UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Sistema ADAS para identificação de distrações e perturbações do motorista na condução de veículos

Rafael Alceste Berri

Tese de Doutorado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Rafael Alceste Berri

Sistema ADAS para identificação de distrações e perturbações do motorista na condução de veículos

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Fernando Santos Osório

USP – São Carlos
Março de 2019

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

B533s Berri, Rafael Alceste
 Sistema ADAS para identificação de distrações e
 perturbações do motorista na condução de veículos /
 Rafael Alceste Berri; orientador Fernando Santos
 Osório. -- São Carlos, 2019.
 141 p.

 Tese (Doutorado - Programa de Pós-Graduação em
 Ciências de Computação e Matemática Computacional) --
 Instituto de Ciências Matemáticas e de Computação,
 Universidade de São Paulo, 2019.

 1. Distrações do Motorista. 2. ADAS. 3.
 Reconhecimento de Padrões. 4. Aprendizado de
 Máquina. 5. Redes Neurais Artificiais. I. Osório,
 Fernando Santos, orient. II. Título.

Rafael Alceste Berri

**ADAS system for recognition of driver's distractions and
disturbances while driving**

Doctoral dissertation submitted to the Institute of
Mathematics and Computer Sciences – ICMC-USP, in
partial fulfillment of the requirements for the degree of
the Doctorate Program in Computer and Mathematical
Sciences. *FINAL VERSION*

Concentration Area: Computer Science and
Computational Mathematics

Advisor: Prof. Dr. Fernando Santos Osório

USP – São Carlos
March 2019

*Este trabalho é dedicado às crianças adultas que,
quando pequenas, sonharam em se tornar cientistas.
Em especial, aos pesquisadores do Instituto de Ciências Matemáticas e de Computação (ICMC).*

AGRADECIMENTOS

Agradeço a Deus, por escolher para mim este caminho e nele me fazer caminhar. A meus pais, Laurindo e Salete, pelo apoio incondicional nos meus estudos, trabalho e demais desafios assumidos. Ao meu irmão Gabriel que, mesmo longe, buscou me auxiliar sempre que possível. Aos inúmeros amigos e colegas que contribuíram ou com um pouco do seu conhecimento, ou mesmo, com apoio no decorrer dos anos de estudos, sou muito agradecido.

Ao professor Dr. Fernando Santos Osório pela amizade, orientação, contribuições e adequações ao trabalho.

À CAPES, pelo auxílio financeiro durante a realização do trabalho.

À PPG-CCMC, pelas oportunidades que o programa traz ao doutorando na aquisição e geração de conhecimento.

Agradeço muito aos 23 colaboradores que disponibilizaram seu tempo para a obtenção dos dados de validação e treinamento do sistema apresentado neste trabalho, e dentre os quais, alguns perderam uma noite inteira de sono para efetuar o teste de direção sonolenta.

Agradeço a professora Ana Paula C. Larocca e ao STI/EESC/USP por disponibilizarem um incrível lugar para coleta de dados envolvendo os motoristas participantes.

E agradeço ao colega Tiago dos Santos que criou um cenário virtual de uma autoestrada para os experimentos, ao qual, o simulador de direção desenvolvido se baseou.

*“O temor do Senhor é o princípio do conhecimento mas
os insensatos desprezam a sabedoria e a disciplina.”
(Bíblia Sagrada, NVI, Provérbios 1:7)*

RESUMO

BERRI, R. A. **Sistema ADAS para identificação de distrações e perturbações do motorista na condução de veículos**. 2019. 141 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

Este trabalho apresenta um sistema que se utiliza de características extraídas de dados provenientes de um sensor Kinect v2 para monitorar o motorista, dados de sensores inerciais, da telemetria do veículo e dados sobre a estrada/faixa de rodagem para reconhecer o estilo de direção, permitindo ao sistema detectar o uso do celular no trânsito, um motorista embriagado e a direção sonolenta, evitando assim, riscos relacionados com a direção. De fato, quando veículos são conduzidos por pessoas em ligações telefônicas, o risco de acidente aumenta de 4 a 6 vezes. Motoristas embriagados causaram 10.497 mortes nas rodovias dos Estados Unidos da América em 2016, segundo o órgão local responsável pela segurança no trânsito (NHTSA). Um Conjunto de Dados Naturalista do Comportamento do Motorista (NDBD) foi criado especificamente para este trabalho e utilizado para o teste e validação do sistema proposto. A solução proposta emprega duas análises dos dados do motorista, os subsistemas de reconhecimento de padrões de Curto e Longo prazos. Assim, pode-se detectar situações de risco na direção. O sistema possui 3 níveis de alerta: “sem alerta”, “alerta baixo” e “alerta alto”. O subsistema de Curto Prazo detecta situações de “sem alerta” e de algum nível de alerta. Já o subsistema de Longo Prazo é responsável por determinar o nível de alerta: baixo ou alto. Classificadores baseados em Aprendizado de Máquina e Redes Neurais Artificiais (RNA) foram utilizados. Um Algoritmo Genético foi empregado para otimizar e selecionar um conjunto de valores que ajustam a entrada de características, função de ativação dos neurônios e topologia/treino da rede neural. O sistema proposto alcançou 79,5% de acurácia nos *frames* do NDBD (conjunto de treinamento e validação obtidos utilizando um simulador veicular próprio), para a detecção conjunta de risco em situações de uso de celular, embriaguez ou condução normal. Para o classificador de Curto Prazo, utilizou-se períodos de 5 *frames* e uma janela de 140 *frames* para o Longo Prazo. Considerando a detecção individualizada dos problemas de condução, no caso específico da embriaguez (usados dados de embriaguez e direção normal) o sistema obteve 98% de acurácia, e especificamente para o uso de celular obteve 95% de acurácia. Na classificação de “sem alerta” (situações sem risco), o sistema obteve apenas 1,5% de predições erradas (falsos positivos), contribuindo assim para o conforto do motorista ao utilizar o sistema.

Palavras-chave: Distrações do Motorista, ADAS, Reconhecimento de Padrões, Aprendizado de Máquina, Redes Neurais Artificiais, Algoritmo Genético, Veículos Inteligentes.

ABSTRACT

BERRI, R. A. **ADAS system for recognition of driver's distractions and disturbances while driving**. 2019. 141 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

In this work, a system has been developed using features from a frontal Kinect v2 sensor to monitor the driver, from inertial sensors, car telemetry, and road lane data to recognize the driving style, enabling to recognize the use of a cell phone while driving, a drunk driver, and drowsy driving, avoiding driving risks. In fact, cars driven by people on phone calls, increases the risk of crash between 4 and 6 times. Drunk drivers caused 10,497 deaths on USA roads in 2016 according to NHTSA. The Naturalistic Driver Behavior Dataset (NDBD) was created specifically for this work and it was used to test the proposed system. The proposed solution uses two analysis of the driver's data, the Short-Term and Long-Term pattern recognition subsystems, thus it could detect the risk situations while driving. The system has 3 levels of alarm: no alarm, lowest alarm, and highest alarm. Short-Term detects between no alarm or some level alarm. Long-Term is responsible for determining the risk alarm level, low or high. The classifiers are based on Machine Learning and Artificial Neural Networks (ANN), furthermore, the values set to adjust input features, neuron activation functions, and network topology/training parameters were optimized and selected using a Genetic Algorithm. The proposed system achieved 79.5% of accuracy in NDBD frames (training and validation sets obtained using a driving simulator), for joint detection of risk in situations of cellphone usage, drunkenness, or normal driving. For the Short-Term classifier, it was used length periods of 5 frames and a window of 140 frames for Long-Term. Considering the individualized detection of driving problems, in the specific case of drunkenness (using data of drunkenness and normal driving), the system achieved 98% of accuracy, and specifically for cell phone usage 95% of accuracy. The best results achieved obtained only 1.5% of “no risk” situation having a wrong prediction (false positives with alarm activation), contributing to the driver comfort when he/she is using the system.

Keywords: Driver Distractions, ADAS, Pattern Recognition, Machine Learning, Artificial Neural Networks, Genetic Algorithm, Intelligent Vehicles.

LISTA DE ILUSTRAÇÕES

Figura 1 – Estrutura deste trabalho.	34
Figura 2 – Visão de alguns Sistemas Avançados de Assistência ao Condutor (ADAS). . .	38
Figura 3 – Exemplos de quadros que possuem sua dificuldade de segmentação por conta da incidência de luz solar.	45
Figura 4 – Ilustração da medição da profundidade em câmeras de luz estruturada. . . .	51
Figura 5 – Visão geral do Kinect v1.	51
Figura 6 – Visão geral do Kinect v2.	52
Figura 7 – Mapa do alcance da medição da profundidade para o Kinect v2.	53
Figura 8 – Classificação dos diferentes tipos de vídeo baseado na necessidade de movimentação e processamento.	54
Figura 9 – Processo de subtração de fundo, onde, N é o número de frames usados na inicialização do fundo, B_t e I_t são, respectivamente, o fundo e o quadro atual no tempo t	55
Figura 10 – Etapas típicas da esqueletização.	63
Figura 11 – Esboço do SVM.	65
Figura 12 – Arquitetura da SVM. A função de kernel K determina o tipo de classificador (ex: Polinomial). Todos os outros parâmetros são encontrados durante o treinamento.	67
Figura 13 – Um diagrama esquemático de uma rede MLP de três camadas.	68
Figura 14 – Exemplo do posicionamento das origens do Volante e da Cabeça do motorista em uma nuvem de pontos, sendo frontal (a) e lateral (b), onde, o eixo X é o vermelho, Y verde e Z azul.	73
Figura 15 – Fluxo do SCLP.	76
Figura 16 – Fluxo do SB.	77
Figura 17 – Equipamentos (a) e um exemplo de tela (b) do <i>Simulator for Distracted Driving Research</i> (SDDR).	82
Figura 18 – Resultado da avaliação do SDDR comparando a uma situação real, feita por 19 participantes, após o utilizar por uma hora.	84
Figura 19 – Imagens obtidas do SDDR.	85
Figura 20 – <i>Drunk Busters Impairment Goggles</i> utilizado na simulação da visão embriagada. .	87

Figura 21 – Gráfico dos melhores <i>Findex</i> ₅ para Qualidade na Direção obtidos em cada comprimento de período. O período com 220 <i>frames</i> possui o melhor <i>Findex</i> ₅ , 80 <i>frames</i> é o melhor <i>Findex</i> ₅ para períodos curtos e 300 é a melhor opção para períodos longos.	93
Figura 22 – Gráfico dos classificadores da Qualidade na Direção obtidos do AG, sendo os dados de treinamento (cores escuras) e dados de validação (cores claras), onde, a barra vermelha é o <i>Findex</i> ₅ , a barra verde é a acurácia e os limites do desvio padrão nos grupos de <i>cross-validation</i> de treinamento são mostrados em cinza.	95
Figura 23 – Resultados dos seis sistemas para detecção de embriaguez considerando o buffer de avaliação (BS), onde, onde, as barras vermelhas são o índice <i>Findex</i> ₅ e as verde a acurácia final. As cores escuras e claras indicam conjuntos de treinamento e validação, respectivamente.	95
Figura 24 – Gráfico dos melhores <i>Findex</i> ₅ obtidos em cada comprimento de períodos no Monitoramento do Motorista.	97
Figura 25 – Gráfico dos <i>Findex</i> ₅ para cada tamanho de grupo de características usando período de 5 <i>frames</i> no Monitoramento do Motorista.	98
Figura 26 – Gráfico dos classificadores para o Monitoramento do Motorista obtidos usando AG, <i>cross-validation</i> e dados de teste, onde, a barra vermelha indica o <i>Findex</i> ₅ , a verde é indicativa de acurácia e os limites de um desvio padrão são mostrados em cinza.	99
Figura 27 – Resultado dos quatro SCLP para detecção de risco do motorista pelo uso de celular enquanto dirige, onde, a barra vermelha indica o <i>Findex</i> ₅ , a azul a acurácia dos sistemas de alertas, em verde a acurácia do alerta baixo e amarelo é a acurácia para alerta alto.	100
Figura 28 – Gráfico para Monitoramento do Motorista e Qualidade na Direção dos melhores <i>Findex</i> ₅ obtidos para cada tamanho de período.	102
Figura 29 – Gráfico para Monitoramento do Motorista e Qualidade na Direção dos melhores <i>Findex</i> ₅ para o número de características selecionadas e em períodos de 140 <i>frames</i>	103
Figura 30 – Gráfico para Monitoramento do Motorista e Qualidade na Direção dos melhores <i>Findex</i> ₅ conforme número de características selecionadas e utilizando períodos de 280 <i>frames</i>	103
Figura 31 – Gráfico para Monitoramento do Motorista e Qualidade na Direção do melhor <i>Findex</i> ₅ conforme número de características selecionadas e utilizando períodos de 5 <i>frames</i>	104
Figura 32 – Gráfico para Monitoramento do Motorista e Qualidade na Direção do melhor <i>Findex</i> ₅ conforme número de características selecionadas e utilizando períodos de 15 <i>frames</i>	105

Figura 33 – Gráfico para Monitoramento do Motorista e Qualidade na Direção dos classificadores obtidos pelo AG usando os dados de treinamento (cores escuras), dados de validação (cores claras) e dados de sonolência (cores claras com um padrão branco), onde, a barra vermelha é do <i>Findex</i> ₅ , a verde é da acurácia, e os limites permitidos pelo desvio padrão nos grupos de <i>cross-validation</i> de treinamento são mostrados em cinza.	106
Figura 34 – Resultado para Monitoramento do Motorista e Qualidade na Direção dos classificadores de curto prazo para <i>npCP</i> entre 2 e 8, onde, as linhas escuras indicam o <i>Findex</i> ₅ obtido, as linhas claras a acurácia, vermelho é indicação de Gaussiana ₅ , verde de Sigmoide ₅ , azul de Gaussiana ₁₅ e amarelo de Sigmoide ₁₅	107
Figura 35 – Resultados para 16 sistemas de detecção de distrações e perturbações do motorista usando características do Monitoramento do Motorista e da Qualidade na Direção, onde, a linha vermelha indica o <i>Findex</i> ₅ , a linha azul é a acurácia do sistema de alerta, a verde indica acurácia para detecção de risco, a linha amarela indica o acerto para “sem alerta” (direção normal), marrom é o acerto do “alerta baixo” e roxo indica o acerto para o “alerta alto”.	108
Figura 36 – Acerto dos alertas em todo o NDBD para Monitoramento do Motorista e Qualidade na Direção nos 8 sistemas utilizando cg5 e cg15 para o teste da direção normal (verde), teste do uso do celular (vermelho), teste de embriaguez (amarelo) e acurácia geral (azul).	109
Figura 37 – Acerto dos alertas para o conjunto de treinamento do Monitoramento do Motorista e Qualidade na Direção nos 8 sistemas utilizando cg5 e cg15 para o teste da direção normal (verde), teste do uso do celular (vermelho), teste de embriaguez (amarelo) e acurácia geral (azul).	109
Figura 38 – Acerto dos alertas para o conjunto de validação do Monitoramento do Motorista e Qualidade na Direção nos 8 sistemas utilizando cg5 e cg15 para o teste da direção normal (verde), teste do uso do celular (vermelho), teste de embriaguez (amarelo) e acurácia geral (azul).	110
Figura 39 – Acerto dos alertas para a sonolência no Monitoramento do Motorista e na Qualidade na Direção nos 8 sistemas utilizando cg5 e cg15 para o teste da direção normal (verde), teste do uso do celular (vermelho), teste de embriaguez (amarelo) e acurácia geral (azul).	110
Figura 40 – Acerto dos alertas do sistema cg5#ls140 agrupando os dados dos grupos de treinamento e validação nos diferentes conjuntos de dados do NDBD, sendo acerto para “sem alerta” indicado pela barra verde, alerta baixo indicado pela vermelha, alerta alto mostrado pela barra amarela e a acurácia geral dos alertas no grupo de treinamento apresentada pela barra azul.	112
Figura 41 – Exemplo de saída do Sistema de Detecção de Distrações e Perturbações. . .	112

Figura 42 – Exemplo de problemas de predição de risco para o cg5#ls140, onde o círculo azul indica a posição da mão direita e o vermelho a posição da mão esquerda. 113

Figura 43 – Características da experiência virtual de embriaguez. 115

LISTA DE TABELAS

Tabela 1 – Tabela de comparação entre os trabalhos relacionados.	44
Tabela 2 – Comparativo entre as especificações do Microsoft Kinect v1 e v2 (AL-NAJI <i>et al.</i> , 2017).	52
Tabela 3 – Todas as características candidatas para o monitoramento do motorista . . .	73
Tabela 4 – Todas as características candidatas na qualidade de direção	74
Tabela 5 – Todas as informações disponíveis em cada <i>frame</i> do NDBD	86
Tabela 6 – Frames do NDBD por pessoa e Seção.	87
Tabela 7 – Valor adotado para ϕ em cada uma das características nas origens <i>VOL</i> e <i>CM</i>	92
Tabela 8 – Parâmetros dos classificadores MLP para Qualidade na Direção.	94
Tabela 9 – Parâmetros da MLP empregados no AG para o Monitoramento do Motorista	98
Tabela 10 – Parâmetros dos classificadores MLP para detecção de celular.	99
Tabela 11 – Todas as características escolhidas para as frequências candidatas a <i>nLP</i> para Monitoramento do Motorista e Qualidade na Direção.	102
Tabela 12 – Todas as características escolhidas para as frequências candidatas a <i>nCP</i> para Monitoramento do Motorista e Qualidade na Direção.	104
Tabela 13 – Limites do AG para os Parâmetros do MLP para Monitoramento do Motorista e Qualidade na Direção	105
Tabela 14 – Parâmetros para o Monitoramento do Motorista e Qualidade na Direção dos classificadores MLP providos pelo AG.	106
Tabela 15 – Tabela de comparação entre os trabalhos relacionados e este trabalho.	124

