系统、疲劳驾驶检测方法，并讨论它们的优缺点和亟待解决的问题。最后，本章简述一下笔者所开发的基于视频的疲劳驾驶检测系统的主要技术路线和创新点，介绍一下后面的各章节安排。

### 4.1 基于视频信号的特征提取

#### 4.1.1 PERCLOS 特征

1996 年，卡耐基梅隆 (Carnegie Mellon) 大学的驾驶研究中心的 Dinges 等研究员首次证明，驾驶员在疲劳驾驶的时候，单位窗口时间内的眼睛闭合时间 (PERCentage of CLOSure, 简称 PERCLOS) 较大程度地反映了驾驶员的疲劳程度 [2]。该特征定义为单位窗口时间内，眼睛闭合一定比例所占的持续时间：

$$PERCLOS_i = \sum_i \frac{\bar{H}_i - H_i}{\bar{H}_i}$$

PERCLOS 判据常常采用单位窗口时间内眼睛闭合程度超过某一阈值的时间占总窗口时间的百分比。具体实验中，有 P70 (眼睛闭合面积 70% 以上的时间百分比)，P80 (眼睛闭合面积 80% 以上的时间百分比)，EM (眼睛闭合速度百分比的均方值) 三种测量方式，其中 P80 被认为最能反映人的疲劳程度。本系统采用 P80 的改进算法作为 PERCLOS 得分。

通过主动外观模型，我们已经可以得到人眼睛的大致位置。但为了精确测量 PERCLOS 特征，在传统主动外观模型的基础上，我们使用了直方图二值化，通过梯度算法寻找灰度值变化的极大点，从而得到上下眼皮精确到像素级别的位置。如图 4-1，首先我们进行图像分割，这一步也可以利用第三章阐述的 PaintSelection 来进行人眼精确分割。其次通过二值化，将上下眼皮的分界点最大化，最后，通过梯度曲线一阶导数的两个波峰，确定上下眼皮的精确位置。

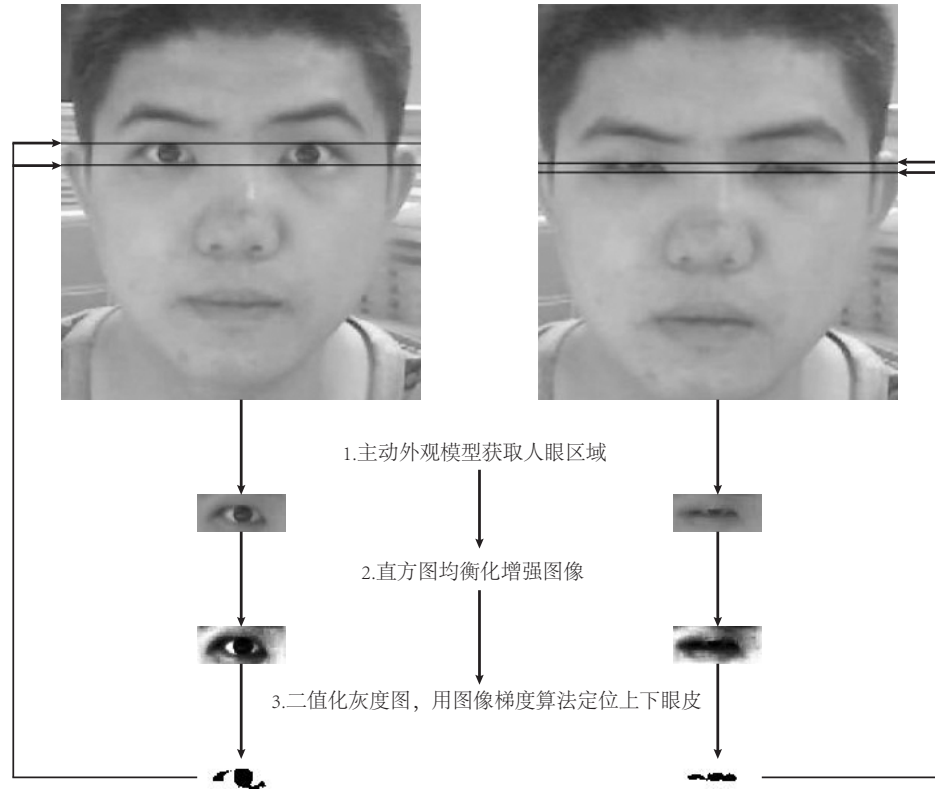


图 4-1 直方图二值化和图像梯度算法流程图

#### 4.1.2 眨眼特征

通过上下眼皮距离变动的频率,我们可以计算是单位时间窗口内的眨眼次数,每次眨眼的平均持续时间(眨眼速度),眼动能量等具体特征。系统记录眼睛刚要闭合,完全闭合,刚要睁开,完全睁开时的时间戳(Time Stamp):  $T_{c1}, T_{c2}, T_{o1}, T_{o2}$ ,进而得到眨眼速率,闭眼速率,眨眼能量,眨眼频率等特征:

$$T_{blink} = T_{o2} - T_{c1}; \quad T_{close} = T_{c2} - T_{c1}$$

$$T_{open} = T_{o2} - T_{o1}; \quad T_{closed} = T_{o2} - T_{c2}$$

$$S_{close} = \frac{\sum_{i=T_{c1}}^{T_{c2}} D_i}{T_{close}}; \quad S_{open} = \frac{\sum_{i=T_{o1}}^{T_{o2}} D_i}{T_{open}}; \quad E_{blink} = \sum_{i=T_{c1}}^{T_{o2}} V_i^2$$

### 4.1.3 打哈欠特征

众所周知，打哈欠是驾驶员困倦的一个明显特征表现，考虑到打哈欠的时间可能会持续一段时间，而且对疲劳驾驶的预测有较深远的影响，哈欠的时间窗口（Time Window）要明显大于眨眼等眼动算法。我们将打哈欠的时间窗口设置为 16 秒，将时间窗口内前  $k$  最小嘴高度记为  $H_m^k$ ，则哈欠频率特征表述为：

$$Y_i = \frac{\sum_{j=i-w}^i (H_j / H_m^k) > C}{w}$$

考虑到实际操作中，驾驶员说话时也会造成嘴部张开百分比变大。行车时禁止驾驶员说话也是不科学的做法。我们注意到：说话与打哈欠这两个动作有着一个明显的区别，嘴巴的上下开合频率不同。说话时，驾驶员因发音不同，嘴部会以较快的频率上下张合；打哈欠时，驾驶员会有 2-3 秒的时间处于嘴巴大幅度张开的状态。通过嘴部高低变动的频率，我们可以过滤到高频的嘴部变化特征，排除因驾驶员说话造成的干扰。

### 4.1.4 头部朝向特征

头部朝向的变化是驾驶员注意力是否集中的重要体现。实验中，我们使用了主动外观模型，通过关键点比例的变化来粗略估测头部朝向的三维变化特征。

$$\alpha = \frac{D(67, 2)}{D(67, 12)}; \theta = \frac{D(31, 0)}{D(36, 14)}; \beta = \frac{D(66, 3)}{D(66, 11)} \quad (4-1)$$

由于实验中，本文没有使用 Kinect 摄像头，现有的三维重建算法无法实时地对人脸进行准确的三维重建，故只能用二维的方法近似估测头部朝向的变化趋势。

## 4.2 基于眼电信号的特征提取

### 4.2.1 传统眼电与前额眼电

在传统眼电的基础上，项目组采用了前额眼电作为传统眼电的改进。如图 4-2 所示，本实验使用了四个电极，以等间距的方式，水平摆放于前额；同时使用了两个电极置于一只眼睛的上下方记录传统垂直眼电，使用两个电极置



于两眼外侧用于记录传统水平眼电，最后两个电极置于耳后，作为参考电极和接地电极。



图 4-2 前额眼电与水平眼电电极放置与比较示意图

#### 4.2.2 独立成分分析

独立成分分析 (Independent Component Analysis, 简称 PCA), 是一种统计学上将信号源分离成统计独立, 非高斯信号源的信号的线性组合。这个问题最早来源于鸡尾酒问题, 即语音学中如何分离出鸡尾酒会中同时讲话的每个人的独立语音信号。独立成分分析要求待分析的信号统计上互相独立。在眼电信号中, 独立成分分析被用于分离互相独立的水平眼电和垂直眼电, 得到了非常好的效果。我们采用了 FastICA[35] 作为独立成分分析的算法。该算法采用最大熵原理, 搜索最优的非线性函数来从混合信号中分离出统计独立的信号, 具有分离效果好, 收敛速度快等特性。