LISTA DE ABREVIATURAS E SIGLAS

@DP — Representa uma característica de desvio padrão para um período/janela de dados nos experimentos

@MED — Representa uma característica de média para um período/janela de dados nos experimentos

acel — Aceleração do veículo simulado

ADAS — Advanced Driver-assistance System

AG — Algoritmo Genético

aX — Componente X do acelerômetro acoplado ao veículo simulado

aY — Componente Y do acelerômetro acoplado ao veículo simulado

aZ — Componente Z do acelerômetro acoplado ao veículo simulado

BS — Buffer Size ou Tamanho do Buffer de previsões para o SB

CaRINA — Carro Inteligente para Navegação Autônoma

CCD — Charged Coupled Device ou Dispositivo de Carga Acoplada, um tipo sensor semicondutor para captação de imagens (câmera)

classificadorCP — Classificador para o Reconhecimento de Curto Prazo do SCLP

classificadorLP — Classificador para o Reconhecimento de Longo Prazo do SCLP

classificadorSB — Classificador para o SB

CM — Origem das coordenadas na Cabeça do Motorista

CMOS — Complementary Metal-Oxide-Semiconductor ou Semicondutor de Metal-Óxido Complementar, uma tecnologia empregada na fabricação de circuitos integrados (câmera)

CO — Número de neurônios na camada oculta da rede de um classificadores MLP

CP — Curto Prazo (nCP frames)

distCF — Distância do veículo simulado para o Centro da Faixa de rodagem

distMD — Distância para a mão direita do motorista

distME — Distância para a mão esquerda do motorista

fAngulo — Ângulo da faixa de rodagem em relação ao veículo simulado

feature_n — Valor normalizado da característica de Monitoramento do Motorista para o frame de dados

feature_v — Valor da característica de Monitoramento do Motorista calculado partindo-se do frame de dados

Findex₅ — Índice utilizado na avaliação do treinamento de classificadores que visa mensurar a presença de falsas detecção de risco no conjunto de treinamento pelo ponderamento da Precisão (precision) e Cobertura (recall) [Adotou-se o *Findex₅* neste trabalho]

fitness — Função objetivo utilizado pelo AG para mensurar a qualidade de treinamento de classificadores

GB — Gigabyte

GPS — Global Positioning System

gX — Componente X do giroscópio acoplado ao veículo simulado

gY — Componente Y do giroscópio acoplado ao veículo simulado

gZ — Componente Z do giroscópio acoplado ao veículo simulado

HMM — Hidden Markov Model ou Modelo Oculto de Markov

HOG — Histogram of Oriented Gradients ou Histograma Orientado a Gradientes, um tipo de descritor em Visão Computacional

ICMC — Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (São Carlos-SP)

interac — Número máximo de interações para o método de aprendizado Rprop (MLP)

IR — Infra-Red ou Infravermelho

LP — Longo Prazo (*nLP frames*)

LRM — Laboratório de Robótica Móvel do ICMC/USP

maxDistMaos — Distância máxima para as mãos do motorista

MB — Megabyte

minDistMaos — Distância mínima para as mãos do motorista

MLP — Multilayer Perceptron

MoG — Mixture of Gaussians ou Mistura de Gaussianas, uma técnica de subtração de fundo de imagens

nCP — Número de frames que compõe um período para o Reconhecimento de Curto Prazo do SCLP

NDBD — Naturalistic Driver Behavior Dataset

nf — Número de frames para um período do SB

NHTSA — *National Highway Traffic Safety Administration of United States*

nLP — Número de frames que compõe uma janela de dados para o Reconhecimento de Longo Prazo do SCLP

npCP — Número de predições para a Avaliação no Reconhecimento de Curto Prazo do SCLP

P — Medida de precisão de dados no reconhecimento de padrões

PAlevel — Nível do pedal do acelerador do joystick usado pelo motorista no controle do veículo simulado

PCA — *Principal Component Analysis* ou *Análise de Componentes Principais*, uma ferramenta de análise exploratória de dados

PERCLOS — *Percentage of Eye Closure*

PFlevel — Nível do pedal de freio do joystick usado pelo motorista no controle do veículo simulado

R — Medida de cobertura (recall) de dados no reconhecimento de padrões

RGB — *Red Green Blue* ou um sistema de cores aditivas em que o Vermelho (Red), o Verde (Green) e o Azul (Blue) são combinados formando cores

RGB-D — *RGB* acrescido de *Depth* (profundidade), ou seja, cada pixel possui informação de cor (*RGB*) e da sua distância para o sensor de captura

RNA / ANN — *Rede Neural Artificial (Artificial Neural Net)*

RP — *Reconhecimento de Padrões*

rReta — Indicação de que o trecho atual da rodovia virtual é reto ou curvo (booleano)

SB — *Sistema Bufferizado de Detecção de Riscos*

SCLP — *Sistema de Curto e Longo Prazos para Detecção de Riscos*

SDDR — *Simulator for Distracted Driving Research*

shoulder_l — Comprimento dos ombros do motorista para o frame de dados

SIFT — *Scale-Invariant Feature Transform* ou *Transformação em Característica Invariante de Escala*, um tipo de descritor em Visão Computacional

SMS — *Short Message Service*

SVM — *Support Vector Machine*

ToF — *Time of Flight*

U.S. CDC — *Centers for Disease Control and Prevention of United States*

velocidade — Velocidade do veículo simulado

VOL — Origem das coordenadas no Volante

VOLangulo — Ângulo do volante do joystick usado pelo motorista no controle do veículo simulado

xMD — Coordenada X da mão direita do motorista (Posição 3D)

xME — Coordenada X da mão esquerda do motorista (Posição 3D)

yMD — Coordenada Y da mão direita do motorista (Posição 3D)

yME — Coordenada Y da mão esquerda do motorista (Posição 3D)

zMD — Coordenada Z da mão direita do motorista (Posição 3D)

zME — Coordenada Z da mão esquerda do motorista (Posição 3D)

LISTA DE SÍMBOLOS

φ — Constante obtida experimentalmente para cada característica de monitoramento do motorista que relaciona a característica ao comprimento dos ombros do motorista

α — Primeiro parâmetro da função de ativação de um classificador MLP

β — Segundo parâmetro da função de ativação de um classificador MLP

SUMÁRIO

1	INTRODUÇÃO	29
1.1	Contextualização e Motivação	29
1.2	Objetivos	31
1.3	Aplicações	33
1.4	Estrutura da tese	33
2	TRABALHOS RELACIONADOS	35
2.1	Sistemas Avançados de Assistência ao Condutor	35
2.2	Sistemas de Detecção de Distrações e Perturbações	37
2.2.1	<i>Detecção de Embriaguez</i>	38
2.2.2	<i>Detecção de Sonolência</i>	39
2.2.3	<i>Detecção de Celular</i>	40
2.2.4	<i>Detecção de Direção Insegura</i>	42
2.3	Considerações	43
3	REFERENCIAL TEÓRICO	47
3.1	Sensores	47
3.1.1	<i>Câmeras de Vídeo</i>	48
3.1.2	<i>Sensores de Posicionamento</i>	49
3.1.3	<i>Sensores Inerciais</i>	49
3.1.4	<i>Sonares</i>	50
3.1.5	<i>Sensores 3D</i>	50
3.2	Subtração de Fundo	53
3.2.1	<i>Câmera Estática</i>	54
3.2.2	<i>Câmera em Movimento</i>	57
3.2.3	<i>Câmera Quase-estática</i>	59
3.2.4	<i>Subtração de Fundo em Câmeras RGB-D</i>	60
3.3	Esqueletização do Corpo Humano	62
3.4	Reconhecimento de Ações Humanas	63
3.5	Reconhecimento de Padrões	65
3.5.1	<i>Máquina de Vetor de Suporte (SVM)</i>	65
3.5.2	<i>Multilayer Perceptron (MLP)</i>	67
3.6	Considerações	68

4	DESENVOLVIMENTO DA SOLUÇÃO	71
4.1	Características	71
4.1.1	<i>Monitoramento do Motorista</i>	72
4.1.2	<i>Qualidade de Direção</i>	74
4.2	Sistema de Curto e Longo Prazos para Detecção de Riscos	75
4.2.1	<i>Reconhecimento de Curto Prazo</i>	75
4.2.2	<i>Reconhecimento de Longo Prazo</i>	76
4.2.3	<i>Controle do Alerta</i>	77
4.3	Sistema Bufferizado de Detecção de Riscos	77
4.4	Considerações	78
5	SIMULADOR E A COLETA DE DADOS PARA EXPERIMENTOS	81
5.1	Simulador de direção para pesquisas sobre distrações e perturbações do motorista no trânsito	82
5.2	Conjunto de dados	84
5.3	Considerações	88
6	EXPERIMENTOS, RESULTADOS E DISCUSSÕES	91
6.1	Qualidade na direção	92
6.2	Monitoramento do Motorista	96
6.3	Monitoramento do Motorista e Qualidade na Direção	101
6.4	Considerações	115
7	CONSIDERAÇÕES FINAIS	119
7.1	Conclusões	119
7.2	Contribuições Científicas	121
7.3	Trabalhos futuros	122
7.4	Lista de Publicações do Autor Relacionadas a esta Tese	123
	REFERÊNCIAS	125

INTRODUÇÃO

1.1 Contextualização e Motivação

O trânsito figura entre as 8 principais causas de mortes de pessoas no planeta (BUCKE-RIDGE, 2015). Olhando-se somente para acidentes fatais no trânsito, 95% destes são causados por erros humanos (AMDITIS *et al.*, 2010). As três principais causas de erros humanos no trânsito são a embriaguez, a sonolência e as distrações do motorista de maneira geral (MURATA *et al.*, 2011).

Considerando-se, em particular, os motoristas embriagados, eles causaram 1.800 mortes nas estradas dos Estados Unidos da América em 2008, número que representa 30% do total de mortes no trânsito neste período (WILSON; STIMPSON, 2010). Apesar de existirem leis rígidas em muitos países, menos de 1% dos motoristas embriagados são abordados pela polícia segundo U.S. CDC (Centro de Controle e Prevenção de Doenças) (CDC, 2013). Pesquisa realizada em bares na cidade de Porto Alegre (RS) constatou que 51% dos motoristas frequentadores desses lugares iriam dirigir seus veículos após a ingestão de álcool e que apenas 9% deles já haviam sido abordados, alguma vez na vida, pela polícia para realizar o teste do bafômetro (BONI *et al.*, 2010). Um estudo mais abrangente realizado em 143 cidades brasileiras em 2009 indicou que beber e dirigir é aceitável para 35% da população (PECHANSKY *et al.*, 2009).

A sonolência, segundo os sociólogos Steve Kroll-Smith e Valerie Gunter, é considerada como a nova embriaguez, já que, é um estado culpável, onde, as pessoas se permitem dirigir quando estão cansadas, como ocorre na direção alcoolizada (WILLIAMS, 2011). Outro indicativo da relação entre a embriaguez e sonolência é feita pela recomendação, da autoridade brasileira de trânsito aos agentes fiscalizadores, de que a sonolência é um dos indicadores da embriaguez (OGAMA, 2014). Além disto, cabe lembrar que os motoristas também podem estar dirigindo sob a ação de substâncias e drogas psicoativas que podem ter como efeitos colaterais comportamentos de sonolência, perda de concentração e reflexos (RAY; FOUGHT; DECKER, 1992; NEUTEL,

1995). Observando números, a sonolência causa cerca de 100.000 acidentes nas estradas dos Estados Unidos, segundo a National Highway Traffic Safety Administration (NHTSA), o que resulta em mais de 1.500 mortos e 71.000 feridos ao ano. Isso equivale a 1,6% de todos os acidentes e de 3,6% dos acidentes fatais (NTSB, 1999). Em pesquisa realizada por Santos *et al.* (2013) com motoristas de 5 empresas de ônibus, cerca de 68% dos motoristas avaliados na cidade de Brasília (DF) e 41% dos motoristas avaliados em Florianópolis (SC) apresentaram sonolência diurna excessiva. Em alguns estados dos EUA como Nova Jersey, um motorista sonolento pode ser considerado um homicida (WILLIAMS, 2011).

A distração ao volante (REGAN; LEE; YOUNG, 2008; PEISSNER; DOEBLER; METZE, 2011), ou seja, uma ação que leve o motorista a desviar a atenção da pista por alguns segundos, representa aproximadamente metade dos casos de acidentes no trânsito. Em rodovias federais brasileiras, cerca de 63.000 acidentes, em 2010, tiveram relação com a desatenção do motorista (SALVADOR, 2011). As distrações do motorista possuem três origens principais: visual (quando os olhos estão orientados para fora da estrada), mãos (mãos fora do volante) e cognitiva (mente dispersa ou focada em outra tarefa) (STRAYER; WATSON; DREWS, 2011). Por exemplo, para um motorista em uma conversa telefônica ocorre a redução no processamento de informações visuais (STRAYER; DREWS; JOHNSTON, 2003), a retirada de uma das mãos do volante e um aumento significativo da distração cognitiva (STRAYER *et al.*, 2013). Nos últimos anos, a distração por uso do celular, vem se mantendo frequente entre os condutores. A cada instante, entre 4% e 6% dos automóveis em rodovias americanas estão sendo conduzidos por pessoas falando ao celular (NHTSA, 2011). Falar ao celular enquanto dirige aumenta o risco de colisão de 4 a 6 vezes (REDELMEIER; TIBSHIRANI, 1997; STRAYER; DREWS, 2004), dessa forma, a distração pelo uso do celular leva o motorista a dirigir de maneira similar a um motorista embriagado (STRAYER; DREWS; CROUCH, 2006). Em alguns países permite-se efetuar ligações telefônicas usando *hands-free* (fones de ouvido ou viva-voz), porém seu uso também pode distrair o motorista de modo similar à utilização do celular ao ouvido (ISHIGAMI; KLEIN, 2009).

Leis vêm sendo criadas nos últimos anos, em muitos países, com o intuito de minimizar estes números de acidentes por distração e condução inadequada de veículos por motoristas imprudentes. Porém, as leis e a fiscalização não são capazes de impedir que motoristas falem ao celular, ou dirijam por longos períodos, ou mesmo, se julguem aptos a desempenhar a direção do veículo. Confiar que os próprios motoristas possam se autoavaliar quanto a sua capacidade de conduzir o veículo é algo arriscado. Em um estudo apresentado por Allstate (2011) mostra que os motoristas americanos acreditam que seu próprio conhecimento, habilidade e hábitos de direção defensiva estão bem acima dos outros motoristas das estradas. Esta mesma pesquisa revela ainda que 64% se autoavaliam como “excelentes” ou “muito bons” motoristas, sendo reduzido significativamente esse percentual para 29% quando os motoristas avaliam amigos próximos.

Os veículos atuais estão cada vez mais equipados com inúmeros sistemas que auxiliam ao motorista no ato de dirigir e visando o aumento da segurança no trânsito. Um passo para resolver o problema da existência de motoristas humanos incapazes de conduzir adequadamente um veículo e, por consequência, ter estradas livres de erros de condução humanos, é dotar os veículos da capacidade de se locomover autonomamente. Existem diversos projetos em desenvolvimento atualmente, como por exemplo o CaRINA (FERNANDES *et al.*, 2014) desenvolvido junto ao LRM do ICMC/USP e o *Google Self-Driving Car* (FISHER, 2014), que mostram o grande esforço científico ocorrido nos últimos anos rumo a veículos mais autônomos. No entanto, enquanto os volantes perdurarem em veículos, mesmo nos autônomos, existirá ainda a opção de um motorista humano requisitar o controle manual do veículo. E esta existência do volante se dá pelo fato de que veículos autônomos precisam de uma infraestrutura preparada para que operem autonomamente com segurança (BADGER, 2015), assim em certos locais, pode ainda não ser possível garantir uma condução totalmente autônoma, sendo necessária a intervenção e a condução do veículo por uma motorista humano. Então, dependendo do destino do usuário, o veículo pode não ser capaz de operar em modo autônomo todo o trajeto ou parte dele. Dessa forma o alto custo de implementação de veículos totalmente autônomos com alto grau de confiabilidade, nos leva a crer que os veículos autônomos precisarão dividir as ruas e estradas com veículos tradicionais conduzidos por humanos, pelo menos por um período ainda relativamente longo de tempo.

Quanto maior for a evolução tecnológica veicular, mais interessante e conveniente é de que o veículo seja capaz de analisar e identificar a aptidão do motorista humano para conduzi-lo. Esta pesquisa visa propor um sistema avançado de assistência ao condutor (ADAS) não intrusivo, capaz de detectar comportamentos de condutores de veículos que indiquem a sua capacidade ou incapacidade de operar o veículo. As distrações e perturbações que farão parte da validação da pesquisa são: o uso de celular, a sonolência e a embriaguez.

O problema a ser tratado é o desenvolvimento de processos e ferramentas para se adquirir e analisar dados do interior e exterior do veículo, que por sua vez, permitam observar o comportamento e o modo de direção de um motorista humano. Contribuir assim para o avanço do estado da arte de sistemas de apoio à direção na diminuição do número de acidentes causados por uma condução imprópria de veículos.