#### 4.2.3 垂直眼电特征提取

我们使用 EEG Lab[36] 中的滤波算法, 对 ICA 分离后的垂直眼电信号进行了 10 Hz 的低通过滤。如图4-3, 我们使用差分算法来提取眨眼特征, 设  $D$

为差分信号， $V$  为滤波后的垂直眼电信号， $R$  为采样率，则有

$$D(i) = (V(i + 1) - V(i)) \times R \quad (4-2)$$

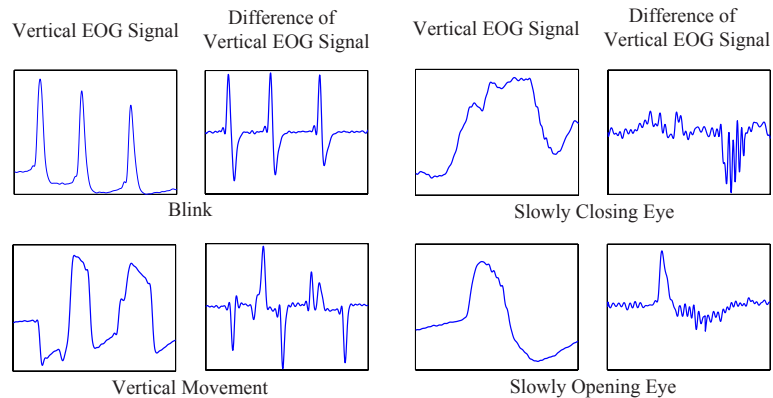


图 4-3 利用差分算法在垂直眼电信号提取特征

#### 4.2.4 水平眼电特征提取

水平眼电上，主要采取了快速傅立叶变化以及小波变换的方法。在傅立叶变换中，我们在 125Hz 采样率的信号上使用了 0.5Hz 的低通滤波器来滤波。使用文献 [19] 中的算法提取水平眼电中的相关眼动特征。

### 4.3 基于握力信号的特征提取

#### 4.3.1 握力信号的采集与转换

如图4-4所示，我们采用了压敏传感器来采集握力信号。首先，将压敏传感器包裹在方向盘上，通过转换电路，将随着压力变化的电阻值转换为相应的电平信号。我们采用了高精度的模拟数字（Analog Digital，简称 AD）采样芯片，集成在嵌入式主板上，实时地采集驾驶员握力信号，供疲劳驾驶检测系统作为源信号来进行分析。

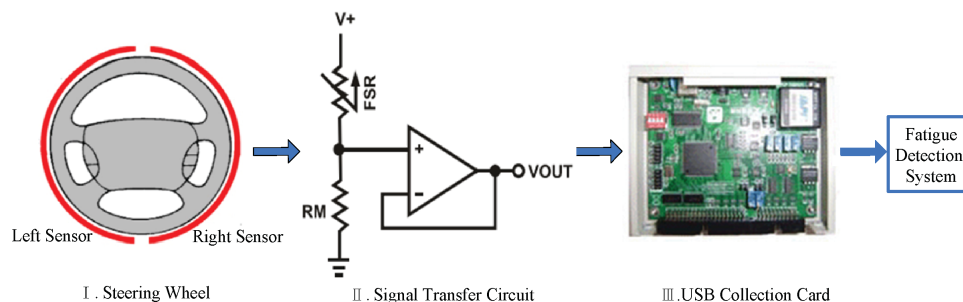


图 4-4 握力信号的采集与转换原理示意图

### 4.3.2 握力信号预处理

在行车过程中，通常驾驶员对于方向盘的握力都是平稳变化的，即使在驾驶员陷入疲劳的状况下，驾驶员的握力也是一个迟缓的过程。因此，握力信号往往都是集中在低频有效。信号进过转换电路和采集电路后往往会引入噪声信号，为此，我们使用高频滤波的方式去除噪音信号。但滤波往往会增大系统的计算量，最终，我们考虑设计专用的硬件滤波电路来有效地去除噪声信号。

### 4.3.3 握力信号中的特征提取

从握力传感器中，我们只能得到方向盘握力大小和握力角度两个一维信号，相比视频信号和言电信号，可供疲劳驾驶检测系统分析的特征较少。为此，我们在时域和频域上对信号进行了进一步分析，来提取疲劳相关特征。为了能够实时地预测疲劳，我们选取当前时刻往前的一个合适的时间窗口的数据进行分析。可以利用短时傅里叶变换和小波变换进行时频域特征的提取。而利用主成分分析（PCA）可以对特征进行降维，来滤除无效的特征量。最终采用了时序上的最大值、最小值、平均值、标准差、自相关系数、小波阶数、各层小波的能量、各层小波正系数的比例作为最终特征。

## 4.4 本章小结

在本章中，我们详细阐述了基于视频信号、眼电信号、握力信号的主要特征提取方法和手段。对于视频特征，我们使用的是 PERCLOS、眨眼、眼动、打哈欠和头部朝向等特征；基于眼电信号，我们使用了独立成分分析抽出水平

眼电和垂直眼电，在时间窗口上分析眨眼和眼动特征，并进一步使用快速傅立叶和小波变换来滤波；基于握力信号，我们采用的是时序上的基本量、小波能量等特征。

## 第五章 基于视频与握力信号疲劳驾驶模拟实验

本章介绍基于真实轿车与虚拟现实技术中的疲劳驾驶检测实验及结果分析，该实验是多模态疲劳驾驶检测系统中关于视频与握力的一个小实验。从时长 270 分钟的视频以及握力信号数据中，我们共提取了 PERCLOS、快速眼动、慢速眼动、哈欠频率、头部朝向等图像特征与 43 维握力的统计、时序、频率特征，通过径向基函数的支持向量机进行拟合。实验表明，视频特征在警觉度的预测结果上好于握力特征，而二者融合的预测可以一定程度上弥补由于光照、晃动因素导致的视频特征损失。

### 5.1 实验概述

#### 5.1.1 实验准备

本实验邀请了五名被试者，其中四名男士，一名女士，均为 23 岁左右。所有被试者均无色盲表现。应实验要求，被试者在实验前一天晚上正常睡眠 7 个小时，实验均在晚饭后 20 点至 23 点进行，用时 90 分钟。这样，被试者一般表现出前半段清醒，后半段进入困倦甚至睡眠状态。实验房间为无声环境，光线柔和，湿度和温度适中。

#### 5.1.2 实验过程

每名被试者被要求在真实轿车内，通过虚拟现实技术，操纵方向盘和刹车踏板来模拟驾驶。每隔 20 秒，正前方会以一定概率随机出现刹车信号。当刹车信号出现，我们要求被试者迅速踩下踏板。通过计算用户的反应时间，适当结合路况，我们可以得出被试者的疲劳程度。方向盘上安置有握力传感器，被试者正前方安置有红外摄像头和 Kinect，这样，握力信号和红外与深度视频信号可以同时采集，便于后续分析。



图 5-1 基于视频与握力信号驾驶员的疲劳驾驶实验环境示意图

## 5.2 实验分析

我们采用了主动外观模型对视频进行实时特征提取，实验效果图参见图5-2：

我们采用 LibSVM[37] 来对被试者数据进行训练与测试。由于设备缘故，前几名被试者测试时握力记录有所缺失，最终我们决定使用同一人的三次试验进行训练与测试，以此作为一个初步的探究试验。我们以 10 秒为一个时间窗，共采集 400 个时间戳的视频与握力特征。其中两次实验数据作为训练集，另外一次实验数据作为测试集进行交叉验证。支持向量机采用径向基核函数 (Radial Basis Function, 简称 RBF Kernel)，参数如下：

$$s = 3(\epsilon - \text{SVR}), \quad t = 2(\text{RBF kernel}), \quad c = 1/2, \quad g = 1/128, \quad p = 1/512$$