1.2 Objetivos

O objetivo geral deste trabalho é pesquisar e desenvolver técnicas de análise do comportamento de condutores de veículos em situações perigosas, permitindo assim, identificar quando o motorista está apto ou não a dirigir adequadamente o veículo e contribuindo para inibir situações de risco causadas pelo uso de celular, sonolência e embriaguez.

Como objetivos específicos e principais contribuições que se espera alcançar nesta tese,

temos os seguintes itens:

- Desenvolver um simulador que permita a coleta de dados para análise do comportamento de condutores em situações as mais realistas possíveis, considerando a distração ao volante, com uso de celular, embriaguez e sonolência. Este simulador deve permitir a captura de dados (criação de “datasets”) e a realização dos estudos sobre o comportamento do motorista e da condução do veículo, sem pôr em risco as pessoas envolvidas neste processo;
- Pesquisar técnicas de integração dos dados referentes ao comportamento do motorista, juntamente com os dados referentes ao comportamento do veículo (por exemplo, usando uma unidade inercial - IMU);
- Definir e propor um conjunto de variáveis e parâmetros capazes de serem usados na análise do comportamento e das ações do motorista. Este conjunto de informações deve monitorar as ações do motorista através do uso de um sensor 3D (dispositivo não invasivo), bem como, deve monitorar os parâmetros do veículo (sensores de velocidade, aceleração/frenagem e inerciais), e também da pose do veículo em relação a via (condução correta deste dentro dos limites e marcações que definem a via);
- Propor e desenvolver métodos de análise das ações do motorista, baseado em um sensor 3D (por exemplo: análise da posição das mãos e do uso de celular);
- Analisar as diferentes informações coletadas e determinar o melhor conjunto de dados que seja representativo e adequado para a identificação de cada um dos diferentes problemas de distrações e perturbações considerados. Avaliar as variáveis e parâmetros considerados, usados para a identificação do uso de celular, embriaguez e sonolência, sejam em separado ou em conjunto;
- Propor e desenvolver classificadores, baseados em Algoritmos de Aprendizado de Máquina, que permitam identificar as distrações e perturbações do motorista. Para isto devem ser consideradas questões referentes a seleção de variáveis, otimização dos parâmetros do classificador (usando Algoritmos Genéticos), organização dos dados e tipo de classificador empregado (classificadores binários ou multi-classe), e questões temporais (duração do período de distração e grau de certeza associado a detecção de cada distração);
- Realizar experimentos e a captura de dados com diferentes motoristas, considerando as diferentes situações de distração e perturbação destes, ou seja, considerando o uso do celular, embriaguez e sonolência;
- Aplicar os métodos desenvolvidos de classificação do comportamento do motorista, junto aos dados capturados, e verificar o desempenho do sistema para um conjunto amplo de diferentes configurações e parâmetros, analisando o desempenho do mesmo através de medidas de acurácia, mas também, levando em consideração a preocupação de evitar e de

minimizar o número de falsos alarmes (Falsos Positivos), que podem ser muito prejudiciais para uma adoção de um sistema ADAS como este proposto.

1.3 Aplicações

A solução resultante deste trabalho objetiva na identificação do comportamento seguro e inseguro de um motorista. A partir do sistema deve-se poder identificar momentos em que o motorista não está apto, por motivos de uso de celular, sonolência ou embriaguez, a desempenhar suas funções de maneira segura a ele e aos demais usuários das estradas. O sistema utiliza a postura corporal do condutor e a qualidade da forma como está conduzindo o veículo para identificar um estado de risco. Este trabalho resume-se a detecção do estado de risco, mas de posse desta informação, pode-se acionar alertas sonoros para que o motorista volte a um estado mais seguro (no caso do uso do celular e sonolência); ou avisar diretamente a polícia/central de uma empresa de transportes da direção perigosa; ou então, ativar outros sistemas presentes no veículo para que assumam o controle do mesmo (como por exemplo, ativar um sistema redutor da velocidade do veículo e de *parking* para parar o veículo em um lugar seguro).

1.4 Estrutura da tese

Este trabalho apresenta um sistema ADAS voltado à identificação de um motorista distraído ou perturbado pelo uso de celular, sonolência ou embriaguez. A estrutura principal desta tese pode ser observada na Figura 1. No Capítulo 2, trabalhos relacionados à detecção do uso de celular no trânsito, à sonolência, à embriaguez, à qualificação e identificação do estilo de direção e outros sistemas ADAS são apresentados. Técnicas, conceitos e sensores empregados a fim de alcançar os objetivos deste trabalho são apresentados no Capítulo 3. O Capítulo 4 apresenta as arquiteturas de sistemas propostos. O simulador criado e dados que possibilitaram a experimentação dos sistemas propostos são descritos no Capítulo 5. Os experimentos e resultados são mostrados no Capítulo 6. As principais conclusões e propostas de trabalhos futuros são mostradas no Capítulo 7.

Figura 1 – Estrutura deste trabalho.

Fonte: Elaborada pelo autor.

TRABALHOS RELACIONADOS

Nesta seção, alguns trabalhos relacionados com a pesquisa proposta são apresentados e discutidos.

2.1 Sistemas Avançados de Assistência ao Condutor

Advanced Driver Assistance Systems (ADAS) ou Sistemas Avançados de Assistência ao Condutor são tecnologias destinadas a auxiliar o motorista em sua tarefa de conduzir o veículo. Vários desses sistemas, que hoje estão sendo pesquisados, são baseados em Visão Computacional ([ARMINGOL *et al.*, 2007](#)). A amplitude do ADAS vai além do próprio veículo, pois deve ser considerado como um conjunto de sistemas e subsistemas que permitam automatismos das autoestradas ([BROOKHUIS; WAARD; JANSSEN, 2001](#)). Assim, o propósito do ADAS é de melhorar a segurança do veículo e, conseqüentemente, a segurança rodoviária, reduzindo ou, até mesmo, eliminando possíveis erros do condutor ([TAPANI, 2007](#)). Com sua utilização obteve-se uma significativa redução no número de vítimas fatais em acidentes rodoviários, nos últimos 30 anos. Houve redução, em países europeus, entre 30% e 40% nos acidentes fatais em decorrência da implantação de tais sistemas ([LEFÈVRE; LAUGIER; IBAÑEZ-GUZMÁN, 2011](#)).

Os ADAS podem ser divididos em subcategorias conforme sua função de apoio à direção. De acordo com [Floudas *et al.* \(2005\)](#), uma categorização que pode ser feita é a seguinte:

Controle Lateral: Estes sistemas procuram prever saídas de pista não intencionais ou mudanças de faixa perigosas. Pode-se citar como exemplos:

- **Aviso de Saída de Faixa** (*Lane Departure Warning*): é utilizado na ajuda aos motoristas em manter-se dentro da faixa de rodagem. Sua função, portanto, é avisar quando o veículo está saindo da pista atual de maneira inadequada ou perigosa, verificando-se para tanto, objetos próximos e o acionamento das luzes de sinalização

(indicando a intenção de troca de faixa) (HSIAO *et al.*, 2009). Estes sistemas são projetados para minimizar colisões causadas principalmente por: erro do motorista, distrações ou mesmo sonolência.

- **Ponto Cego** (*Blind Spot*): monitora a área lateral ao veículo, local onde os condutores não conseguem ver pelos retrovisores exteriores (CARDARELLI, 2012). Se o sistema detectar um veículo nesta região, sinais acústicos e/ou visuais alertam o motorista do risco de colisão.

Controle Longitudinal: Abrange controle de velocidade, de distâncias e tempos de reação. Exemplos desses tipos de aplicação são:

- **Controle de velocidade de Cruzeiro Adaptativo** (*Adaptive Cruise Control*): ajusta automaticamente a distância para o veículo a frente e a velocidade, utilizando-se de sensores para medir a distância longitudinal entre os veículos (MANIKOTH *et al.*, 2012). Seu objetivo é melhorar o conforto na condução, reduzir acidentes de trânsito e aumentar o fluxo de tráfego (velocidade média).
- **Assistente de Luz Alta** (*High-beam Assist*): o sistema detecta a luz dos veículos que circulam em sentido contrário ou os que estão à frente, alternando entre luz alta e média (ALCANTARILLA *et al.*, 2011).
- **Reconhecimento de Sinais de Trânsito** (*Traffic Sign Recognition*): tem por objetivo ajudar o motorista a detectar e classificar os sinais de trânsito que definem perigos e limitações das estradas (MARENCO *et al.*, 2009). Com este sistema pode-se reconhecer, por exemplo, o limite de velocidade ou curvas acentuadas à frente.

Auxílio ao Estacionamento e Marcha Ré: São sistemas de detecção de obstáculos em situações de baixa velocidade. O Assistente de Estacionamento (*Park Assist*) é um exemplo de abordagem dessa categoria, sendo um recurso que permite ao veículo estacionar sozinho, envolvendo direção automática e controle de velocidade (TOTZKE *et al.*, 2011).

Melhoramento da Visão: Dá suporte ao motorista em situações de visibilidade reduzida. Como exemplo, pode-se citar a Visão Noturna (*Night Vision*) cujo objetivo é ampliar a capacidade do motorista em visualizar obstáculos durante as horas escuras do dia, utilizando-se de câmeras sensíveis a radiação infravermelha (MANIKOTH *et al.*, 2012). Enquanto a maioria dos faróis dos automóveis pode iluminar a estrada cerca de 60 metros à frente do veículo, este sistema permite aos condutores obter informações sobre o tráfego de até 150 metros de distância.

Anti-colisão: São sistemas que visam evitar ou minimizar os possíveis problemas de um acidente em curso, atuam quando o motorista pode não ter mais condição de reação. Como exemplos dessa categoria de ADAS, podem-se citar:

- **Airbags:** São sacos inflados, muito rapidamente, com ar no momento que ocorre a colisão do veículo (TANGO; CARREA; GOBETTO, 2000). Os *airbags* mais comuns são os frontais e laterais que protegem os ocupantes do veículo.
- **Aviso de Colisão Frontal** (*Forward Collision Warning*): o sistema fornece avisos (visuais, auditivos ou vibrações) ao condutor quando detectado um provável acidente iminente com o veículo à frente (HOJJATI-EMAMI; DHILLON; JENAB, 2012).
- **Detecção de Objeto** (*Object Detection*): visa detectar objetos, os mais diversos a se colocar na trajetória do veículo. Um exemplo desse tipo de abordagem é a detecção de pedestres a fim de alertar o condutor. Em alguns casos pode ainda executar frenagem automática ou mesmo abrir *airbags* externos (no caso de uma colisão inevitável) (GERONIMO; SAPPA; LÓPEZ, 2009).

Monitoramento da Superfície da Estrada: Tem como principal objetivo alertar o condutor em caso de más condições da pista. O sistema pode emitir avisos ou estar diretamente relacionado ao sistema de controle de velocidade, por exemplo, ajudando o motorista a manter-se na velocidade apropriada para as condições atuais da pista, ou até mesmo, detectando buracos na via de rodagem (ERIKSSON *et al.*, 2008).

Monitoramento do Motorista: Concentra-se no estado fisiológico e comportamental do motorista, informando quando a tarefa de dirigir não pode ser adequadamente executada pelo condutor. O monitoramento do estado condutor pode ser dividido em duas ramificações principais: a detecção de distrações e a identificação de sonolência (KUTILA, 2006, p. 23). Esta tese é direcionada para a geração de soluções para esta categoria de ADAS.

Na Figura 2, são mostradas possíveis localizações de sensores ou câmeras de aquisição de vídeo do automóvel de ADAS. Pode-se ainda ver que os sistemas de Aviso de Saída de Faixa, Assistente de Luz Alta, Reconhecimento de Sinais de Trânsito, Aviso de Colisão Frontal e Detecção de Objeto podem utilizar uma mesma entrada de vídeo, partindo-se de uma câmera montada frontalmente. Embora, cada um dos sistemas tenha diferentes exigências de resolução e de processamento.

Nas subseções seguintes encontram-se trabalhos relacionados ao monitoramento do motorista e direção insegura.

2.2 Sistemas de Detecção de Distrações e Perturbações

Esta subseção trata de trabalhos que se relacionam diretamente com os objetivos deste trabalho: detecção de embriaguez (Seção 2.2.1), sonolência (Seção 2.2.2) e do uso de celular pelo motorista (Seção 2.2.3). Além disso, sistemas avaliadores e classificadores da condução do veículo são apresentados na Seção 2.2.4.

Figura 2 – Visão de alguns Sistemas Avançados de Assistência ao Condutor (ADAS).

Fonte: Adaptada de [Kisacanin \(2011\)](#).

2.2.1 Detecção de Embriaguez

Há diferentes métodos para se detectar um motorista embriagado. O sistema de detecção de motorista embriagado desenvolvido por [Haile \(1987\)](#) e aprimorado em [Haile \(1992\)](#) utiliza-se de um caminho entre sensores (dispostos em paralelos) por onde o motorista precisa conduzir o veículo. Um motorista normal consegue se manter em uma região segura/normal dentro do caminho. Os sensores são portáteis, ociosos e devem ser posicionados sobre o solo. Quando o pneu do veículo fica sobre um sensor seu volume interno é reduzido e lâmpadas de sinalização são acionadas, ou mesmo, outros aparelhos, indicando assim, a inaptidão do condutor.

Tratando-se de sistemas de controle de ignição, pode-se citar o sistema proposto por [Carroll, Bellehumeur e Carroll \(2013\)](#) que verifica o teor de álcool no sangue do motorista empregando um sensor transdérmico incorporado ao volante do veículo. Estando acima de um limiar, o sistema impede que o veículo ligue. A medição do teor de álcool é feita por algoritmos neurais previamente treinados a receber as informações dos sensores. Outra proposta desta natureza é a de [Edmonds e Hopta \(2001\)](#) que funciona de forma similar, porém, o teor de álcool é medido pela intensidade dos comprimentos de ondas que emergem do dedo do condutor. Um microprocessador correlaciona a intensidade coletada com o teor de álcool. Para garantir que um dedo do principal condutor do veículo é o que está sendo analisado, o sistema faz a identificação da impressão digital.

O sistema de [Kathar e Bhuyar \(2016\)](#) baseia-se em um sensor para respiração humana que detecta se o motorista está embriagado ou não (acima de um limite permitido de álcool na respiração). O sensor está instalado e mede a concentração do álcool na região próxima ao acento do motorista.

[Murata et al. \(2011\)](#) apresentaram um sistema não intrusivo que detecta se um motorista está sob influência de álcool pela medição do pulso do motorista. Uma janela de 180 segundos é utilizada para comparar os dados do motorista adquirido com uma base de dados de pessoas em

estado normal e embriagado. Assim, o sistema consegue predizer entre um motorista embriagado e um normal.

A solução apresentada por [Dai et al. \(2010\)](#) foca na detecção, utilizando sensores inerciais, de direção perigosa efetuada por um motorista embriagado. Um programa instalado no celular computa dados provenientes de um acelerômetro e um sensor de orientação, comparando então com um conjunto de dados com padrões típicos de um motorista embriagado para identificar direção imprópria. O sistema pode alertar a polícia local automaticamente por chamada telefônica em locais onde há cobertura de celular.

A proposta de [Shirazi e Rad \(2014\)](#) identifica se um motorista está embriagado pelo modo como conduz o veículo na faixa de rodagem. O sistema localiza o centro da faixa de rodagem e a curvatura da via, comparando-os ao ângulo do volante em um período, é possível distinguir o comportamento de um indivíduo sóbrio de um alcoolizado. O modelo de distinção entre sóbrio e alcoolizado foi obtido utilizando um conjunto de dados obtidos por motoristas em um simulador. O método gera modelos personalizados (sóbrio e embriagado) para cada motorista, ou seja, o motorista precisa ser identificado no momento em que começar a guiar o veículo. Assim, com modelos individuais, o sistema alcançou uma alta fidelidade. O método utiliza técnicas *onlines* que atualizam os modelos dos motoristas durante a condução do veículo. No caso de geração de modelos sóbrios e embriagado genéricos (servindo para qualquer motorista), o método apresenta alta incerteza na sua detecção.

2.2.2 Detecção de Sonolência

Entre os trabalhos da área, destaca-se a abordagem de [Akrouit e Mahdi \(2013\)](#) que utiliza uma câmera frontal ao motorista. A proposta localiza a face e em seguida os olhos do motorista usando a técnica de detecção de regiões Viola-Jones ([VIOLA; JONES, 2004](#)). Usando a Transformada Circular de Hough ([CAUCHIE; FIOLET; VILLERS, 2008](#)) na região dos olhos é possível identificar a íris e as duas pálpebras. O estado dos olhos é obtido pela classificação de duas características, uma não linear e outra não estacionária, permitindo assim a detecção da sonolência.

[Ebrahim et al. \(2014\)](#) apresentam um sistema que extrai características da amplitude ocular, como a própria amplitude, a energia, as velocidades máximas e médias de fechamento e da abertura dos olhos, e características atreladas à duração, como frequência de fechamento e abertura, tempo de permanência fechado e o PERCLOS ([KNIPLING; WIERWILLE, 1994](#)). PERCLOS é a taxa que os olhos do motorista permanecem fechado em um intervalo de tempo (uma das características mais presentes em sistemas de detecção de sonolência), sendo considerado risco, PERCLOS maior ou igual a 80%. Utilizando-se dados provenientes de experimentação real e simulada com 43 indivíduos em 67 horas no total, foi possível atingir a acurácia de 82% na detecção da sonolência. Os testes reais foram realizados com 18 pessoas que tiveram seus dados coletados utilizando um trajeto de 260 km em rodovias da Alemanha, durante o período

diurno (entre 9-13 horas) e com pouco tráfego nas autoestradas.

A proposta de [Dkhil et al. \(2015\)](#) detecta a sonolência com base em sinais fisiológicos (eletrodos colocados na cabeça do motorista) e na detecção do tempo de fechamento ocular. Os olhos são detectados através de Viola-Jones ([VIOLA; JONES, 2004](#)) partindo-se de imagens adquiridas de uma câmera. Em cada quadro é classificado se os olhos do motorista estão abertos ou fechados. Uma técnica *Fuzzy* é empregada para analisar os dados fisiológicos e de fechamento para detectar a sonolência.

Outro sistema é apresentado por [Lenskiy e Lee \(2012\)](#), o qual, efetua segmentação do condutor por cor da pele e utiliza o SURF (*Speeded Up Robust Feature*) ([BAY et al., 2008](#)) como modelo de extração de características faciais, utilizadas para determinar se o olho está fechado e sua frequência de fechamento. O método utiliza essas duas informações para detectar a sonolência.

O algoritmo proposto por [Kumar et al. \(2012\)](#) utiliza morfologia matemática na segmentação dos olhos do motorista, sendo a detecção do estado de olho, como sendo aberto ou fechado, possível com auxílio da utilização do filtro de Gabor (distribuição de frequência espacial local). A detecção da sonolência ocorre utilizando PERCLOS.

No trabalho sobre a detecção de sonolência de [Berri et al. \(2013\)](#), o estado dos olhos é identificado pela presença ou não da retina. A morfologia matemática é empregada na detecção do olho aberto ou fechado, e PERCLOS é utilizado para constatar a sonolência.

2.2.3 Detecção de Celular

Há esforços ativos no desenvolvimento de métodos para detecção do uso de celulares por motoristas enquanto dirigem. O modo mais simples de resolver este problema é ajudando o motorista a não responder ligações telefônicas enquanto dirige. Usualmente isto pode ser obtido utilizando um aplicativo específico para celular, mas neste caso o motorista precisa colaborar com o sistema ativamente para que funcione. O sistema “Mãos no Volante” ([CIDADES, 2012](#)), desenvolvido pelo governo brasileiro, bloqueia qualquer notificação em um celular por um período de tempo, ou seja, enquanto o motorista está conduzindo o veículo. As ligações são ainda automaticamente respondidas via mensagens de texto entre celulares, informando a pessoa que tenta contato com o motorista, de que ele está momentaneamente indisponível, pois está dirigindo. Quanto ao recebimento de SMS, o aplicativo bloqueia as notificações de mensagem. Ou seja, se o motorista receber uma mensagem de texto, não será notificado, podendo visualizá-la assim que desabilitar o aplicativo. No entanto, quem enviou o SMS não recebe a notificação de que a pessoa está dirigindo. O início e duração do bloqueio são indicados pelo motorista. O sistema por si só não evita o uso indevido do celular.

A abordagem de [Yang et al. \(2011\)](#) utiliza um *software* funcionando no celular para capturar e processar sinais sonoros de alta frequência, enviados pelo equipamento de som do

veículo. Os sinais sonoros são usados para mensurar a posição onde o celular está, assim, pode-se saber se é o motorista quem o está utilizando, bloqueando assim a operação do aparelho. A proposta obteve precisão de classificação de mais de 90% em detectar o motorista, sendo feita experimentações com dois modelos de celulares distintos e empregando dois veículos diferentes. Mas este sistema depende do sistema operacional do celular e um *software* precisa estar continuamente habilitado. A grande vantagem da técnica é o seu funcionamento mesmo utilizando o celular com fones de ouvido (*hands-free*).

O método de [Park et al. \(2018\)](#) utiliza um *software* no celular do motorista e sensores do aparelho (acelerômetro, giroscópio, bússola e microfone) para detectar eventos, os eventos possibilitam saber se um motorista está querendo utilizar o celular enquanto dirige. Exemplo de eventos detectados são “andando em direção do veículo”, “em pé próximo do veículo e abrindo uma das portas do veículo”, “entrando no veículo”, “fechando a porta” e “iniciando o motor”. Para diminuir a chance de falsa detecção dos eventos, o método emprega uma máquina de estados da sequência lógica dos eventos. A posição, por onde o aparelho celular entrou no veículo, é identificada também, portanto, o sistema é capaz de distinguir entre um motorista e um passageiro. Todas as atividades são identificadas por meio da análise e fusão dos sensores, possibilitando o bloqueio total do aparelho caso uma distração do motorista em curso seja detectada.

O método de [Deshmukh e Dehzangi \(2017\)](#) é um método intrusivo e utiliza eletroencefalografia (sinais fisiológicos) para identificar quando o motorista está falando ao celular ou não. Porém, este sistema precisa de uma configuração inicial dos sensores e alguns eletrodos (intrusivo) distribuídos pelo peito do motorista.

O trabalho feito por [Veeraraghavan et al. \(2007\)](#) tem por objetivo alcançar a detecção e classificação de atividades do condutor de um automóvel utilizando um algoritmo de visão computacional que detecta movimento relativo baseado na segmentação da pele do motorista. Outro estudo utilizando segmentação de pele é o apresentado por [Berri et al. \(2016\)](#), onde a pele do motorista é detectada através de uma segmentação adaptativa. Da imagem binarizada da pele do motorista, são extraídas duas características, o Percentual de Mão (PM) e o momento de Inércia (primeiro Momento de HU) (HU, 1962). O PM é a taxa de *pixels* de pele em duas regiões específicas. Por meio dessas duas características, pode-se prever o uso do celular em cada quadro. Sendo observada a presença de celular em pelo menos 65% dos quadros processados em um período de 3 segundos, é declarado o uso do celular pelo motorista (no período) pelo sistema. O sistema é híbrido e utiliza ainda um sistema para detecção do movimento das mãos/braços (fluxo ótico) como parametrizador do sistema de reconhecimento de padrões. Ambos os métodos usam uma câmera para adquirir uma sequência de imagens do motorista, mas o primeiro emprega uma câmera lateral do motorista enquanto que o segundo utiliza uma câmera frontal a ele. Sistemas baseados em imagens tem algumas limitações devido a problemas com iluminação e sombras.

Os métodos de [Artan et al. \(2014\)](#) e [Seshadri et al. \(2015\)](#) detectam o uso do celular utilizando câmeras infravermelhas de vigilância, na tentativa de evitar alguns problemas com a iluminação. Os sistemas localizam a face do motorista (ROI). A ROI é classificada entre o uso e não uso do celular enquanto dirige utilizando um classificador para vetores de Fisher (FV) ([JAAKKOLA; HAUSSLER, 1999](#)) (características) no primeiro sistema e o segundo utiliza HOG ([DALAL; TRIGGS, 2005](#)) como características extraídas.

O método RGB-D de [Craye e Karray \(2015\)](#) para detecção do motorista falando ao celular emprega um Microsoft Kinect v2 posicionado em frente do motorista como sensor. Características são extraídas da posição dos braços (à frente, à direita, à esquerda, acima ou abaixo), face (orientação, boca e mandíbula) e olhos (direção do olhar e fechamento). Este sistema obteve 90% de acurácia no reconhecimento e detecção da distração.

2.2.4 Detecção de Direção Insegura

O sistema apresentado por [Lee et al. \(2006\)](#) é capaz de monitorar o motorista usando duas câmeras de captura de imagens, uma posicionada frontalmente ao motorista e outra na frente do veículo. Usando a câmera interna, o sistema detecta a direção para onde o condutor está olhando. A direção da pista de rodagem à frente é detectada utilizando a segunda câmera (externa) no veículo. Estas duas orientações extraídas são convertidas para um mesmo sistema de coordenadas e correlacionadas. Assim pode-se inferir o estado da direção entre: segura, de risco e muito risco. Um Modelo Oculto de Markov (HMM) ([MITROVIC, 2005](#)) analisa a mesma correlação para detectar 4 padrões de direção, que são: pista em linha reta, curva na pista, mudança de faixa de rodagem e fazendo a volta.

O sistema detector de estilo de direção de [Johnson e Trivedi \(2011\)](#) classifica o motorista do veículo em típico (não agressivo) e agressivo, utilizando para isso um *software* em funcionamento em um celular devidamente posicionado no para-brisa do automóvel. O celular precisa conter uma câmera, um acelerômetro, um giroscópio, um magnetômetro e um GPS para que o sistema opere. O sistema detecta a agressividade ou não do motorista efetuando o mapeamento não linear (*Dynamic Time Warping*) entre os dados adquiridos atualmente pelos sensores e os dados provenientes de um pequeno conjunto de treinamento (modelos).

O aplicativo DriveSafe ([BERGASA et al., 2014](#)), disponível para celular do tipo iPhone, detecta comportamentos desatentos e de direção perigosa. O *software* ainda mensura a qualidade da direção e efetua alerta quando essa pontuação é baixa. A definição da qualidade da direção é baseada na detecção de sonolência e distrações. A sonolência é detectada através do modo como o motorista conduz o veículo pela pista de rodagem (posição e saída involuntária da pista), que é detectada através da câmera traseira do celular posicionada para observação das faixas de rodagem a frente do veículo. A detecção de distração utiliza o acelerômetro, o giroscópio e o GPS, assim consegue-se detectar os níveis de aceleração, frenagem e das guinadas efetuadas pelo veículo. O sistema não opera em trânsito urbano de cidades, apenas em rodovias, por isso

funciona apenas quando a velocidade do veículo é superior a 50 km/h.

O sistema proposto por [Pinilla, Quintero e Premachandra \(2014\)](#) diagnostica a direção do motorista utilizando um GPS para se adquirir a posição, velocidade, aceleração e as guinadas efetuadas pelo veículo. Esses dados são comparados com regras para a condução, na região, previamente catalogadas, sendo as regras baseadas na lei e em técnicas de segurança. Um sistema baseado em Lógica *Fuzzy* é responsável por qualificar a direção efetuada.

2.3 Considerações

Este capítulo apresentou trabalhos relacionados à detecção do uso de celular no trânsito, sonolência, embriaguez e direção insegura dentro do escopo definido para este trabalho. Os sistemas levantados nesta seção atuam no monitoramento do motorista, na qualificação da direção ou antes do veículo ser ligado, permitindo assim a detecção de perturbações e distrações. A Tabela 1 mostra a comparação entre os sistemas apresentados nesta seção.

Alguns dos trabalhos levantados ([AKROUT; MAHDI, 2013](#); [DKHIL *et al.*, 2015](#); [LENS-KIY; LEE, 2012](#); [KUMAR *et al.*, 2012](#); [BERRI *et al.*, 2013](#); [VEERARAGHAVAN *et al.*, 2007](#); [SESHADRI *et al.*, 2015](#); [BERRI *et al.*, 2016](#); [LEE *et al.*, 2006](#); [CRAYE; KARRAY, 2015](#)) que monitoram o motorista empregam câmeras RGB para esse fim. Essas câmeras dependem de iluminação e, além disso, de certa constância e homogeneidade da iluminação para detectar e segmentar o motorista corretamente. A acurácia da segmentação do motorista pode ser prejudicada pela presença de partes internas do veículo com cores próximas a pele humana. Em uma situação real a incidência da iluminação pode variar o que torna este um problema muito relevante. Por exemplo, as partes onde a luz do sol atinge, os *pixels* da imagem adquirida saturam, ou seja, apresentam a tendência de apresentar a cor branca como pode ser observado na Figura 3. O movimento do veículo causa ainda o deslocamento da região incidência de luz. Todos esses problemas dificultam a utilização de um sistema baseado em uma câmera na aquisição de dados do motorista para ambiente real.

Uma câmera é capaz de capturar informações 2D da cena, ou seja, não temos noção da profundidade dos *pixels* em relação a câmera. A utilização de um Sensor 3D que já utilize uma iluminação própria não visível e tolerante a incidência do sol é interessante para esse tipo de solução. Assim, pode-se utilizar dados 3D (nuvem de pontos) para rastrear os movimentos e a atuação do motorista dentro do veículo, sem qualquer influência de cor ou intensidade de luz capturada por um dispositivo passivo. O Sensor 3D pode ser instalado em uma posição frontal ao motorista e que não prejudique a dirigibilidade do veículo, mas dotando assim o sistema da capacidade de observar todas as partes superiores do corpo humano (cabeça, pescoço, ombros, braços e mãos) em suas posições habituais na condução do veículo, possibilitando assim, a percepção de poses e ações não usuais para um motorista em perfeitas condições de dirigibilidade. [Craye e Karray \(2015\)](#) é um exemplo de método que emprega um Sensor 3D (Kinect v2) porém

Tabela 1 – Tabela de comparação entre os trabalhos relacionados.

		Carroll, Bellehumeur e Carroll (2013)	Murata <i>et al.</i> (2011)	Kathar e Bhuyar (2016)	Dai <i>et al.</i> (2010)	Haile (1992)	Shirazi e Rad (2014)	Dkhl <i>et al.</i> (2015)	Ebrahim <i>et al.</i> (2014)	Akrouf e Mahdi (2013)	Lenskiy e Lee (2012)	Kumar <i>et al.</i> (2012)	Berri <i>et al.</i> (2013)	Cidades (2012)	Yang <i>et al.</i> (2011)	Park <i>et al.</i> (2018)	Artan <i>et al.</i> (2014)	Seshadri <i>et al.</i> (2015)	Craye e Karray (2015)	Deshmukh e Dehzangi (2017)	Veeraraghavan <i>et al.</i> (2007)	Berri <i>et al.</i> (2016)	Lee <i>et al.</i> (2006)	Johnson e Trivedi (2011)	Bergasa <i>et al.</i> (2014)	Pinilla, Quintero e Premachandra (2014)															
Detecção	Embriaguez	X	X	X	X	X	X																																		
	Sonolência							X	X	X	X	X	X													X															
	Celular ao ouvido													X	X	X	X	X	X	X	X	X																			
	Celular (viva voz)													X	X	X				X																					
	Direção insegura (geral)																						X	X	X	X															
Abordagem	Segmentação de pele										X										X	X																			
	Viola-Jones							X	X		X	X									X	X	X																		
	Característica invariantes (SIFT, SURF, HOG, FV)										X						X	X																							
	Transformada Circular de Hough									X	X																														
	Morfologia matemática											X	X																												
	Mapeamento não linear entre duas séries temporais (DTW)																								X																
	Expressão facial																			X																					
	Amplitude da boca											X												X																	
	Comportamento dos olhos							X	X		X	X	X							X			X																		
	Orientação da cabeça																			X			X																		
	Detecção de movimento																				X	X	X	X	X	X															
	Rastreamento da posição do veículo na pista de rodagem																						X		X																
	Baseado em sensores de movimento/orientação			X												X								X	X	X															
	Pose e esqueletização 3D															X									X	X															
	Posição dos braços/mãos																X																								
	Avaliação da direção				X	X	X																	X	X	X															
	Avaliação de sinais biológicos	X	X	X				X																																	
	Utiliza dados da telemetria do veículo																										X														
	Sons de alta frequência															X																									
	Bloqueio do uso do celular por um determinado período														X																										
	Intrusiva							X	X												X																				
Meio	Monitoramento do motorista	X	X	X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X																	
	Qualidade da direção				X	X	X																X	X	X	X															
	Motorista é testado antes de ligar o veículo	X																																							
Restrições	Impede passageiros ao lado do motorista				X	X		X	X					X	X	X					X																				
	Precisa do consentimento do motorista para operar				X	X		X	X					X	X	X					X				X	X															
	Precisa de iluminação (passiva)								X	X	X	X			X			X	X		X	X	X	X	X	X															
	Trabalha sem possibilidade de reeducação do motorista distraído																X	X																							
	Não garante que o motorista está sendo avaliado	X												X							X																				
	Grande tempo ao volante altera o resultado do sistema		X																																						
Utiliza modelos personalizados para cada motorista						X																																			
Sensores		I	T	P	I	A	#	M	I	C	8	E	I	C	I	C	I	C	I	!	I	A	II	I	C	K	6	E	I	C	I	C	2	C	I	C	I	!	@		
						I	G			#	E									I	G	I	B	I										I	A	!	I	@	#	V	
																																				I	B	I	A	I	@
																																				I	B	I	A	I	@
																																				I	B	I	A	I	@

Legenda: X - Sim, C - Câmera (RGB), I - Câmera (IR), A - Acelerômetro, G - Giroscópio, @ - GPS, T - Transdérnico
 K - Kinect, M - Sensor de massa, B - Bússola, P - Medidor de Pulso, ! - Microfone, E - Eletrodos,
 V - Telemetria do veículo, # - vários.

utiliza fusão de sensores e, por isso, ainda é dependente de dados RGB e da incidência do sol.

O uso do celular, em determinadas posições, pode não ser adequadamente detectado pelo sensor interno empregado no monitoramento do motorista. Por exemplo, ao enviar um

Figura 3 – Exemplos de quadros que possuem sua dificuldade de segmentação por conta da incidência de luz solar.

Fonte: [Berri et al. \(2014\)](#).