每名被试者通过支持向量机预测的疲劳度与踏板实验错误率的相关系数与均方误差如表5-1：

以 1、2 组实验为训练集，第 3 组实验为测试集为例，如图5-3，可以看到，视频方法已经可以提供一定的预测精度，然而，在光线突然变暗且伴随驾驶员进入困倦状态时，握力信号反而比视频信号更加可靠 (图5-4)。如图5-5，通

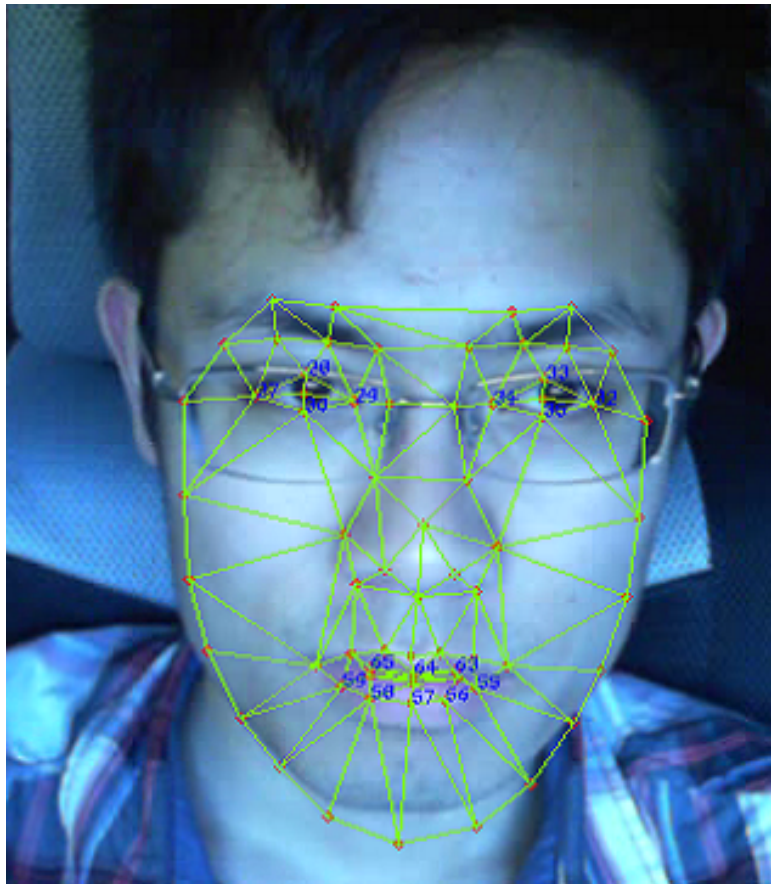


图 5-2 训练后的主动外观模型实验效果图

表 5-1 基于视频、握力的视频握力的疲劳度预测相关系数与均方误差

测试数据	视频方法	握力方法	二者结合
1	0.771/0.0516	0.642/0.0815	<b>0.852/0.0109</b>
2	0.822/0.0229	0.791/0.0212	<b>0.867/0.0075</b>
3	0.861/0.0159	0.738/0.0703	<b>0.909/0.0059</b>
平均值	0.818/0.0368	0.723/0.0843	<b>0.876/0.0081</b>

过结合握力信号与视频信号的信息，可以一定程度上提升总体疲劳驾驶检测的准确率。

### 5.3 本章小结

在本章，我们阐述基于视频与握力信号的驾驶员疲劳驾驶模拟实验，该实验证明了：我们的疲劳驾驶检测系统可以对被试者真实警觉度和操作错误率提供较准确的预测值。此外，我们通过将信息更加丰富的视频信号和表征生理特征的握力信号结合，相对于通过单一信号源进行疲劳驾驶检测的方法，有着更加精准的预测结果。今后，我们希望结合更多的信息，如眼电信号、深度信息来开展实验，进一步增强疲劳驾驶检测的鲁棒性与准确性。