SMS o motorista pode deixar o celular em uma região fora do campo de visão do sensor, ou mesmo, ocluído por outros objetos. Conversas telefônicas utilizando *hands-free* podem não ser detectáveis pelos dados internos do veículo (movimentação do motorista). Apesar do foco da detecção do uso de celular deste projeto utilizar o modo convencional (o de levar o celular ao ouvido), é possível diferenciar algumas situações de distração cognitiva causada pelo envio do SMS ou conversa por *hands-free* utilizando as características externas que permitam qualificar a condução do veículo e, portanto, perceber a alteração do padrão de condução.

A posição do sensor de captura dos dados no monitoramento do motorista é extremamente importante, pois permite o correto funcionamento do sistema e a obtenção de dados do motorista dentro da região mais precisa do sensor. Além disso, uma escolha inadequada pode inviabilizar a utilização do sistema em situações do dia a dia. O trabalho de [Veeraraghavan et al. \(2007\)](#) utiliza uma câmera lateral ao motorista, posicionada na porta do passageiro (parte interna do veículo). Por conta dessa escolha, não é possível haver um passageiro sentado na parte frontal do veículo (entre a câmera e o motorista), o que resulta na redução da capacidade de transporte do veículo. Por essa razão, a posição frontal ao motorista adotada no trabalho de [Berri et al. \(2016\)](#), é a posição do sensor mais usual dentre os trabalhos de monitoramento do motorista observados.

A utilização de um sistema intrusivo, como por exemplo, adotado por [Ebrahim et al. \(2014\)](#), [Dkhil et al. \(2015\)](#) e [Deshmukh e Dehzangi \(2017\)](#), também deve ser evitado. Já que esse tipo de sistema pode causar desconforto ao motorista por necessitarem de eletrodos ou outros equipamentos em contato com sua pele, além do que, necessitam obrigatoriamente do consentimento do motorista para sua operação.

A questão do consentimento do motorista é muito importante, já que não se deve necessitar de qualquer auxílio do motorista ([DAI et al., 2010](#); [HAILE, 1992](#); [DKHIL et al., 2015](#); [EBRAHIM et al., 2014](#); [CIDADES, 2012](#); [YANG et al., 2011](#); [PARK et al., 2018](#); [DESHMUKH; DEHZANGI, 2017](#); [JOHNSON; TRIVEDI, 2011](#); [BERGASA et al., 2014](#)), ou seja, o motorista não deve participar de maneira direta no sistema de identificação de distrações e perturbações.

O interessante é que o condutor do veículo nem perceba que o sistema está em operação e observando suas ações, tendendo assim, a efetuar ações mais naturais e espontâneas.

Um outro ponto relevante em perturbações e distrações do motorista é a reeducação. Um sistema ADAS não deveria apenas se limitar a detectar uma condução indevida pelo motorista, ou seja, o motorista precisa tomar conhecimento em tempo real da sua avaliação. O alerta ao motorista de sua dificuldade de direção momentânea, por conta do uso do celular ou qualquer outra distração, abre caminho para a conscientização do mesmo de que precisa conduzir o veículo com mais responsabilidade. Os trabalhos de [Artan *et al.* \(2014\)](#) e [Seshadri *et al.* \(2015\)](#) impossibilitam a reeducação instantânea do motorista por funcionarem como um radar de trânsito, ou seja, o motorista seria alertado posteriormente ao receber os avisos de infração de trânsito.

Outro ponto primordial para um sistema de detecção de momentos de risco do motorista é o de garantir que o motorista está sendo avaliado enquanto dirige. Sistemas como o de [Cidades \(2012\)](#), não possuem nenhuma verificação de ser um motorista que o está utilizado. Já o caso de [Deshmukh e Dehzangi \(2017\)](#) os eletrodos responsáveis pela detecção de distração poderiam ser instalados em um passageiro, permitindo que situações de risco na direção passem despercebidas pelo sistema. O sistema de bloqueio da ignição de [Carroll, Bellehumeur e Carroll \(2013\)](#) permite que a concentração de álcool do sangue de um passageiro seja medida, ou seja, um passageiro pode se passar por um motorista no momento em que o veículo é ligado e depois de ligado o sistema não efetua mais testes.

A constituição de um sistema generalista, ou seja, capaz de operar avaliando dados de motoristas não utilizados no treinamento e na constituição do mesmo é interessante. Sistemas como o de [Shirazi e Rad \(2014\)](#) somente operam com motoristas previamente cadastrados, ou seja, todos os novos motoristas precisam passar pela criação de um modelo personalizado que irá possibilitar a detecção de riscos na direção. Já no sistema de [Murata *et al.* \(2011\)](#) observa-se uma piora na acurácia do motorista quando fica um tempo demasiado dirigindo, ou seja, o padrão observado pelo sistema se altera prejudicando a detecção de perturbações e distrações do motorista.

O monitoramento do motorista e de seu modo de conduzir o veículo pode beneficiar diversos negócios, como por exemplo, empresas de ônibus e de aluguéis de veículos (seguradoras). Assim, pode-se ter subsídios *on line* ou *offline* (caixa preta) de como funcionários ou clientes estão conduzindo as frotas de veículos. Portanto, pode-se penalizar motorista que se expõem a riscos desnecessários.

No próximo capítulo, é apresentada a fundamentação teórica que permite a criação de um sistema de apoio ao motorista capaz de monitorar distrações (uso de celular) e perturbações (sonolência e embriaguez) de maneira não intrusiva e sem qualquer necessidade de consentimento do motorista. Mostram-se ainda os principais sensores (Seção 3.1) que se pode utilizar para a captura de dados, interna e externa do veículo.

REFERENCIAL TEÓRICO

Nesta seção apresentam-se conceitos importantes para o desenvolvimento da proposta apresentada neste trabalho.

3.1 Sensores

Um sensor é um dispositivo que responde a um estímulo físico/químico de maneira específica e que pode ser transformado em outra grandeza física para fins de medição e/ou monitoramento. Na prática, transdutor é um dispositivo que converte uma informação de algum fenômeno físico captada por um sensor em um sinal detectável, que pode ser elétrico, mecânico, ótico, entre outros.

Equipamentos sensores utilizam-se de transdutores na transformação da energia medida em outro formato de apresentação, mais simples para a utilização (MURPHY, 2000). Por exemplo, um sensor pode captar o som, a pressão ou mesmo a luz, convertendo em um sinal analógico ou digital que possa ser usado por um robô.

Existem diversos tipos de sensores (WOLF *et al.*, 2009), porém podemos classificar em dois tipos principais segundo seu método de medição: os sensores ativos emitem energia no ambiente e medem seu retorno; e os sensores passivos somente captam a energia já disponível no ambiente (SIEGWART; NOURBAKHS; SCARAMUZZA, 2011).

A tarefa a ser desempenhada deve ser levada em conta no momento da escolha do sensor, já que cada sensor possui características próprias, como por exemplo, faixa de atuação, sensibilidade, precisão e exatidão (ROMERO *et al.*, 2014). Quando não existem disponíveis sensores com características compatíveis com a tarefa, uma alternativa adotada é a fusão de múltiplos sensores. A fusão sensorial pode ser classificada em Romero *et al.* (2014): (i) redundante ou competitiva, onde, mais de um sensor com características iguais ou distintas fazem medições do ambiente para posterior confronto (normalmente utilizada quando o problema a ser tratado é

a imprecisão); (ii) complementar, sensores que medem informações distintas e que juntos são capazes de eliminar falsas detecções; e (iii) coordenada na qual se utiliza sensores em uma ordem específica, como por exemplo, sendo detectado som em alguma sala, então outro tipo de sensor é utilizado para se confirmar mais informações, como uma câmera.

Nas subseções a seguir, alguns tipos de sensores são apresentados.

3.1.1 Câmeras de Vídeo

Câmeras de vídeo são sensores que apresentam uma rica quantidade de informações (ROMERO *et al.*, 2014; MATARIC, 2014). Com este ganho de informação, é possível, por exemplo, dotar um robô de um sistema de percepção similar ao humano, calculando distâncias pelo uso de sistemas de visão estéreo (bem como através de sensores *laser* e ultrassom) e identificando a presença de elementos na cena, reconhecendo posturas e gestos, etc. As câmeras atuais capturam a informação utilizando a tecnologia CCD ou CMOS. O CCD (*charged coupled device*) possui uma matriz sensível à luz (fotodiodo) ou *pixels*. A energia projetada sobre o *pixel* é acumulada por certo tempo, sendo então congelada e transmitida a um registrador. Um circuito anexo ao registrador processa uma linha da matriz de *pixels* por vez, amplificando o sinal de cada *pixel* e convertendo para digital, gerando assim a imagem final. A tecnologia CMOS (*Complementary Metal-Oxide Semiconductor*) possui uma matriz de *pixel* similar ao CCD, mas tendo como diferença, a existência para cada *pixel*, de um registrador dedicado, que amplifica e digitaliza o sinal capturado. Não necessitando, portanto, de congelamento ou transmissão de informações para um registrador externo. As duas tecnologias vêm competindo pela preferência dos fabricantes de câmeras a alguns anos, mas o que deve ser levado em conta no momento de se optar por uma ou por outra é que a CCD possui maior sensibilidade à luz e menor apresentação de ruídos; já o CMOS pode trabalhar com uma frequência de captura mais elevada, possui menor custo de fabricação e menor consumo de energia (SIEGWART; NOURBAKHSH; SCARAMUZZA, 2011). CCD, portanto, possui uma qualidade de imagem melhor, mas CMOS vem apresentando significativas melhorias e já é possível encontrar câmeras CMOS com qualidade similar à CCD (ROMERO *et al.*, 2014).

As câmeras de vídeo permitem a aquisição de imagens, que são mais ricas em termos de informações, porém usualmente mais complexas de serem tratadas. Os sistemas de visão computacional buscam explorar as informações capturadas pelas câmeras e assim contribuir na percepção do ambiente, seja usando imagens estáticas ou através de sequências de imagens (vídeos). As câmeras permitem identificar elementos pela sua forma e contornos, cor, textura e até mesmo estimar a proximidade de obstáculos.

A iluminação é algo a se considerar quando se adota câmeras, pois esses dispositivos são passivos, ou seja, dependem de uma iluminação externa. Algumas câmeras podem ser acompanhadas de um sistema de iluminação que funcione em conjunto, sendo que o mais usual nestes casos a utilização de luz infravermelha e, portanto, não visível ao olho humano. Assim,

pode-se evitar desconforto do motorista, por exemplo, ao incidir a iluminação diretamente na sua face. As câmeras preparadas para captar a luz infravermelha geram imagens em níveis de intensidade (tons de cinza), e algumas destas podem capturar imagens RGB sob a atuação de iluminação visível (por exemplo, a luz solar).

As imagens providas por uma câmera são representações 2D da cena, ou seja, não apresentam noção de profundidade dos *pixels* capturados em relação ao equipamento. Para se obter uma descrição 3D da cena, existem diferentes abordagens, onde pode ser necessária a utilização de um par de câmeras (câmeras estéreo) e de um processo complexo para reverter esta perda da informação de profundidade, a ser abordado na Seção 3.1.5.

3.1.2 Sensores de Posicionamento

Um sensor de posicionamento muito usado em ambientes externos é o *Global Positioning System* (GPS), ou Sistema de Posicionamento Global. O GPS utiliza-se dos satélites de posicionamento que estão na órbita da terra. Esses satélites emitem mensagens contendo a hora de partida da mensagem. Assim, o GPS estima a distância que a mensagem percorreu para chegar até o receptor e, tendo recebido mensagens de pelo menos 2 satélites (SIEGWART; NOURBAKHSH; SCARAMUZZA, 2011), é possível estimar a posição do sensor na superfície da terra. O GPS é um sensor absoluto, ou seja, cada cálculo de posicionamento é independente do anterior, portanto, seu erro é local e não cumulativo (ROMERO *et al.*, 2014).

3.1.3 Sensores Inerciais

O giroscópio é um dispositivo capaz de informar a mudança de orientação. Isso faz com que seja usado para auxiliar em navegação de helicópteros e aviões (principalmente no piloto automático). Existem dois tipos de giroscópios, os mecânicos e os óticos (SIEGWART; NOURBAKHSH; SCARAMUZZA, 2011). Os mecânicos são baseados no princípio da inércia, usando um rotor suspenso por um suporte formado por dois círculos articulados. Os óticos medem a velocidade angular entre dois feixes de luz ou *lasers* emitidos de uma mesma fonte.

Um acelerômetro mede a ação de forças externas sobre ele, inclusive a gravidade. Atualmente, os acelerômetros são pequenos sistemas Microeletromecânicos (MEMS) que, através de uma pequena coluna (massa) balançante, pode medir estas forças externas que atuam sobre o sensor, medindo inclusive a atuação da força da gravidade sobre o mesmo.

Existem dispositivos que agregam um giroscópio e um acelerômetro internos, podendo até fazer fusão de mais sensores, como bússola (magnetômetro). Assim são capazes de estimar a posição relativa, velocidade e aceleração do movimento de um veículo ou robô. Esses dispositivos integrados são chamados de IMUs (*Inertial Measurement Units*), porém carregam erros cumulativos da integração do giroscópio e acelerômetro. É possível, no entanto, adotar alguma referência externa como um GPS para minimizar esse problema, bem como aplicar filtros.

3.1.4 Sonares

Os sonares são sensores ativos capazes de detectar a distância entre o sensor e objetos/obstáculos. A detecção ocorre pela emissão de um pacote sonoro ultrassônico, onde se mede o tempo em que o som demora para refletir em alguma superfície e retornar. Os sonares são equipamentos, usualmente de baixo custo, e possuem um alcance curto, detectando normalmente objetos entre 30 centímetros e 5 metros (ROMERO *et al.*, 2014). A precisão da medição também depende do ângulo em que as ondas atingem uma superfície, quanto mais agudo este ângulo maior é o erro da medição, bem como depende também do tipo de material (por exemplo, elasticidade, porosidade e rugosidade) que irá refletir as ondas sonoras. Um robô pode utilizar vários sonares, porém precisam ser acionados de modo sequencial para minimizar eventuais interferências entre eles.

3.1.5 Sensores 3D

A aquisição de dados em 3D vem possibilitando o surgimento de novas técnicas de rastreamento de objetos, esqueletização do corpo humano, reconhecimento de formas, enfim, permite uma maior compreensão da cena por meio de informações de distância. Os principais sensores que atualmente permitem a obtenção dos dados 3D são: sensores a laser, câmeras estéreo, câmeras de luz estruturada e câmeras *time-of-flight* (ToF).

Os sensores a laser são utilizados em diversos projetos em que a precisão é primordial. O Velodyne¹, por exemplo, é um sensor laser multicamadas (com múltiplos feixes) e vem sendo utilizado em veículos autônomos, como o CaRINA (FERNANDES *et al.*, 2014) e o *Google Self-Driving Car* (FISHER, 2014), para se detectar obstáculos no entorno do veículo, ou mesmo, para a construção de mapas. Porém, sensores laser possuem um alto custo (STOYANOV *et al.*, 2013).

Uma câmera estéreo é um sensor que simula a visão binocular humana e por isso, é formada por duas câmeras defasadas, que ao ser aplicado algoritmos de processamento das imagens usando os dois pontos de vistas (normalmente próximos), podem ser gerados mapas de disparidade (BRADSKI; KAEHLER, 2008). Portanto, a relação entre o mesmo ponto visto de diferentes posições de uma mesma linha de base permite estimar a profundidade dos elementos da cena (proporcional a distância/disparidade dos *pixels*). Além do mapa de profundidade, a câmera estéreo, provê a cor (RGB) da cena. As câmeras são componentes que captam a luz do ambiente (sensor passivo) e, portanto, necessitam de iluminação para conseguir operar em locais com baixa iluminação.

A câmera de luz estruturada utiliza técnica de visão estéreo ativa, ou seja, projetam uma sequência conhecida de padrões na cena e captam a deformação geométrica da forma destes padrões sobre os objetos, utilizando para tanto, uma câmera em posição diferenciada ao

¹ Mais informações sobre o Velodyne estão disponíveis em <<http://velodynelidar.com/>>.

projektor (SARBOLANDI; LEFLOCH; KOLB, 2015). O mapa de disparidade, portanto, é gerado analisando-se as distorções nos padrões. Na Figura 4, ilustra-se a medição da profundidade em câmeras de luz estruturada. Em 2010, foi lançado o Microsoft Kinect v1, um sensor baseado em luz estruturada, que é capaz de disponibilizar dados de profundidade (entre 0,8 e 4 metros) e imagem (RGB-D) a alta taxa de 30 Hz e a baixo custo (SARBOLANDI; LEFLOCH; KOLB, 2015; FANKHAUSER *et al.*, 2015). Na Figura 5, mostra-se o Kinect v1 e o posicionamento de suas câmeras e projetor. No entanto, o Kinect v1 não é capaz de medir profundidade sob a incidência de luz solar, já que seu projetor utiliza um comprimento de onda que está também presente na luz do sol e, por isso, a sua utilização é limitada a ambientes internos ou em período noturno (EL-LAITHY; HUANG; YEH, 2012).

Figura 4 – Ilustração da medição da profundidade em câmeras de luz estruturada.

Fonte: Sarbolandi, Lefloch e Kolb (2015).

Figura 5 – Visão geral do Kinect v1.

Fonte: Elaborada pelo autor.