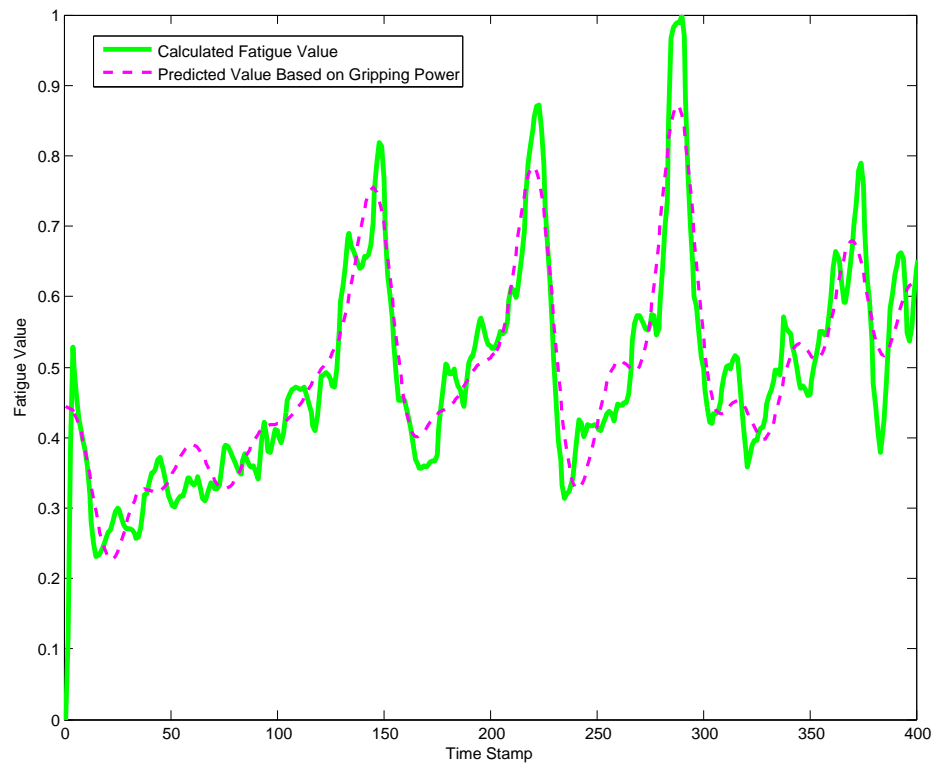


图 5-3 基于视频特征疲劳驾驶检测结果分析图

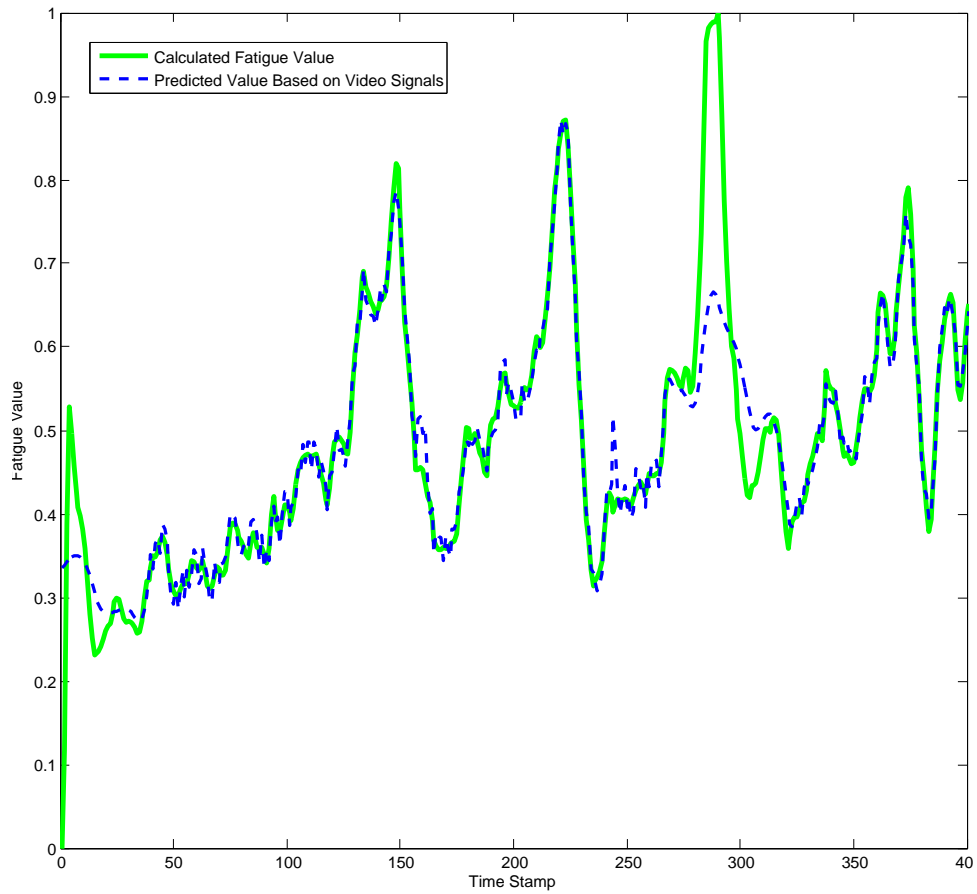


图 5-4 基于握力特征疲劳驾驶检测结果分析图

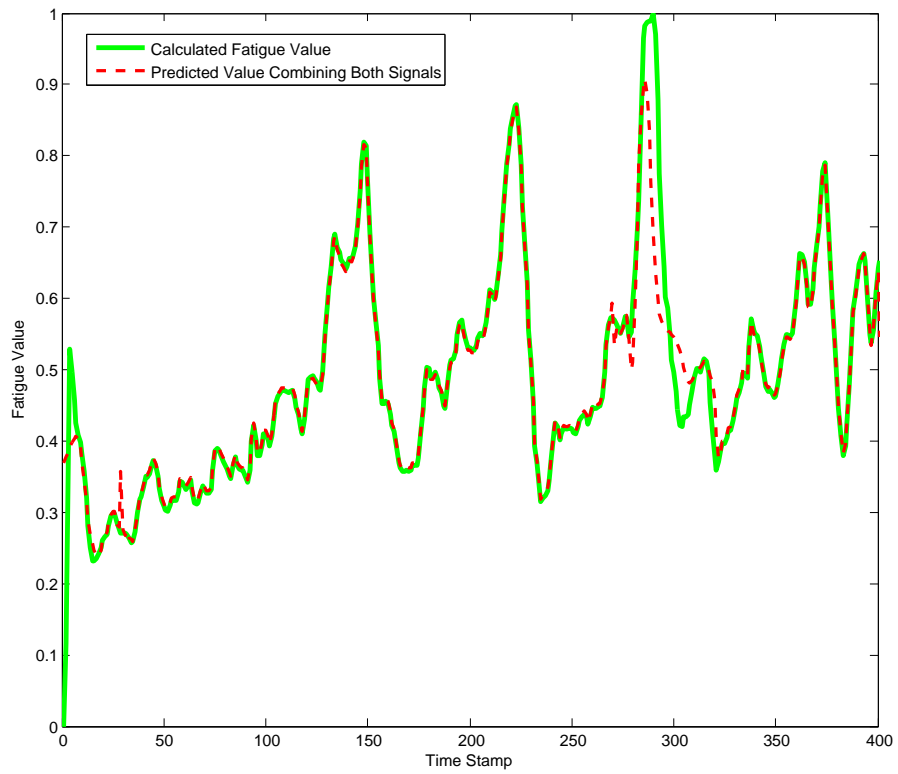


图 5-5 结合视频特征与握力特征疲劳驾驶检测结果分析图

## 第六章 基于视频与眼电信号的疲劳驾驶模拟实验

本章介绍基于视频与眼电信号开展的疲劳驾驶检测实验及结果分析，该实验是多模态疲劳驾驶检测系统中关于视频与眼电的开创性实验，从时长 380 分钟的 5 组视频以及眼电信号、按键测试数据中，我们共提取了 17 维图像特征与 48 维眼电特征，通过径向基函数的支持向量机进行拟合。实验表明，在按键测试实验中，眼电特征对于被试者警觉度的预测结果上好于视频特征，而二者融合的预测可以一定程度上弥补由于光照、晃动因素导致的视频特征损失。

### 6.1 实验概述

#### 6.1.1 实验准备

本实验邀请了五名被试者，其中四名男士，一名女士，均为 23 岁左右。所有被试者均无色盲表现。应实验要求，被试者在实验前一天晚上正常睡眠 7 个小时，实验当天早晨早起。实验均在中午午饭后进行，用时 67 分钟。这样，被试者一般表现出前半段清醒，后半段进入困倦甚至睡眠状态。实验房间为无声环境，光线柔和，湿度和温度适中。

#### 6.1.2 实验过程

每名被试者被要求在屏幕前做按键测试实验。通过 NeuroScan 软件，我们可以得到被试者在实验中的错误率和相应的眼电信号；与此同时，我们通过红外线摄像头采集被试者的视频信号。实验前，我们需要打开红外摄像头，调整方向，使得被试者完整出现在镜头中；同时在 NeuroScan 中观察各眼电电极阻抗是否为蓝色或者黑色，即阻抗达到要求，最后将 NeuroScan 软件切换到信号图，观测实时信号是否正常，并开始记录。

系统流程图如图 6-1 所示。

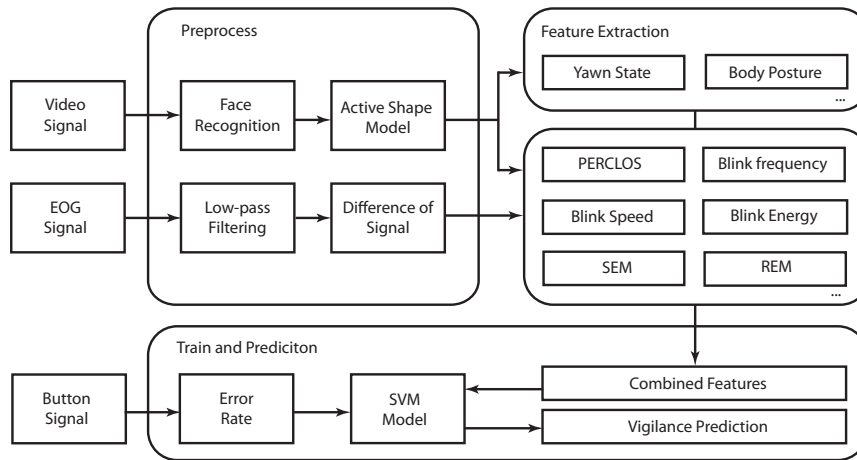


图 6-1 系统流程图

## 6.2 实验分析

### 6.2.1 实验结果

采集数据后，我们按照前面章节的算法进行特征提取，使用线性动力系统 (Linear Dynamic System, 简称 LDS) 进行特征平滑。线性动力系统是一个非监督学习的算法，可以增强特征中的主要部分，大幅降低噪音，从而显著提升特征与警觉度的相关系数。石立臣等通过线性动力系统将无标号的脑电信号噪声进行了很好的除去。在动力系统模型中，空间中的点随着时间的变化状况存在着固定的规则；而线性动力系统中观测变量和隐变量都是符合高斯分布的连续变量，可以对有噪声的信号进行很好的过滤：

$$P(x_t|z_t) = N(x_t|z_t + \bar{w}, Q)P(z_t|z_{t-1}) = N(z_t|Az_{t-1} + \bar{v}, R) \quad (6-1)$$

因此，考虑到视频图像中本身会有获取过程中的高斯噪声，我们采用了线性动力系统来解决，针对眨眼特征的效果图如图6-2：

最后，我们采用 LibSVM[37] 来对被试者数据进行训练与测试。每名被试者以 10 秒为一个时间窗，共采集 400 个时间戳的视频与眼电特征。所有数据被平均分成等长的两份，一份作为训练集，一份作为测试集。支持向量机采用