A tecnologia ToF é baseada na mensuração do tempo em que a luz emitida pelo sensor leva para chegar a um objeto e retornar ao sensor (LEFLOCH *et al.*, 2013). Essa ideia é empregada também em alguns sensores *lasers*, mas se torna mais acessível quando empregadas em câmeras ToF. Uma câmera ToF emite iluminação (projektor) não visível a observadores humanos (ex: luz

infravermelha) periodicamente, onde, por meio de uma câmera capaz de capturar a luz emitida, verifica-se o tempo de retorno da luz refletida e, com essa informação, pode-se gerar o mapa de profundidade (SARBOLANDI; LEFLOCH; KOLB, 2015). Em 2014, a Microsoft lançou o Kinect v2, agora empregando a tecnologia ToF para mensurar a profundidade, apesar disso, manteve-se como sensor de baixo custo (FANKHAUSER *et al.*, 2015). A Figura 6 mostra o Kinect v2 e as novas posições para as câmeras e projetor. Yang *et al.* (2015) avaliou a detecção de profundidade do Kinect v2 e concluiu que sua faixa de utilização é entre 0,5 e 4 metros e, mesmo dentro deste alcance, o erro da medição varia, sendo menor que 2 mm em uma região mais precisa (entre 0,5 até 3 metros) ou até maior que 4 mm em outra, conforme mostrado pela Figura 7. Comparando-se ao seu antecessor, o Kinect v2 possui melhor resolução de captura de imagens para RGB e IR, sua precisão e acurácia é maior para mensurar profundidade e é capaz de operar sob a incidência de luz solar ao alcance máximo de 2 metros (PINTO *et al.*, 2015). A comparação entre as características do Kinect v1 e Kinect v2 pode ser observada na Tabela 2.

Figura 6 – Visão geral do Kinect v2.

Fonte: iFixit (2013).

Tabela 2 – Comparativo entre as especificações do Microsoft Kinect v1 e v2 (AL-NAJI *et al.*, 2017).

Características	Kinect v1	Kinect v2
Tipo de sensor de profundidade	Luz estruturada	ToF
Câmera RGB	640 × 480, 30 fps	1920 × 1080, 30 fps
Câmera IR	320 × 240, 30 fps	512 × 424, 30 fps
Campo de visão (imagem RGB)	62° × 48,6°	84,1° × 53,8°
Campo de visão (imagem profundidade)	57° × 43°	70° × 60°
Alcance da mensuração de profundidade	0,8-4 metros (Padrão) 0,4-3,5 metros (Próximo)	0,5-4,5 metros
Definição da esqueletização	20 <i>joints</i>	25 <i>joints</i>
Número máximo de esqueletos rastreáveis	2	6

Comparando-se o Kinect com sensores baseados em *laser* para a obtenção de dados 3D da cena, o Kinect apresenta algumas limitações, como por exemplo, menor precisão e

Figura 7 – Mapa do alcance da medição da profundidade para o Kinect v2.

Fonte: Yang *et al.* (2015).

menor região de operação, no entanto, caso o alcance e precisão do Kinect sejam satisfatórios, torna-se atrativo pelo baixo custo (HAMMERLE *et al.*, 2014). De maneira geral, Sensores 3D monoculares tendem a apresentar melhor qualidade de mensuração de profundidade (WANG; LIU; WU, 2014c).

3.2 Subtração de Fundo

A análise e compreensão de uma sequência de vídeo é um campo ativo da pesquisa, pois, muitas pesquisas se deparam com a necessidade básica de detectar objetos que se movem na cena. Os objetos que se movem são denominados de *foreground* e são extraídos de uma informação estática (ou quase estática) chamada de fundo (*background*), por este motivo o processo é denominado de subtração de fundo.

Existem diversas aplicações para a subtração de fundo. Por exemplo, pode-se utilizar em câmeras de monitoramento e vigilância (sistemas CCTV), reconhecendo novidades na cena (intrusos). No caso específico desta pesquisa, estas técnicas são muito relevantes para o monitoramento do motorista, uma vez que este se encontra em um cenário de fundo relativamente fixo (veículo) e se deseja monitorar seus movimentos (elementos que se movem na cena).

O processamento da subtração de fundo em vídeos pode ser classificado segundo a movimentação da câmera (estática ou móvel) e conforme o requisito temporal de processamento

(quadros sequenciais ou não sequenciais). A Figura 8 mostra as quatro categorias para sistemas de subtração de fundo e apresenta alguns exemplos. Há, na literatura, técnicas que são efetivas na subtração de fundo tanto para processamento sequencial e não sequencial de vídeos (ELGAMMAL; ELQURSH, 2014). Nas subseções a seguir, são discutidos métodos e técnicas de subtração de fundo em câmeras estáticas (Seção 3.2.1), câmeras em movimento (Seção 3.2.2), câmeras “quase-estáticas” (Seção 3.2.3) e para câmeras RGB-D (Seção 3.2.4).

Figura 8 – Classificação dos diferentes tipos de vídeo baseado na necessidade de movimentação e processamento.

		Mais Geral	
		Quadros não sequenciais	Quadros sequenciais
Mais Geral	Câmera estática	Processamento posterior para vídeos de vigilância	Processamento de câmeras de vigilância em tempo real (câmera de monitoramento do motorista)
	Câmera em movimento	Fundos dinâmicos ou que não se conhece antecipadamente o tipo de fundo, mas possibilitam o processamento posterior (Youtube)	Streaming de vídeo (câmera de celular, transmissão de TV, robótica, câmera acopladas em veículos (externa), Google Glass, etc)

Fonte: Elgammal e Elqursh (2014).

3.2.1 Câmera Estática

A maneira mais simples de se entender a subtração de fundo em câmeras estáticas é através da subtração de duas imagens, uma chamada de imagem de fundo (B) e a outra o frame atual processado (I), sendo a imagem resultante, portanto, detentora da distância entre os *pixels* das imagens. Aplicando-se um limiar (T) nesta imagem é possível se determinar os *pixels* da imagem em que houve alterações (*foreground* ou F). O processo de obtenção da máscara de *foreground* usando apenas um canal de codificação de cor é descrito pela Equação 3.1, sendo que para mais de um canal é preciso aplicar uma função de medição de distância (d) entre os *pixels* das imagens como mostrado na Equação 3.2.

$$F(x, y) = |B(x, y) - I(x, y)| > T \quad (3.1)$$

$$F(x, y) = d(B(x, y), I(x, y)) > T \quad (3.2)$$

Para tanto, precisa-se fazer algumas suposições: a câmera e seus parâmetros permanecem fixos; não ocorreu qualquer variação de iluminação; o fundo é visível e estático, ou seja, a imagem de fundo possui apenas fundo; o *foreground* pode ser separado do fundo apenas aplicando um limiar na imagem resultante da diferença entre eles (BOUWMANS, 2014c).

Na prática, seria necessário um ambiente completamente controlado para que seja possível garantir a iluminação constante, por essa razão, os sistemas de subtração de fundo de câmeras estáticas geram a imagem de fundo durante o processamento de uma sequência de quadros, e assim, adaptam-se as mudanças de iluminação. A Figura 9 mostra uma visão geral do processo de subtração de fundo ao qual inclui as seguintes etapas: (1) módulo de inicialização do fundo provê a primeira imagem de fundo dos N quadros de treinamento; (2) detecção de *foreground* que consiste na classificação dos *pixels* como *foreground* ou fundo, é gerada pela comparação da imagem de fundo e o quadro atual; (3) módulo de manutenção do fundo atualiza a imagem de fundo com base na imagem de fundo anterior, o quadro atual e a binarização do *foreground* detectado. Portanto, a etapa (1) é executada até se obter a primeira imagem de fundo (inicialização do sistema) e as etapas (2) e (3) são executadas até o término do processamento.

Figura 9 – Processo de subtração de fundo, onde, N é o número de frames usados na inicialização do fundo, B_t e I_t são, respectivamente, o fundo e o quadro atual no tempo t .

Fonte: Bouwmans (2014a).

A etapa de Inicialização do Fundo geralmente ocorre utilizando um conjunto de quadros para “treinamento” (geração, extração ou construção), que pode conter ou não apenas *pixels* de fundo. Os tipos de algoritmos que podem ser aplicados nesta etapa são (BOUWMANS, 2014c): em lote usando N quadros de treinamento (consecutivos ou não) (MOLINIER; HÄME; AHOLA, 2005); incremental com N conhecido; e progressivo com N desconhecido, onde, o processo gera fundos parciais, continuando a geração até que a imagem de fundo esteja completamente obtida (CHUNG; WANG; CHEN, 2002; COLQUE; CAMARA-CHAVEZ, 2011). Um dos principais desafios para a obtenção da primeira imagem de fundo se dá quando mais da metade das áreas dos quadros de treinamento contém objetos em movimento, ou seja, não pertencentes ao fundo.

A etapa de Manutenção do Fundo concentra-se em adaptar a imagem de fundo as mudanças, que ocorrem na cena, ao longo do tempo. As questões que permeiam esta etapa são:

- *Esquema de manutenção*: Os esquemas de manutenção da imagem de fundo encontrados na literatura são: o cego, o seletivo e os esquemas adaptativos *fuzzy* (BAF; BOUWMANS; VACHON, 2008). Os esquemas cego e seletivo utilizam normalmente um filtro IIR (*Infinite Impulse Response*), como descrito na Equação 3.3, onde, τ é a taxa de aprendizado (entre 0 e 1), B_t e I_t são a imagem de fundo e o quadro atual no tempo t , respectivamente. Para o cego, o valor de τ é constante para todos os *pixels* analisados; já para o seletivo, há um valor de τ (maior) aplicado sobre os *pixels* que são fundo e outro valor (menor)

aplicado aos *pixels* que são *foreground*. A motivação do desenvolvimento do esquema seletivo foi de sanar a principal desvantagem do método cego que é o de computar fundo e *foreground* com igual importância, poluindo assim, a imagem de fundo resultante. Porém, o esquema seletivo depende da correta classificação do *pixel* (*foreground*/fundo) para apresentar vantagens, podendo até zerar o valor de τ nos *pixels foreground* se o erro de classificação for desprezível. Este problema pode ser minimizado, ainda, utilizando o esquema adaptativo *fuzzy*, ao qual mensura a incerteza da classificação e aplica uma graduação na regra de atualização, usando para tanto, o resultado da detecção de *foreground*. A Equação 3.4 rege o esquema de manutenção *fuzzy*, onde, μ_F e μ_B são respectivamente os valores de confiança *fuzzy* para o *pixel* ser *foreground* ou fundo, sendo que podem variar entre 0 (confiança mínima) e 1 (máxima), e τ é constante para todos os *pixels* conforme esquema cego. O μ_F é calculado pela utilização da equação $\mu_F = 1 - \mu_B$.

$$B_{t+1}(x, y) = (1 - \tau)B_t(x, y) + \tau I_t(x, y) \quad (3.3)$$

$$B_{t+1}(x, y) = \mu_F B_t + \mu_B ((1 - \tau)B_t(x, y) + \tau I_t(x, y)) \quad (3.4)$$

- *Taxa de aprendizado*: A taxa de aprendizado (τ) dita a velocidade de adaptação para as mudanças na cena (por exemplo, a alteração na iluminação) (BOUWMANS, 2014b). Essa taxa pode ser: fixa (ZANG; KLETTE, 2003; WHITE; SHAH, 2007), ajustada por ferramentas estatísticas (ZANG; KLETTE, 2003; LINDSTRÖM *et al.*, 2006) ou método *fuzzy* (POURREZA, 2008; MADDALENA; PETROSINO, 2010).
- *Mecanismo de manutenção*: É interessante se certificar de que cada *pixel* declarado como fundo é realmente fundo, para então, atualizar este *pixel* na imagem de fundo ($B_t(x, y)$). Com essa ideia em mente, algumas soluções de subtração de fundo (ZHANG *et al.*, 2005; WANG, 2005) contam o número de vezes consecutivas em que um *pixel* é classificado como fundo, sendo um número representativo, a atualização do fundo é efetuada. Outra ideia aplicável ao mecanismo de manutenção do fundo é a de buscar por componentes (grupos de *pixels* vizinhos) em que sejam transferidos de *foreground* para fundo depois de certo tempo (LINDSTRÖM *et al.*, 2006), ou seja, aplicam-se detectores CUSUM, por exemplo, nos componentes *foreground* e os componentes que permanecem estáticos por alguns ciclos de processamentos são incorporados a imagem de fundo.
- *Frequência de atualização*: A manutenção do esquema pode ser efetuada a cada novo quadro processado, porém, existem soluções que buscam identificar significativas alterações na cena para, então, permitir a atualização dos *pixels* do fundo (PORIKLI; TUZEL, 2003; MAGEE, 2004).

A etapa de detecção de *foreground* consiste em rotular (classificar) os *pixels* em fundo ou *foreground*. A classificação dos *pixels* ocorre usualmente com a aplicação da Equação 3.2,

onde, a distância entre os *pixels* do fundo e do quadro atual são medidos e, sendo acima de um limiar de corte, o *pixel* é considerado *foreground*. É interessante efetuar um processamento anterior à classificação dos *pixels*, onde se pode, por exemplo, ajustar a geometria e intensidade (RADKE *et al.*, 2005) do quadro analisado. É possível empregar algumas estratégias de limiar de corte. O limiar pode ser: fixo, conforme utilizado na Equação 3.2, onde, o mesmo limiar é aplicado a todos os *pixels* classificados; um limiar adaptativo para cada *pixel* pode ser obtido através da variância no *pixel* (WREN *et al.*, 1997); ou mesmo, pode-se utilizar limiares *fuzzy* (CHACON-MURGUIA; GONZALEZ-DUARTE, 2012). Após a aplicação do limiar, é obtida a máscara de *foreground*, ao qual, pode ser melhorada em um processamento posterior. Esse processamento pode utilizar operadores morfológicos (STRINGA, 2000) (removendo *pixels* isolados) ou outro conceito *fuzzy* (SIVABALAKRISHNAN; MANJULA, 2010) (inferência entre as máscaras atual e anterior).

A segmentação entre elementos de fundo e de frente em imagens pode ser realizada em 2D (imagens tradicionais), porém também podem ser aplicados algoritmos de segmentação de elementos em nuvens de pontos 3D, a fim de separar os elementos da cena (*foreground* e *background* 3D). No caso de mapas de profundidade ou nuvens de pontos 3D, a segmentação pode considerar a profundidade dos elementos da cena, facilitando o processo de segmentação. A subtração de fundo em câmeras RGB-D será detalhada na Seção 3.2.4.

3.2.2 Câmera em Movimento

Para o caso de processamento de vídeo disponibilizado por câmeras que se encontram em movimento, a ideia aplicada no processamento é utilizar algoritmos de detecção direta do *foreground*. Estes algoritmos detectam objetos específicos na cena, normalmente, processando quadro a quadro do vídeo, sem fazer qualquer restrição temporal (não sequencial), ou seja, cada quadro é processado de maneira independente como sendo imagens sem qualquer relação. Contudo, alguns métodos usam a informação temporal para reduzir o número de falsas detecções (ELGAMMAL *et al.*, 2002). A constituição de uma imagem de fundo e a sua posterior subtração do quadro atual para se obter o *foreground* passa a ser descartada, já que o fundo e o *foreground* podem se movimentar de um quadro para outro.

Outra questão relevante, para sistemas que fazem uso de câmeras móveis, é que sua robustez para aplicações no mundo real é menor em relação às soluções que empregam câmeras estáticas, e o desempenho é ainda mais prejudicado quanto menor forem os objetos alvo ou quando há oclusões destes (ELGAMMAL; ELQURSH, 2014). Exemplos de sistemas que utilizam câmeras móveis são a detecção de faces (JAIN; LEARNED-MILLER, 2011) e a detecção de pedestres (DOLLAR *et al.*, 2009), com acurácias declaradas para o estado da arte entre 60% e 70% e com pelo menos um falso positivo por frame processado em ambiente real.

O sucesso da detecção e subtração de fundo em câmeras móveis depende de se utilizar e integrar vários sinais perceptíveis na segmentação do *foreground*, que são (ELGAMMAL;

ELQURSH, 2014): *descontinuidade de movimentação*, onde, a ideia é detectar objetos e partes da imagem que estão se movendo independentemente do restante da cena; *descontinuidade de profundidade*, ou seja, detectar regiões que estão mais próximas da câmera que outras; *descontinuidade de aparência* visa detectar objetos salientes ou regiões que são visualmente diferentes; e *formas familiares*, onde, regiões que possuam uma forma esperada (como por exemplo: pessoas ou veículos) são selecionadas como *foreground*.