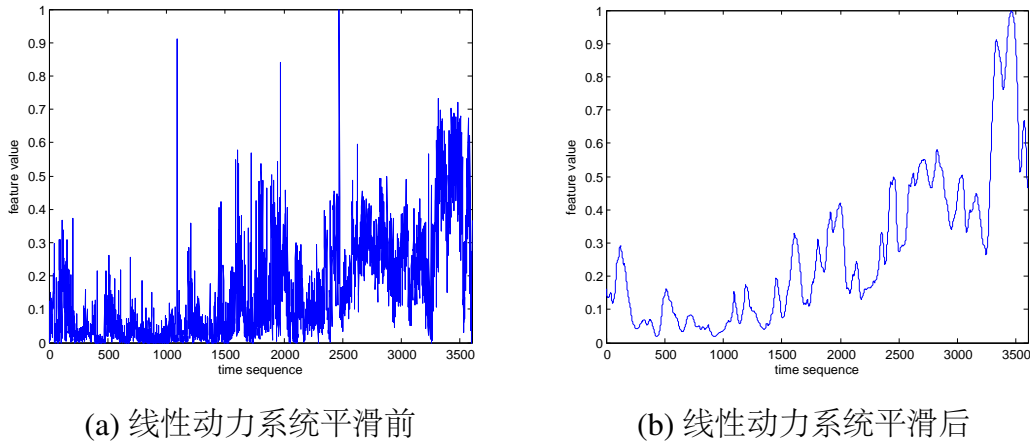


图 6-2 线性动力系统特征平滑示意图

径向基核函数 (Radial Basis Function, 简称 RBF), 参数如下:

$$s = 3(\epsilon - \text{SVR}), \quad t = 2(\text{RBF kernel}), \quad c = 8, \quad g = 1/64, \quad p = 1/1024$$

每名被试者通过支持向量机预测的疲劳度与按键测试错误率的相关系数与均方误差如下表:

表 6-1 基于视频、眼电、视频眼电的疲劳度预测相关系数与均方误差

被试者	视频方法	眼电方法	二者结合
1	0.731/0.0256	0.843/0.0136	<b>0.852/0.0117</b>
2	0.778/0.0129	0.892/0.0064	<b>0.919/0.0170</b>
3	0.750/0.0151	0.866/0.0148	<b>0.882/0.0111</b>
4	0.750/0.0175	0.929/0.0091	<b>0.937/0.0045</b>
5	0.756/0.0170	0.809/0.0051	<b>0.921/0.0072</b>
Average	0.752/0.0882	0.88/0.0098	<b>0.898/0.0089</b>

以被试者 1 为例, 分别通过视频特征、眼电特征, 以及二者特征结合的预测曲线与实际按键测试的误差如图 6-3 所示:

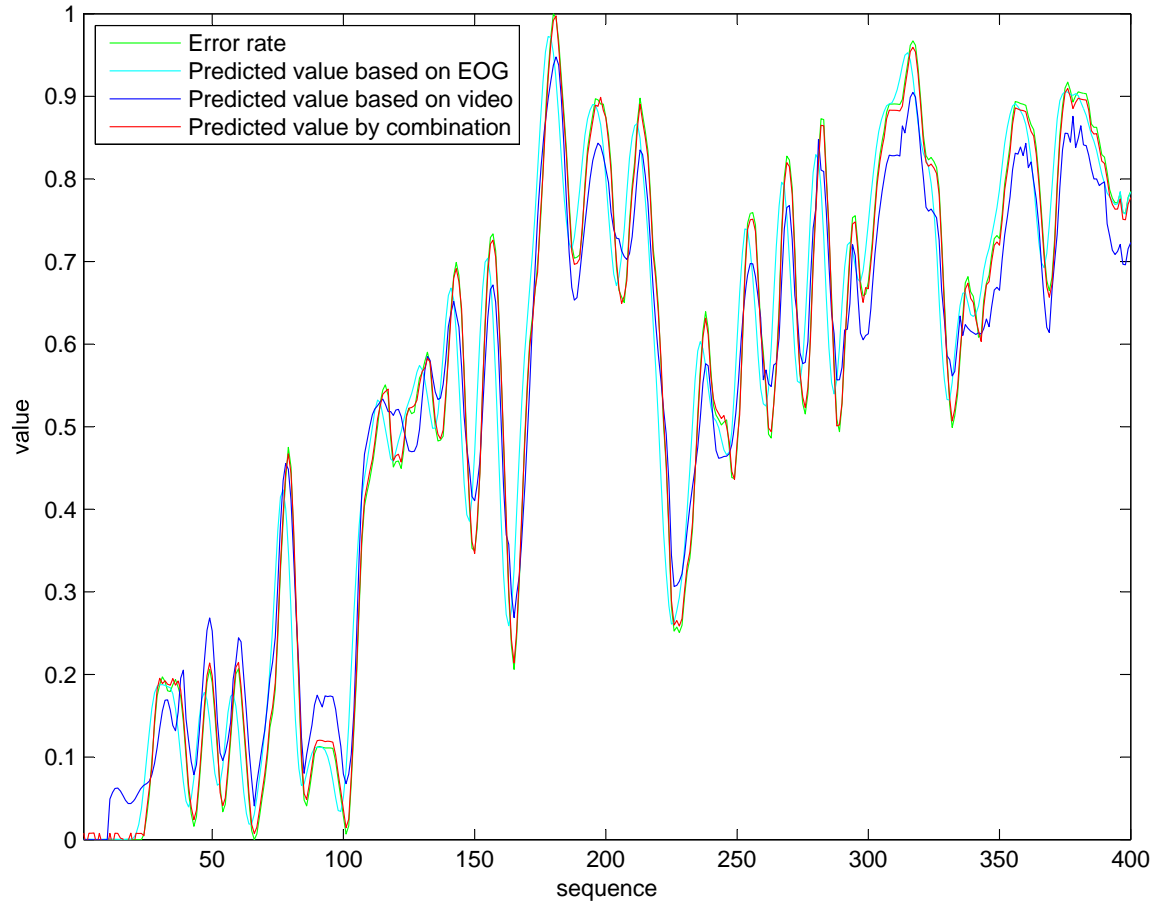


图 6-3 基于视频、眼电、视频眼电的疲劳度预测曲线

### 6.3 本章小结

该实验证明了：我们的疲劳驾驶检测系统可以对被试者真实警觉度和操作错误率提供较准确的预测值。此外，我们将准确的眼电信号与信息更加丰富的视频信号结合，相对于通过单一信号源进行疲劳驾驶检测的方法，有着更加精准的预测结果。今后，我们希望结合更多的信息，如深度、握力信号来开展实验，进一步增强疲劳驾驶检测的鲁棒性与准确性。

## 第七章 全文总结

### 7.1 主要结论

本文主要研究基于视频的疲劳驾驶检测系统，主要涵盖了两年来笔者主持或参与的三大实验：驾驶员情绪分类实验、基于视频与握力信号的疲劳驾驶模拟实验、基于视频与眼电信号的疲劳驾驶模拟实验；工作重心主要在特征提取、分类器训练与融合、以及实时系统研发。

在驾驶员情绪分类实验中，我们在 CK+ 数据库上对 123 名被试者的 593 组视频片段进行了训练以及测试。主要用到的算法涵盖以下几种：使用类 Haar 特征作为输入，Adaboost 算法将弱分类器融合的方式进行人脸检测；使用主动外观模型（ASM）进行人脸对齐和提取几何特征，使用 Gabor 变换提取纹理特征，通过多类别的 Adaboost 算法降维并创新地使用了模糊积分进行分类的融合，最终，我们的实验表明，通过分类器融合几何特征和纹理特征，在该数据集上取得了更好的分类效果。

在基于视频与眼电信号的疲劳驾驶模拟实验中，我们对 5 名被试，在 NeuroScan 系统上分别进行了 67 分钟的眼电、视频同时录制的按键测试实验。笔者第一个设计并完成该融合视频与眼电特征的实验，并在视频特征上提取了打哈欠、头部朝向等，增加了尽可能多的信息；在眼电上，采用水平眼动、垂直眼动、眨眼能量等多种特征进行实验，最终证明了：眼电信号具有表征准确的特性，视频信号具有表义丰富的特性；通过二者结合，可以大幅提升疲劳驾驶检测系统的准确率。该系统同时与基于眼电、脑电信号的疲劳驾驶检测系统参展 2011 年中国国际工业博览会获二等奖。

在基于视频与握力信号的疲劳驾驶模拟实验中，我们对 3 名被试者，在基于真实轿车的疲劳驾驶模拟系统中进行了 90 分钟结合握力与视频的实验。由于可用的实验次数较少，实验信息噪声较大，难以得到相对可靠的结论。初步实验表明：视频信号由于信息丰富，可靠性优于握力信号，将握力信号考虑进疲劳驾驶检测，可以一定程度上提升疲劳驾驶检测系统的鲁棒性和准确性。



## 7.2 创新点

1. 本文完成第一个综合 PERCLOS 特征，主动外观模型几何特征，打哈欠、头部朝向、眼动信息等共 17 种特征的基于视频的在线疲劳驾驶检测系统，并开发了智能手机端系统。
2. 第一个设计并完成了综合眼电与视频信号进行疲劳驾驶检测的按键测试实验，证明了眼电信号的可靠性与视频信息的丰富性，第一个提出综合二者特征可显著提高疲劳驾驶检测准确率的实验结论，该文章已被 ICONIP 2012 收录。
3. 参与并完成综合几何特征与纹理特征的表情识别实验，并通过特征融合极大提高了预测准确率。
4. 参与并完成综合握力与视频信号进行疲劳驾驶检测的实验。

## 7.3 前景展望

本文尽管在基于视频的疲劳驾驶检测系统的研究中得出了一些结论，但是仍然存在需要深化和改进的地方：

从算法效果和实时性上来说，鉴于深度学习一定程度上更容易学习到生理上疲劳驾驶的本质特征，可能会取得更好的实验效果。迫于时间关系，笔者没能完成该项实验。然而，实时性和准确性是一把双刃剑，从目前的技术角度来看，深度学习等很难应用到基于视频的实时系统；反而，现今的基于视频的实时系统已经可以起到较好的预防作用和报警作用，相信随着移动端中央处理器性能的提高，疲劳驾驶检测系统会真正走进千家万户。

此外，普通与红外视频技术仍旧有一定局限性。在真实驾车环境中，佩戴墨镜、强光照射、隧道驾驶等可能会对系统造成或多或少的影响，在光照不足的环境中，手机摄像头很难捕捉到面部细节清晰的画面，容易导致系统的误判。因此，针对复杂的光照、背景和装饰品，我们需要设计更加鲁棒的特征提取和机器学习算法。遇到道路非常颠簸的情况所捕捉的面部会出现运动模糊的情况，此时，人脸识别和器官定位算法都不可避免地出现不精确的情况。因此，在检测状态中，如果系统长时间无法捕捉到清晰的人脸，系统应根据驾车

时间长短来提醒司机注意预防疲劳驾驶。但相信随着高速摄像头和高分辨率感光片的普及，我们提出的基于视频的疲劳驾驶检测系统的普适性会越来越高。

从社会效益和经济效益来看，考虑到智能手机的高普及率以及疲劳驾驶现象的广泛性，未来基于智能手机的疲劳驾驶检测系统应该是一个非常具有前景的市场。根据中国公安部交通管理局 [38] 统计数据，截至 2012 年上半年，全国机动车保有量为 2.33 亿，汽车保有量达到 1.04 亿辆；根据中国工信部统计数据 [39]，截至 2013 年 3 月底，中国手机用户数量达到 11.46 亿，远远超过了机动车保有量。假设三分之一的机动车配备有价值 100 元的基于智能手机的疲劳驾驶检测系统，该产业盈利可超过 70 亿元，边际成本几乎为零；与此同时，全球机动车保有量超过了 10 亿元，该系统的国际市场同样不可小觑。此外，如果手机生产厂商在摄像头设计时可以加入一定的控制方法并相应地开发摄像头控制软件。通过软件将手机摄像头的感光片 CCD 中的 IR 部分滤除掉，并控制摄像头进入红外模式或普通摄像模式。这样，在驾驶过程中，司机就可以完全的将摄像头当作疲劳驾驶的监控器了。无论是在白天还是在黑夜驾驶车辆的时候，驾驶员只要打开该软件，在行驶过程中感到疲劳的时候，手机就会自动提醒他，从而使手机本身能发挥更大的作用。只要我们将基于多传感器的疲劳驾驶检测系统进一步实时运行在移动装备上，或者与汽车导航仪结合，该产品定将走进千家万户，为世界上每一位驾驶员保驾护航！

最后，这是一个三维的时代，计算机视觉正向着三维的方向跨步前行。笔者在微软亚洲研究院从事了半年的三维重建工作，深深认识到未来的视觉技术应该越来越利用三维信息进行识别。比如，通过二维提取 SIFT, ORB 等特征，相邻图之间进行特征匹配，可以进一步深化我们对人脸朝向的认识。这应该成为未来研究的主流方向。

## 附录 A 基于视频的疲劳驾驶检测系统开发大事记

**2011年7月1日** 我加入智能计算与智能系统微软-教育部重点实验室(仿脑计算与人工智能中心), 师从吕宝粮教授, 着手学习计算机视觉与机器学习相关算法。

**2011年10月23日** 我使用 ASM 与 SVM 算法, 为第 14 届中国国际工业博览会完成第一个版本的基于视频的疲劳驾驶检测系统演示程序。

**2011年11月6日** 参与整合基于视频、眼电、脑电的综合疲劳驾驶检测系统, 参展第 14 届中国国际工业博览会获得二等奖。

**2012年3月15日** 参与基于视频的情绪分类实验, 合作发表论文“Facial Expression Recognition by Combining Texture and Geometrical Features, Service Science and Engineering”。

**2012年4月1日** 着手与眼电组合作, 设计并开展基于 NeroScan 系统的综合眼电和视频的疲劳驾驶模拟实验, 着手撰写论文。

**2012年7月1日** 完成实验分析, 以第一作者身份于 ICONIP 2012 上发表论文“Online Vigilance Analysis Combining Video and Electrooculography Features”。

**2012年7月15日** 加入微软亚洲研究院, 师从高级研究员李志伟, 学习图像分割与三维重建相关算法, 研发基于移动设备的三维重建与增强现实系统, 投稿 CVPR 2013 与 CHI 2013 未果, 转投 ICCV 2013。

**2013年3月1日** 着手与握力组与眼电组合组, 开展基于真实轿车与虚拟现实技术的疲劳驾驶模拟实验。

**2013年4月15日** 完成基于智能设备的疲劳驾驶检测系统雏形, 开展基于深度学习的人脸识别实验。

**2013年6月17日** 完成基于真实轿车与虚拟现实技术的疲劳驾驶模拟实验分析, 参与毕业论文答辩。

## 参考文献

- [1] GONZALEZ C, PEREZ M E, SAGON S. IRF World Road Statistics 2011: Data 2004-2009[J]. World Road Statistics, 2012.
- [2] DINGES D, GRACE R. PERCLOS: A Valid Psychophysiological Measure of Alertness as Assessed by Psychomotor Vigilance[R].[S.l.]: Office of Motor Carriers, Federal Highway Administration, 1998.
- [3] 杨钧, 刘钊. 中华人民共和国道路交通事故统计年报 (2010 年度)[R].[S.l.]: 公安部交通管理局, 2010.
- [4] KNIPLING R, WANG J, KANIANTHRA J. Current NHTSA Drowsy Driver: Research and Development[C]//International Technical Conference on the Enhanced Safety of Vehicles. .[S.l.]: [s.n.] , 1996:366–374.
- [5] 孟建柱. 机动车驾驶证申领和使用规定 [EB/OL].(2012-9-12)[2013-5-31]. <http://www.mps.gov.cn/n16/n1282/n3493/n3823/n442207/3386943.html>.
- [6] ADAMS J. Vigilance in the Detection of Low-intensity Visual Stimuli[J]. Journal of Experimental Psychology, 1956, 52(3):204.
- [7] BAKER C. Attention to Visual Displays During a Vigilance Task[J]. British Journal of Psychology, 1959, 50(1):30–36.
- [8] PEIRIS M, JONES R, DAVIDSON P, et al. Identification of Vigilance Lapses Using EEG/EOG by Expert Human Raters[C]//27th Annual International Conference of the Engineering in Medicine and Biology Society. .[S.l.]: [s.n.] , 2006:5735–5737.
- [9] JOHNS M, TUCKER A, CHAPMAN R. A New Method for Monitoring the Drowsiness of Drivers[C]//International Conference on Fatigue Management in Transport Operations. .[S.l.]: [s.n.] , 2005, 23.