As técnicas existentes para subtração de fundo em câmeras móvel podem ser agrupadas em quatro grupos:

- *Compensação do movimento*: são baseadas em técnicas que estimam o movimento utilizando transformação geométrica entre dois quadros consecutivos (MURRAY; BASU, 1994; ARAKI *et al.*, 2000) ou entre cada quadro e um mapa estendido do fundo (ROWE; BLAKE, 1996; HAYMAN; EKLUNDH, 2003; JIN *et al.*, 2008). O mapa estendido do fundo é a representação do fundo na cena e pode ser gerado por algum modelo estatístico. Um método tipicamente utilizado para compensar a movimentação é a Homografia (HARTLEY; ZISSERMAN, 2004) que é uma transformação projetiva planar ao qual é possível mapear pontos de um plano para outro plano, e neste uso, de uma posição anterior da câmera para a atual. A Homografia entre duas imagens é estimada com base em quatro pontos correlacionados nas imagens. Para se obter pontos correlacionados se utiliza, tipicamente, RANSAC (FISCHLER; BOLLES, 1981) ou uma de suas variantes, como por exemplo, MILESAC (TORR; ZISSERMAN, 2000).
- *Segmentação do movimento*: refere-se a problemas de segmentação que se baseiam no movimento. A segmentação é efetuada utilizando a dispersão do movimento, através de mudança da informação nos *pixels* (como por exemplo, de cor ou textura) ou utilizando as duas estratégias anteriores (híbrido). As soluções pertencentes a este grupo podem ser classificadas em quatro categorias (ELGAMMAL; ELQURSH, 2014): estatísticas (GRUBER; WEISS, 2004), em que, a partir de algumas trajetórias escolhidas randomicamente em um subespaço, buscam-se regiões com grande número de trajetórias similares; baseadas em fatoração da matriz de trajetórias (ICHIMURA, 1999; KANATANI, 2001); algébricas que utiliza um algoritmo genérico para separação do subespaço (VIDAL; HARTLEY, 2004); e agrupamento espectral (BROX; MALIK, 2010; OCHS; BROX, 2012; YAN; POLLEFEYS, 2006), ao qual, utiliza a informação no entorno das trajetórias para computar uma matriz de similaridade.
- *Segmentação do movimento em camadas*: são técnicas que modelam a cena como um conjunto de camadas em movimento. Dentre as técnicas adotadas, pode-se citar o fluxo ótico (JEPSON; BLACK, 1993; WANG; ADELSON, 1994) e a utilização de camadas baseadas nos planos 3D (TORR; SZELISKI; ANANDAN, 2001) presentes na cena.

- *Segmentação de movimento com preservação do fundo*: utilizam a segmentação do movimento para separar os objetos que se movem independentemente, e assim, mantendo as representações do fundo e *foreground* da cena. A representação do *foreground* e fundo podem ser gerados de maneiras distintas, mas todos os métodos deste grupo baseiam-se na movimentação da cena. Por exemplo, pode-se utilizar níveis de dispersão da movimentação para gerar as fronteiras (SHEIKH; JAVED; KANADE, 2009). Outros métodos empregam filtros Bayesianos, onde, a representação anterior do fundo e *foreground* são confrontadas com a movimentação no quadro atual, obtendo-se assim, a nova representação para o fundo e *foreground*, podendo efetuar buscas na cena a nível de bloco (KWAK *et al.*, 2011) ou *pixel* (ELQURSH; ELGAMMAL, 2012).

3.2.3 Câmera Quase-estática

Nas operações de vigilância e monitoramento em geral, os fundos, por concepção, deveriam se comportar de forma estática, mas portam-se de maneira dinâmica em alguns ambientes, resultando na necessidade de tratativas especiais. Uma visão dos possíveis problemas de dinamismo do fundo pode ser observada em câmeras instaladas em regiões de costa, por exemplo. Ao se processar e buscar subtrair o fundo proveniente de imagens de uma câmera fixa instalada nestas regiões observa-se os seguintes desafios provenientes deste ambiente (BLOISI, 2014): mudanças de iluminação (gradual ou súbita), sombras, instabilidade ou vibração (proveniente principalmente do vento), movimento de elementos do fundo (por exemplo, folha de árvores balançando ou o agitar das águas do mar) e mudanças na geometria do fundo (por exemplo, carros estacionados ou movimentações de outros objetos). O que nos leva a denominar esse tipo de situação de “quase-estática” já que a câmera é estática (fixa) mas as dificuldades do ambiente necessitam de algum tratamento especial.

Todos os algoritmos criados para trabalhar com câmeras móveis (Seção 3.2.2) podem ser utilizados em câmeras “quase-estáticas”, já que são genéricos o suficiente para operarem sob qualquer tipo de fundo. Porém, essa decisão ocasiona algumas perdas na acurácia da segmentação, muito por conta de uma maior susceptibilidade a falsas detecções de *foreground* e, muitas vezes, o abandono de uma análise temporal (entre quadros sucessivos).

A utilização da criação e manutenção de uma imagem de fundo (B) e a geração da máscara de *foreground* (F), onde, a partir da diferença entre B e o quadro atual obtém-se F (Seção 3.2.1) pode ocasionar ruídos na máscara F em situações de “quase-estático”. Ainda não existem técnicas do estado-da-arte capazes de lidar de maneira efetiva e em tempo real com câmeras estáticas e fundos dinâmicos de qualquer grau (BLOISI, 2014). Pode-se, no entanto, adotar algumas ideias que se adaptem a um problema específico e conduzem a robustez necessária para o funcionamento a contento da solução específica. Um ponto de partida é a adoção de um modelo de fundo adequado. Segundo Cristani *et al.* (2010) existem quatro classes de métodos para modelagem do fundo que podem ser adotados:

- *Por pixel*: cada *pixel* do fundo é independente. Esta classe possui o menor custo computacional e pode ser resolvido em tempo real. Costuma falhar quando aplicada a imagens com ricas e dinâmicas texturas (DALLEY; MIGDAL; GRIMSON, 2008);
- *Por região*: divide os quadros em blocos e extraí aspectos específicos de cada bloco para obter o *foreground*. Ao utilizar um conjunto de *pixels* para gerar o modelo, ganha-se tolerância a oscilações ou pequenos movimentos. Estes métodos se baseiam, por exemplo, em histogramas e extração de bordas dentro da região;
- *Por quadro*: olha para mudanças globais da cena, ou seja, que ocorre em todo o quadro. Um exemplo de problema que pode ser minimizado com esta ideia é a mudança repentina de iluminação observada em todo o quadro (por exemplo, o sol sendo obstruído por uma nuvem);
- *Multifase*: combina os métodos anteriores em um processo serial, capaz de refinar o resultado final.

Pode-se, segundo Cheung e Kamath (2004), seguir uma ideia recursiva ou não recursiva quanto a análise temporal. Os algoritmos recursivos fazem a atualização do modelo do fundo a cada novo quadro processado, resultando efetivamente em alguma influência dos quadros processado em um passado distante sobre o modelo atual do fundo (ELHABIAN; EL-SAYED; AHMED, 2008). Já técnicas não recursivas mantem os N últimos quadros adquiridos e estima o fundo utilizando análise estatística destes quadros (limita-se a influência do passado).

A acurácia de um sistema de subtração de fundo em fundos dinâmicos é aumentada pela adoção de um módulo supressor de ruídos adaptado ao ambiente de aquisição (BLOISI, 2014). Dependendo do grau das perturbações da câmera, pode-se melhorar satisfatoriamente o resultado final apenas com a adoção de filtros morfológicos (GONZALEZ; WOODS, 2008).

Dentro do escopo deste trabalho espera-se algumas dificuldades em câmeras e sensores 3D embarcados no veículo. Por exemplo, um sensor 3D utilizado na obtenção dos dados do motorista, é fixado em uma posição conveniente dentro do veículo, porém, mesmo fixo está sujeito a vibrações ocasionadas pelo motor do veículo ou mesmo por algum desnível ou pequeno obstáculo no caminho do veículo. O vento também pode causar certa turbulência e perturbação na medição das distâncias internas do veículo, caso as janelas do veículo estejam abertas. Nas janelas, objetos externos visíveis podem vir a acrescentar ruídos ao fundo.

3.2.4 Subtração de Fundo em Câmeras RGB-D

Atualmente é possível, além da obtenção da cena em cores (RGB), receber de um sensor (Seção 3.1.5) a distância medida para cada *pixel* da imagem, ou seja, o chamado RGB-D onde D é a profundidade (*depth*) do *pixel*. Existem na literatura algumas soluções adaptadas e preparadas para se beneficiar desta informação adicional.

Camplani e Salgado (2014) realizaram alguns experimentos para identificar vantagens que a utilização dos dados da profundidade pode acrescentar a subtração do fundo, comparando-se a somente utilizar a cor (RGB). Foi empregado o algoritmo de Mistura de Gaussianas (MoG) (STAUFFER; GRIMSON, 1999) sobre a cor e profundidade, separadamente, para a obtenção das máscaras de *foreground*. De fato, a detecção de *foreground* obtida com MoG em dados de profundidade dos *pixels* não sofre influência de sombras ou camuflagem por cor (o objeto que se movimenta possui cor muito próxima ao fundo) e não é afetado por mudanças de iluminação. Contudo, utilizar os dados de profundidade apresenta seus próprios desafios que são: lidar com a “camuflagem por profundidade”, ou seja, ocorre quando o *foreground* está muito próximo ao fundo; e com os ruídos provenientes do sensor de profundidade. Por essas razões, a maioria das estratégias de subtração de fundo em câmeras RGB-D combinam cor e profundidade (CAMPLANI; SALGADO, 2014).

Um exemplo de estratégia baseada tanto em cor quanto em profundidade é a apresentada por Gordon *et al.* (1999), ao qual, adaptou-se o algoritmo MoG para processar todos os dados combinados. Outro exemplo similar é apresentado por Leens *et al.* (2009) que, usando uma câmera *time-of-flight* (ToF), aplica-se o algoritmo ViBe (BARNICH; DROOGENBROECK, 2011) na cor e profundidades adquiridas, os resultados são combinados e processados por operações morfológicas para se obter a detecção de *foreground* final. O sistema de vigilância de Clapés, Reyes e Escalera (2013) é baseado no sensor Microsoft Kinect e utiliza uma distribuição Gaussiana multivariada (utilizando as 4 dimensões do RGB-D) para obter o fundo, e então, segmentar os objetos e pessoas presentes na cena.

A fusão de dados da cor e da profundidade pode ser ampliada. A abordagem apresentada por Camplani e Salgado (2014), emprega classificadores estatísticos de fundo baseados em cor e profundidade, separadamente, e aliados a detecção de bordas, que, por sua vez, utiliza a combinação de cor e profundidade da cena. O *foreground* é obtido por meio de uma média ponderada adaptativa das detecções de fundo (por *pixel*), onde, os dados provenientes da cor são mais influentes (recebe maior peso) que a profundidade em *pixels* situados na borda, e ocorrendo o inverso em *pixels* sem borda detectada. Assim, é possível reduzir o problema de ruído na mensuração da profundidade.

Outras propostas combinam dados de profundidade e infravermelho. Um exemplo deste tipo de abordagem é apresentado por Störmer, Hofmann e Rigoll (2010), onde, as regiões de *foreground* são estimadas de maneiras independentes e o *foreground* resultante é região onde as duas estimações concordam.

Existem, ainda, na literatura soluções que utilizam apenas dados de profundidade para segmentar *foreground* na cena. O sistema de reconhecimento de gestos de Mahbub *et al.* (2013) extrai o *foreground* utilizando algoritmo de segmentação Otsu (OTSU, 1979) sobre o mapa de profundidade da cena, porém, é eficiente apenas em ambientes controlados (fundo constante). A proposta de Stone e Skubic (2011) utiliza um período de treinamento, em que as profundi-

dades mínimas e máximas capturadas e arquivadas em cada *pixel*, e assim, é possível extrair o *foreground* dos quadros subsequentes.

Nesta tese, o uso da profundidade da cena é empregado no monitoramento do motorista (ver Seção 4.1.1). Por essa razão, dá-se atenção aos dados da cena pertinentes ao motorista, sendo assim, pode-se descartar elementos fora deste foco de atenção e interesse. Além disto, o processo de esqueletização (Seção 3.3) é de grande importância no contexto desta tese, pois o elemento alvo e central do processo de segmentação é uma figura humana (motorista) e a sua postura. Por meio da esqueletização, portanto, pode-se obter as posições de partes relevantes do corpo do motorista, descartando todos os dados de profundidade restantes da cena.

3.3 Esqueletização do Corpo Humano

O rastreamento humano em cenas vem ganhando mais atenção e robustez nos últimos anos. Assim, sua aplicabilidade vem se expandido e já é utilizado em jogos de computador, vigilância, telepresença, soluções para reabilitação de pessoas, enfim, vem permitindo e facilitando a interação homem/máquina. Como é possível notar em pesquisas bibliográficas sobre o tema, no início dos anos 2000, o rastreamento de partes do corpo humano utilizava basicamente uma câmera de vídeo, como pode ser visto em trabalhos levantados por Moeslund, Hilton e Krüger (2006) e Poppe (2007). Já depois da popularização de Sensores 3D (Seção 3.1.5), houve uma migração do estado da arte da esqueletização para o emprego de mapas de profundidade, um exemplo desse tipo de técnica é o sistema apresentado por Shotton *et al.* (2013) que permite detectar e oferece dados do esqueleto humano em tempo real. Inúmeras outras propostas vêm sendo apresentadas com base na profundidade da cena e Lun e Zhao (2015) nos apresentam algumas.

Segundo Lun e Zhao (2015), as etapas típicas para a esqueletização humana utilizando dados provenientes de sensores 3D são: Extração de *Foreground*, Detecção de Pose e a Estimação do Esqueleto. A primeira etapa do processamento é a Extração de *Foreground* ao qual são utilizados algoritmos de subtração de fundo (Seção 3.2) e somente os *foregrounds* detectados como humanos são considerados nas etapas subsequentes. A segunda etapa é a Estimação da Pose, onde, comumente se faz uso de técnicas de aprendizado de máquina treinadas na detecção das partes do corpo humano. Sendo conhecida a localização das partes do corpo, pode-se então inferir os pontos de articulações ou *joints* que, ao serem unidos por linhas, obtém-se a esqueletização humana. O número de pontos de articulações pode variar de conforme a aplicação ou técnica adotada. Na Figura 10, mostra-se um exemplo de estimação de esqueleto para um quadro.

Figura 10 – Etapas típicas da esqueletização.

Fonte: Elaborada pelo autor.

3.4 Reconhecimento de Ações Humanas

O ser humano possui a habilidade de perceber ações efetuadas por outras pessoas somente pela informação visual, como por exemplo, identificar pessoas falando ao celular. Esta habilidade está intimamente ligada à capacidade de localização de pessoas e objetos, percebendo partes do corpo que estão em movimento e, por fim, analisando a interação entre humanos e objetos na cena para buscar compreender o que as pessoas estão fazendo.

Nos últimos anos, vem sendo estudados meios de se dotar sistemas inteligentes, da capacidade de reconhecer ações humanas automaticamente. Um sistema de reconhecimento de ações não é responsável somente por adquirir informações da aparência e movimentação de vídeos, mas requer sofisticados modelos de aprendizado de máquina, para então, compreender o significado semântico da grande quantidade de informação extraída da cena (WANG; LIU; WU, 2014b).

Com progresso das tecnologias dos sensores 3D (Seção 3.1.5), vêm se desenvolvendo técnicas que adotam os mapas de profundidade na obtenção da informação visual ao invés do convencional RGB, permitindo inclusive, a operação com baixo nível de iluminação e invariância à textura e a cor da cena.

Os sistemas de reconhecimentos de atividades se dividem em duas categorias (LUN; ZHAO, 2015): os baseados em modelos (a classificação da atividade ocorre por comparação com modelos previamente catalogados) e baseados em algoritmos (utilizam um conjunto de

regras manualmente definidas). Segundo Wang, Liu e Wu (2014b), as duas principais fontes de características utilizadas pelo estado da arte são: a esqueletização de pessoas (Seção 3.3) e o próprio mapa de profundidade.

O rastreo de partes do corpo e, por consequência, sua movimentação é uma fonte importante de informação para distinguir muitas ações do cotidiano humano. Empregando a esqueletização, Yang e Tian (2012) computam a diferença de posições entre todas as articulações rastreadas no quadro atual, considerando também entre o atual e o anterior e entre o atual e o inicial, capturando assim a configuração temporal da pose e, ao empregar PCA, se obtém a informação mais importante da pose para a classificação da ação. SMIJ (*Sequence of most informative joints*) (OFLI et al., 2014) é um método que encontra as seis movimentações de articulações com maior relevância para a classificação de atividades, empregando para isso, a variância do ângulo e a velocidade angular. Existem, ainda, métodos que são supervisionados como o conjunto de *actionlet* de Wang, Liu e Wu (2014c) ou representação de interações por subconjunto de articulações. Este método utiliza modelos sob supervisão humana para discriminar ações, como por exemplo, o ato de levar o celular ao ouvido necessita de uma movimentação preponderantemente nas articulações de um dos braços.

Muitas ações não podem ser detectadas apenas com base na esqueletização, pois o ato de “comer” ou “beber”, por exemplo, possuem características bem semelhantes em termos de movimentação da pessoa. A ideia, então, é extrair características do mapa de profundidade que possam enriquecer a compreensão da cena. Um dos meios mais comuns na literatura é adaptar, ou mesmo, aplicar técnicas para imagens níveis de cinza na extração de informações da profundidade da cena. Dentre as técnicas 2D empregadas, destacam-se: HOG (*histograms of oriented gradients*) (DALAL; TRIGGS, 2005), SIFT (*scale-invariant feature transform*) (LOWE, 2004), STIP (*space-time interest points*) (LAPTEV et al., 2008), HOF (*histograms of optic flow*) (LAPTEV et al., 2008) e descritores de *kernel* (BO et al., 2011). A proposta de Yang e Tian (2012) aplicou HOG sobre mapas de diferença de profundidades (DMM-HOG), que são computados utilizando dois quadros consecutivos e agregando a diferença com o passar do tempo, sendo diferenciadas atividades presentes em vídeos com aplicação de DMM-HOG em três pontos de vistas distintos (frontal, lateral e superior). Um padrão local de profundidade (LDP) (ZHAO et al., 2012) pode ser extraído de uma região, ao qual é subdividida em células, onde, calcula-se o valor médio da profundidade e, então, o vetor das diferenças (LDP) é gerado computando-se todas as diferenças entre pares de células. Pode-se, ainda, converter o mapa de profundidade dos quadros em uma nuvem de pontos (3D), que por sua vez pode ser expandida para 4D ao se analisar uma sequência de quadros para determinar a ação (análise temporal). A proposta de Wang, Liu e Wu (2014d) extrai amostras randômicas de subvolumes 4D com diferentes tamanhos e localizações, onde, em cada subvolume é obtida a taxa de ocupação normalizada via função sigmoide. Assim, pode-se utilizar um classificador previamente treinado que utilize a ocupação nessas amostras para determinar a atividade.