- [10] DINGES D F, GRACE R. PERCLOS: A Valid Psychophysiological Measure of Alertness as Assessed by Psychomotor Vigilance[J]. Federal Highway Administration. Office of motor carriers, Tech. Rep. MCRT-98-006, 1998.
- [11] KNIPLING R F, WIERWILLE W W, AMERICA I. Vehicle-based Drowsy Driver Detection: Current Status and Future Prospects[M].[S.l.]: National Highway Traffic Safety Administration, Office of Crash Avoidance Research, 1994.
- [12] SAITO H, ISHIWAKA T, OKABAYASHI S. Applications of Driver's Line of Sight to Automobiles - What Can Driver's Eye Tell[C]//Vehicle Navigation and Information Systems Conference. .[S.l.]: [s.n.] , 1994:21–26.
- [13] WANG Q, YANG J, REN M, et al. Driver Fatigue Detection: A Survey[J]. The Sixth World Congress on Intelligent Control and Automation (WCICA), 2006, 2:8587–8591.
- [14] CAFLISCH L, COLOMBO A, CUSANO C, et al. FaceLab: A Tool for Performance Evaluation of Face Recognition Strategies[C]//Electronic Imaging. .[S.l.]: [s.n.] , 2006:60610V–60610V.
- [15] MAY J F, BALDWIN C L. Driver Fatigue: The Importance of Identifying Causal Factors of Fatigue When Considering Detection and Countermeasure Technologies[J]. Transportation Research Part F: Traffic Psychology and Behaviour, 2009, 12(3):218–224.
- [16] ASSMUS J D, GALLEN C, KELLY T, et al. Quantitative EEG Analysis of Sleep Chronophysiology: A Comparison Between Root Mean Square (RMS) Estimation and Fast Fourier Transformation (FFT).[R].[S.l.]: DTIC Document, 1993.
- [17] LIN C T, WU R C, LIANG S F, et al. EEG-based Drowsiness Estimation for Safety Driving Using Independent Component Analysis[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2005, 52(12):2726–2738.
- [18] DUAN R N, WANG X W, LU B L. EEG-based Emotion Recognition in Listening Music by Using Support Vector Machine and Linear Dynamic System[C]//Neural Information Processing. .[S.l.]: [s.n.] , 2012:468–475.

- [19] WEI Z P, LU B L. Online Vigilance Analysis based on Electrooculography[C]//The 2012 International Joint Conference on Neural Networks (IJCNN). [S.l.]: [s.n.], 2012:1–7.
- [20] S LAL P B L K, A CRAIG, NGUYEN H. Development of An Algorithm for an EEG-Based Driver Fatigue Countermeasure[J]. Journal of Safety Research, 2003, 34(3):321–328.
- [21] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[J]. Computer Vision and Pattern Recognition (CVPR), 2001, 1:1–511 – I–518.
- [22] VIOLA P, JONES M J. Robust Real-Time Face Detection[J]. International Journal of Computer Vision (IJCV), 2004, 57(2):137–154.
- [23] COOTES T F, TAYLOR C J, COOPER D H, et al. Active Shape Models: Their Training and Application[J]. Computer Vision and Image Understanding, 1995, 61(1):38–59.
- [24] COOTES T F, EDWARDS G J, TAYLOR C J. Active Appearance Models[M]//European Conference on Computer Vision.[S.l.]: Springer, 1998:484–498.
- [25] LUCEY P, COHN J F, KANADE T, et al. The Extended Cohn-Kanade Dataset (CK+): A Complete Dataset for Action Unit and Emotion-Specified Expression[C]//Computer Vision and Pattern Recognition Workshops (CVPRW). [S.l.]: [s.n.], 2010:94–101.
- [26] ROTHER C, KOLMOGOROV V, BLAKE A. Grabcut: Interactive Foreground Extraction Using Iterated Graph Cuts[C]//ACM Transactions on Graphics (TOG). [S.l.]: [s.n.], 2004, 23:309–314.
- [27] BLAKE A, ROTHER C, BROWN M, et al. Interactive Image Segmentation Using an Adaptive GMMRF Model[M]//European Conference on Computer Vision (ECCV).[S.l.]: Springer, 2004:428–441.
- [28] McCULLOCH W S, PITTS W. A Logical Calculus of the Ideas Immanent in Nervous Activity[J]. The Bulletin of Mathematical Biophysics, 1943, 5(4):115–133.

- [29] McCLELLAND J L, RUMELHART D E, GROUP P R, et al. Parallel Distributed Processing[J]. Explorations in the Microstructure of Cognition, 1986, 2.
- [30] RUMELHART D E, McCLELLAND J L. Parallel Distributed Processing: Explorations in the Microstructure of Cognition[M], Vol. 1.[S.l.]: MIT press, 1987.
- [31] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4):541–551.
- [32] HINTON G E, OSINDERO S, TEH Y W. A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2006, 18(7):1527–1554.
- [33] LECUN Y, CORTES C. The MNIST Database of Handwritten Digits[EB/OL]. <http://yann.lecun.com/exdb/mnist/>.
- [34] LUO P, WANG X, TANG X. Hierarchical Face Parsing via Deep Learning[C]//Computer Vision and Pattern Recognition (CVPR). .[S.l.]: [s.n.] , 2012:2480–2487.
- [35] KOLDOVSKY Z, TICHAVSKY P, OJA E. Efficient Variant of Algorithm FastICA for Independent Component Analysis Attaining the Cramér-Rao lower bound[J]. IEEE Transactions on Neural Networks, 2006, 17(5):1265–1277.
- [36] DELORME A, MAKEIG S. EEGLab: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis[J]. Journal of Neuroscience methods, 2004, 134(1):9–21.
- [37] CHANG C C, LIN C J. LIBSVM: A Library for Support Vector Machines[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3):27.
- [38] 公安部交管局. 全国机动车保有量 2.33 亿辆 [EB/OL].(2013-01-29)[2013-06-04]. [http://v.youku.com/v\\_show/id\\_XNDI4NzQ3MTky.html](http://v.youku.com/v_show/id_XNDI4NzQ3MTky.html).
- [39] 工信部. 中国手机用户数量达到 11.46 亿 [EB/OL].(2013-04-25)[2013-06-04]. <http://net.zol.com.cn/369/3691822.html>.

- [40] DINGES D, GRACE R. PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance[J]. Federal Highway Administration. Office of motor carriers, Tech. Rep. MCRT-98-006, 1998.
- [41] BULLING A, WARD J, GELLERSEN H, et al. Eye movement analysis for activity recognition using electrooculography[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011(99):1-1.



## 攻读学位期间发表的学术论文目录

- [1] DU R F, LIU R J, WU T X, LU B L. Online Vigilance Analysis Combining Video and Electrooculography Features[C]. Neural Information Processing, 2012, 447:454.
- [2] LIU R J, DU R F, LU B L. Facial Expression Recognition by Combining Texture and Geometrical Features[C]. Service Science and Engineering, 2012.

## 攻读学位期间参与的项目

- [1] 973 项目 “2009CB320901”
- [2] 自然科学基金项目 “61272248”
- [3] 第五期国家级大学生创新训练项目 “S030ITP5022”
- [4] 第四期上海交通大学大学生创新实践计划 “S030IPP004016”
- [5] 第 20 期 PRP “T140PRP20001”
- [6] 第 18 期 PRP “T200PRP18002”

## 致 谢

衷心感谢我的实验室导师吕宝粮教授：感谢吕老师带领我走进 BCMI 这样一个温暖的大家庭，走进您深入浅出的神经网络课程，为我在学术之路迷惘的时候指点迷津。从您的身上，我学到了治学的严谨、求实与脚踏实地。在未来的道路上，我会秉承您的教诲，一步一个脚印，不断攀登学术的高峰。

特别感谢俞勇老师以及俞老师所建立的 ACM 班。我现在取得的一切都离不开俞老师四年来为我们的付出。俞老师四年的栽培是我和 09 级 ACM 班一生的财富。从每学期感人肺腑的学子讲坛，到大三开始的科研实践，每一次的总结不仅是对自己过去人生的审视，更坚定了我对梦想的执着。

还要感谢微软亚洲研究院的李志伟高级研究员、蔡锐博士、张磊博士、曹翔博士和 Koji Yatani 博士。是你们，带领我走进了三维重建这样一个对我而言全新的领域，抑或带领我感受人机交互这样一个五彩斑斓的世界。

感谢 BCMI 实验室图像组的李冰、吴天翔、刘人杰、汤龔鸣、周元剑、申姗姗等师兄师姐，作为论文和项目的合作者，我们一同书写了最绚丽的青春；感谢石立臣、魏正平、彭勇、郑伟龙等师兄，在实验室一次又一次地为我答疑解惑。感谢 ACM 班的同学四年来对我的支持、激励与鞭策。

最后，感谢我的父母，你们用无私的爱，让我感受到了家庭的温暖，亲情的力量；感谢我大学期间所有的任课教师，让我在这里汲取知识的源泉；感谢我的母校——上海交通大学，让我懂得了饮水思源，让我在这里扬帆起航！

## **RESEARCH ON FATIGUE DRIVING DETECTION SYSTEM BASED ON VIDEO SIGNALS**

Fatigue driving is one of the major causes of traffic accident. Additionally, the prediction of fatigue state is of crucial significance for pilots, car and train drivers. Unfortunately, almost every existing vigilance analysis system suffers from poor illumination, horizon of the cameras, together with various appearance and behaviors of the subjects. On one hand, most previous works based on video signals suffer greatly from varying environments such as dim lighting, different skin color and unstable camera. Nevertheless, we could utilize video to extract features not merely on eyes but on yawn states, head movements and body posture as well. On the other hand, features extracted from electrooculography (EOG) signals are scarcely affected by a diversity of environments, but only features of eyes like blinks and movements are utilized.

In this thesis, I am the first to propose a novel, easy-to-use, online system to analyze vigilance level combining both video and Electrooculography (EOG) features. Additionally, I conducted experiments on video and gripping power as well. This system adopts Haar Cascade Adaboost for face detection and Active Shape Model for alignment, calculates PERCLOS by binarization and image gradient algorithms, calculates SEM, REM, yawn frequency and extracts boosted Gabor features from texture. Finally, we apply Linear Dynamic System to denoise and Support Vector Machine for regression. As a result, we improve the accuracy greatly by combining the features from EOG and video. The experiments indicate that our system could serve as an easy-to-use, economical tool for drivers thus benefiting the entire society.

For one thing, the video features extracted from an infrared camera include percentage of closure (PERCLOS) and eye blinks, slow eye movement (SEM), rapid eye movement (REM) are also extracted from EOG signals. For another, other features like yawn frequency, body posture and face orientation are extracted from the video by using Active Shape Model (ASM). The features are described as follows:

- PERCLOS of eyes

The PERCLOS of eyes, which refers to the percentage of eye closure, is an efficient feature to estimate vigilance [40] defined as follows:

$$PERCLOS_e = \frac{\overline{H}_e - H_e}{\overline{H}_e}$$

where  $\overline{H}_e$  indicates the average open eye height above a certain threshold. Another PERCLOS feature is calculated according to the areas of eyes. Finally, the proportions of fully-closed and half-closed eyes and mouth during a certain time window are also regarded as PERCLOS features.

- Blink frequency

The frequency of blinks also has a strong relationship to vigilance. We setup four thresholds  $H_{c1}, H_{c2}, H_{o1}, H_{o2}$ , which indicate the relative height when eyes are about to close, already closed, about to open and fully open. This procedure suggests a complete blink. The times of eye blinks during a time window is calculated as vigilance features.

- Eye Movement

The relative position of the pupil can be recognized by ASM. Thus the moving frequency of eyes are recorded as an important feature. During each time window, we calculate the movement of eye pupils and its the amplitude. The speed of the eye movement is also calculated as a feature.

- Yawn frequency

Since an action of yawn suggest significantly that the subject has already been fatigued. The window size  $w$  for yawn state should be large enough such as 16 seconds  $\times$  30 frames. Denote the average of the least  $k$  heights of mouth as  $H_m^k$

$$Y_i = \frac{\sum_{j=i-w}^i (H_j / H_m^k) > C}{w}$$

Here  $C$  is a threshold and indicates the ratio between open mouth height and

normal mouth height when the subject is about to yawn.

- Body Posture

We estimate the posture of body by locating the relative position of eyes, nose and mouth on the face. Denote orientation of the face as  $\alpha$ ,  $\theta$  and  $\beta$ , corresponding to different reference of eyes, nose and mouth. This degree can be calculated as follows:

$$\alpha = \frac{D(67, 2)}{D(67, 12)}; \theta = \frac{D(31, 0)}{D(36, 14)}; \beta = \frac{D(66, 3)}{D(66, 11)}$$

where points 67, 66, 31, 36 denote the center of the nose, mouth, left and right pupil separately while the others indicate the left and right side of the face horizontally and correspondingly.

- Blink Features

After the preprocess of EOG signals, every blink is marked at four time points  $c1$ ,  $c2$ ,  $o1$  and  $o2$ , which indicate the time when the eye is to close, closed, to open and opened. Denote  $V$  as the signal,  $D$  as the difference of signal, we have the following features:

$$\begin{aligned} T_{blink} &= T_{o2} - T_{c1}; & T_{close} &= T_{c2} - T_{c1} \\ T_{open} &= T_{o2} - T_{o1}; & T_{closed} &= T_{o2} - T_{c2} \\ S_{close} &= \frac{\sum_{i=T_{c1}}^{T_{c2}} D_i}{T_{close}}; & S_{open} &= \frac{\sum_{i=T_{o1}}^{T_{o2}} D_i}{T_{open}}; & E_{blink} &= \sum_{i=T_{c1}}^{T_{o2}} V_i^2 \end{aligned}$$

where  $T$  indicates the time during a window size,  $S$  indicates the speed, and  $E$  indicates the energy of blinks.

- Eye Movements

Two kinds of eye movements, Slow Eye Movement (SEM) and Rapid Eye Movement (REM) are extracted, according to different kinds of time threshold in [41]. In order to get these features more accurately, two methods of Fourier transformation and wavelet transformation are used. In the Fourier transformation method,

we use a band-pass filter with frequency 0.5Hz and 2Hz to process the horizontal EOG signal. The sampling rate is 125Hz and the period is 8 seconds.

The contribution of this thesis includes the following four aspects:

1. To our knowledge, this is the first thesis to propose an online fatigue detection system combining PERCLOS, ASM geometry features, yawn states and Gabor texture features. Additionally, we built a system on mobile phones.
2. To our knowledge, we are the first to design and complete experiments taking both EOG signals and video signals into account, which proves the accuracy of EOG features and the richness of video signals. We prove that combining both features could effectively improve the overall accuracy of fatigue detection system. Our paper is published on ICONIP 2012.
3. We conducted experiments on facial expressions recognition by combining both geometry and texture features. We use fuzzy integral to improve the overall accuracy on CK+ database.

There are totally five healthy subjects for our experiments, including four men and 1 woman, all of whom are around 23 years old. Particularly, we ensure that none of the subjects is color-blind. Each subject is asked to have sufficient sleep at night and get up early in the morning. Each experiment is conducted after lunch and lasts for 67 minutes so that the subject behaves sober at first and sleepy after a period of about half an hour in the experiment. The room is in silence and the light is soft.

From the experimental results, we can arrive at the conclusion that our new system offers a good prediction of the actual vigilance level. Moreover, this method outperforms the existing approaches using either video features or EOG features alone, since our proposed method utilizes both the accuracy of EOG signals but the yawn state and body postures provided by video as well. The initial result from video and gripping power experiments indicate similar results. In the future, we plan to utilize comprehensive features including depth information to get a better performance. Besides, more experiments will be conducted and the stability and robustness of the algorithms are expected to be improved.