Apesar dos avanços do estado da arte de reconhecimento de atividades ocorridos nos últimos anos, Wang, Liu e Wu (2014a) elencam alguns desafios ainda não vencidos, como: tratamento de oclusões; reconhecimento das atividades em tempo real; e aplicação em semântica complexa, ou seja, em ambientes com diversos tipos de objetos e interações (humano/objeto) possíveis.

3.5 Reconhecimento de Padrões

O Reconhecimento de Padrões (RP) preocupa-se com a detecção automática de regularidades (padrão) nos dados por meio da utilização de algoritmos de computador que possibilitam a classificação dos dados em classes ou categorias distintas (BISHOP; NASRABADI, 2006; THEODORIDIS; KOUTROUMBAS, 2009). RP pode ser integrado a sistemas de Visão Computacional dotando-os da capacidade de classificar (reconhecer) objetos ou padrões capturados.

Nas subseções a seguir, ferramentas de suporte a classificação em RP são apresentadas.

3.5.1 Máquina de Vetor de Suporte (SVM)

A SVM (Máquina de Vetor de Suporte) foi introduzida por Vapnik em 1995 e é uma ferramenta de classificação binária (VAPNIK, 1995). Considerando um conjunto de dados $\{(\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n)\}$ de entrada $x_i \in R^d$ (sendo d o espaço dimensional) e saída $y_i \in \{-1, +1\}$. A ideia central da técnica é gerar um hiperplano ótimo para separar duas classes de objetos. O hiperplano é escolhido para maximizar a separação entre as duas classes, gerada com base em vetores de suporte (WANG, 2005). A fase de treinamento consiste na escolha dos vetores de suporte entre os dados de treinamento previamente rotulados. Na Figura 11, tem-se uma visão geral do funcionamento do SVM.

Figura 11 – Esboço do SVM.

Fonte: Adaptada de Meyer (2012).

Com a SVM é possível utilizar algumas funções de *kernels* para tratar dados não lineares (não possuem margem de separação linear entre classes). A função de *kernel* transforma os dados

originais em um espaço de características de alta dimensionalidade, onde as separações não lineares podem se apresentar sob forma linear (STANIMIROVA *et al.*, 2010). Dentre os *kernels* existentes pode-se citar: Linear (Equação 3.5), Polinomial (Equação 3.6), *Radial basis function* (Equação 3.7) e Sigmoide (Equação 3.8). A escolha de uma função adequada (mais acurada) e de parâmetros (γ , *degree* e *coef0*) corretos são importantes passos para atingir a alta acurácia do sistema de classificação.

$$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i \cdot \vec{x}_j \quad (3.5)$$

$$K(\vec{x}_i, \vec{x}_j) = (\gamma(\vec{x}_i \cdot \vec{x}_j) + \text{coef0})^{\text{degree}}, \gamma > 0 \quad (3.6)$$

$$K(\vec{x}_i, \vec{x}_j) = e^{-\gamma\|\vec{x}_i + \vec{x}_j\|^2}, \gamma > 0 \quad (3.7)$$

$$K(\vec{x}_i, \vec{x}_j) = \tanh(\gamma(\vec{x}_i \cdot \vec{x}_j) + \text{coef0}) \quad (3.8)$$

Após a escolha da função de *kernel* adequada para o classificador, pode-se iniciar o seu treinamento. Na Equação 3.9, precisa-se maximizar os valores do vetor $\vec{\omega}$. Este é um problema em otimização de programação quadrática (HEARST *et al.*, 1998) e sujeito as restrições (para qualquer $i = 1, \dots, n$, sendo n a quantidade de dados de treinamento): $0 \leq \omega_i \leq C$ e $\sum_{i=1}^n \omega_i y_i = 0$. O parâmetro de penalização C é um controle da relação entre a complexidade do algoritmo e o número de amostras do treinamento classificados erroneamente.

$$W(\vec{\omega}) = \sum_{i=1}^n \omega_i - \frac{1}{2} \sum_{i,j=1}^n \omega_i \omega_j y_i y_j K(\vec{x}_i, \vec{x}_j) \quad (3.9)$$

O cálculo do limiar b se dá por meio da Equação 3.10. Deve-se calcular para todos os vetores de suporte $\vec{x}_j$ ($0 \leq \omega_j \leq C$) e adota-se a média de todos os cálculos como valor de b .

$$b = y_j - \sum_{i=1}^l y_i \omega_i K(\vec{x}_i, \vec{x}_j) \quad (3.10)$$

Para executar a classificação de um novo vetor de característica $\vec{x}$ utiliza-se a Equação 3.11, sendo $\lambda_i = y_i \omega_i$ e a função matemática *sign* responsável pela extração do sinal de um número real, apresentando -1 para valores negativos, 0 para valor igual a zero e $+1$ para valores positivos. Na Figura 12, mostra-se, de maneira geral, a arquitetura da SVM.

$$f(\vec{x}) = \text{sign}\left(\sum_i \lambda_i K(\vec{x}, \vec{x}_i) + b\right) \quad (3.11)$$

Figura 12 – Arquitetura da SVM. A função de kernel K determina o tipo de classificador (ex: Polinomial). Todos os outros parâmetros são encontrados durante o treinamento.

Fonte: [Schölkopf \(1997\)](#).

3.5.2 Multilayer Perceptron (MLP)

Redes Neurais Artificiais (RNAs) são sistemas computacionais cuja estrutura é inspirada por estudos do cérebro humano. Em 1958, Rosenblatt inventou o Perceptron que é uma rede neural de alimentação simples ([MEHROTRA; MOHAN; RANKA, 1996](#)). Com um Perceptron pode-se resolver problemas linearmente separáveis. Perceptrons com mais de uma camada de ligações variavelmente ponderadas são referidos como Perceptrons multicamadas (MLP) ([JAIN; MAO; MOHIUDDIN, 1996](#)). A rede MLP pode resolver problemas não lineares e é uma ferramenta de classificação multi-classes.

Existem três tipos de camadas presentes em uma rede MLP ([JAIN; MAO; MOHIUDDIN, 1996](#)). A primeira camada é responsável por receber os dados, não realizando nenhum tipo de processamento e seu número de neurônios é igual a quantidade de variáveis (características). A última camada é denominada camada de saída, pois é a camada responsável por retornar a resposta calculada pela rede. A quantidade de neurônios desta camada depende da quantidade de *bits* de saída da rede. Entre estas pode haver uma ou mais camadas intermediárias (ocultas), o número depende da complexidade do problema. Na Figura 13, é mostrado um exemplo de uma rede MLP com 4 neurônios na camada de entrada, 3 neurônios na camada intermediária e 2 neurônios na camada de saída.

Individualmente, um neurônio tem ligações de entrada (que recebem da camada anterior) e ligações de saída (passa a resposta para a camada seguinte). Cada valor de entrada do neurônio é multiplicado pelo seu respectivo peso e os resultados são somados e adicionados ao *bias* (peso do neurônio) ([MEHROTRA; MOHAN; RANKA, 1996](#)). A soma resultante é transformada utilizando uma função de ativação ($f(x)$). Nas Equações 3.12, 3.13 e 3.14 são mostradas, respectivamente, as funções de ativação: Identidade, Sigmoid Simétrica e Gaussiana. A camada de entrada utiliza a função de ativação Identidade, pois esta camada possui função apenas de propagação dos valores de entrada (características) para a camada seguinte. A escolha da função de ativação correta e dos parâmetros (α e β) ajustados ao problema, nas demais camadas, podem

Figura 13 – Um diagrama esquemático de uma rede MLP de três camadas.

Fonte: Elaborada pelo autor.

influenciar e contribuir para melhorar a acurácia do classificador.

$$f(x) = x \quad (3.12)$$

$$f(x) = \frac{\beta(1 - e^{-\alpha x})}{(1 + e^{-\alpha x})} \quad (3.13)$$

$$f(x) = \beta e^{-\alpha x^2} \quad (3.14)$$

Uma rede MLP precisa ser treinada, através do ajuste dos pesos da rede, para estar apta a classificar. Um dos algoritmos de treinamento de redes MLP possíveis é o denominado *backpropagation* (WERBOS, 1994). Este processo visa descobrir os pesos que melhoram a acurácia da rede e se dá pela redução gradual do erro da rede. Os erros são estimados utilizando-se os valores de treinamentos já previamente rotulados, em um processo denominado de aprendizado supervisionado. Existem outros algoritmos de aprendizado neural (de ajuste dos pesos), de melhor desempenho e derivados do *backpropagation*, como por exemplo, o algoritmo de treinamento denominado de *Rprop - Resilient Propagation* (RIEDMILLER; BRAUN, 1992).

3.6 Considerações

Neste capítulo, foram apresentados conceitos, técnicas e sensores que podem ser empregados na identificação de distrações e perturbações do motorista na condução de veículos. Porém, sensores passivos como as câmeras de vídeo têm limitações que podem ser impostas por problemas de iluminação e de variação da iluminação do ambiente, sendo que as câmeras monoculares fornecem apenas uma representação plana (projeção 2D) do espaço observado.

Imagens RGB-D podem ser adquiridas de sensores 3D, assim, pode-se obter a profundidade do ambiente interno do veículo, como por exemplo, de uma cabine de caminhão. A representação 3D possui vantagens sobre a representação 2D, por permitir uma descrição mais exata da cena analisada. Além disto, alguns dispositivos de captura da profundidade da cena funcionam adequadamente mesmo sob condições de pouca ou nenhuma iluminação (dispositivos ativos de percepção).

Algoritmos de subtração de fundo podem ser aplicados em imagens RGB assim como em dados de profundidade da cena (*depth*), porém em certas situações faz-se necessária a pesquisa e o desenvolvimento de algoritmos melhor adaptados ao problema tratado, como por exemplo, na cabine de um veículo onde a câmera apesar de ser fixa pode sofrer com vibrações e mesmo o motorista estando estático pode também ter um certo “balanço” devido ao movimento do veículo, sendo portanto um tema de pesquisa em aberto. Com a subtração de fundo pode-se obter a região onde o motorista se encontra e, assim, rastrear as partes de seu corpo (esqueletização). Utilizando a esqueletização do motorista e o mapa de profundidade é possível reconhecer se o motorista está desempenhando alguma atividade prejudicial a direção do veículo, como por exemplo, utilizar o celular. Assim, pode-se monitorar os movimentos do motorista.

Outros sensores como acelerômetro, giroscópio podem ser úteis para mensurar a qualidade da direção efetuada pelo motorista, uma vez que suas ações são também um reflexo da situação e do ambiente externo em que ele se encontra.

Com o mapa de profundidade, a esqueletização, o reconhecimento de atividades paralelas efetuadas pelo motorista e informações sobre a qualidade da direção empregada no veículo o objetivo é de extrair características que permitam o monitoramento do comportamento do motorista. Espera-se que a importante tarefa de monitorar e reconhecer padrões de comportamento permita identificar se o motorista se encontra em situação de risco pelo uso do celular, sonolência ou embriaguez seja executada por reconhecimento de padrões, utilizando para tanto, o conjunto de características extraído a partir de dispositivos como os sensores RGB-D e de movimentação do veículo (IMU).

DESENVOLVIMENTO DA SOLUÇÃO

Este trabalho propõe uma abordagem para tratar sinais obtidos de sensores e de outros sistemas de apoio à direção disponíveis em um veículo, permitindo assim, a análise comportamental do seu condutor para a identificação de momentos de inaptidão do motorista na condução do veículo. O uso de celular no trânsito, a embriaguez e a sonolência são os principais focos de detecção e de análise comportamental. A análise comportamental inclui dados com uma captura em 3D dos movimentos e ações do motorista, bem como, dados de monitoramento do comportamento do próprio veículo, que refletem o modo de conduzir do motorista. Este trabalho propõe duas arquiteturas para sistemas ADAS para identificação de distrações e perturbações do motorista na condução de veículos. Os dois sistemas propostos são não intrusivo e devem estar operantes durante toda a condução do veículo.

Nas subseções seguintes, os sistemas propostos são descritos nas Seções 4.2 e 4.3, sendo as possíveis características para os sistemas ADAS candidatos, mostradas na Seção 4.1.

4.1 Características

O sistema possui duas fontes de informação: o monitoramento do motorista e a qualidade da direção. O monitoramento do motorista tem características relevantes da postura do motorista, sendo empregadas características baseadas no deslocamento 3D das mãos do motorista. A qualidade da direção apresenta características capazes de mensurar a direção efetuada pelo motorista. Os dois grupos de fonte de características permitem determinar o quão “normal” é a direção ou postura do motorista, em outras palavras, as características permitem observar “boas práticas” pelo motorista, tanto de sua postura (boas poses) como do seu modo de conduzir o veículo. O estudo realizado e a abordagem proposta permitiram definir quais são as características relevantes, onde o modelo de distinção entre uma direção com e sem “risco” é obtido experimentalmente (veja Seção 6) usando aprendizado de máquina e um conjunto de dados (Seção 5.2).

Nas duas próximas subseções, as características candidatas para o Monitoramento do Motorista e Qualidade de Direção são mostradas.

4.1.1 Monitoramento do Motorista

O monitoramento do motorista utiliza o Kinect v2¹ (Seção 3.1.5) para captura de informações do motorista. Com este sensor, pode-se obter o mapa de profundidade entre a distância de 0,5 e 4 metros do sensor. O Kinect é posicionado frontalmente ao motorista e na parte superior da cabine do veículo, permitindo assim, uma melhor precisão na obtenção das distâncias já que o motorista tenderá a estar na região com maior precisão na aquisição de dados pelo sensor. Esta posição não prejudica a dirigibilidade do veículo e evita oclusões da detecção de partes do corpo do motorista.

O rastreamento do motorista está baseado na esqueletização (Seção 3.3) do motorista. A esqueletização possibilita o rastreio de partes do corpo do motorista (postura). O método de esqueletização de Shotton *et al.* (2013)² foi empregado neste trabalho. Este método prediz a pose humana e a respectiva posição 3D de cada parte do corpo rastreada, utilizando para isso, um único mapa de profundidade adquirido do Kinect. Portanto, a esqueletização é obtida somente utilizando dados de um *frame* de informação do Kinect. O processo de obtenção das partes do corpo do motorista (por exemplo: braços, mãos, cabeça, pescoço, ombros e tronco) inicia-se com técnicas eficientes de aprendizado de máquina, que classificam cada *pixel* do mapa de profundidade segundo a parte do corpo correspondente. Embora a classificação das partes do motorista ocorra *pixel a pixel*, na prática o treinamento do classificador ocorreu de tal forma que regiões (grupo de *pixels* próximos) do mapa de disparidade são classificadas como sendo a mesma parte e buscando respeitar uma pose factível a um humano. Finalmente, a posição 3D de cada uma das partes do corpo é calculada utilizando o resultado da classificação. Este método identifica 25 partes do corpo e suas posições espaciais. Apesar da esqueletização ocorrer utilizando um único *frame* de informação, no rastreamento das partes (esqueleto) emprega-se um conjunto de *frames* consecutivos para observar uma pessoa presente, evita-se assim, falsas detecções.

É possível que o sistema de monitoramento perceba mais de uma pessoa (esqueleto) em um mesmo *frame*. Nestes casos, o esqueleto com a cabeça mais próxima ao Kinect é adotado como o do motorista, pois espera-se que o motorista seja a pessoa mais próxima ao Kinect, dada a sua posição frontal.

São cinco as partes do corpo do motorista utilizadas pelo extrator de característica do monitoramento do motorista proposto, sendo elas: a cabeça, ombro esquerdo, ombro direito, mão esquerda e mão direita. As posições 3D das mãos esquerda e/ou direita são base para a

¹ Mais informações sobre o Kinect V2 estão disponíveis em <<https://developer.microsoft.com/pt-br/windows/kinect>>.

² Kinect for Windows SDK 2.0 está disponível em <https://go.gl/YcC2GD>.

obtenção de todas as características de monitoramento do motorista. Duas origens referenciais são criadas. A primeira origem é a do volante do veículo, e sua localização é estabelecida em uma calibração inicial do sistema. A segunda origem está na cabeça do motorista, sua posição é a mesma obtida da esqueletização do motorista e sendo utilizada como orientação a orientação do volante rotacionada no eixo X em 180 graus. Na Figura 14, mostram-se dois exemplos de nuvens de pontos contendo suas respectivas origens, do volante e da cabeça do motorista.

Figura 14 – Exemplo do posicionamento das origens do Volante e da Cabeça do motorista em uma nuvem de pontos, sendo frontal (a) e lateral (b), onde, o eixo X é o vermelho, Y verde e Z azul.

Fonte: Elaborada pelo autor.

São 20 as características candidatas extraídas do monitoramento do motorista em cada esqueletização (*frame*), sendo 10 características extraídas para cada origem. A Tabela 3 mostra todas as características candidatas e suas abreviações. Para calcular as características, as posições 3D de cada mão do motorista são convertidas para o sistema de coordenadas do Volante e da Cabeça do Motorista (*VOL* e *CM* são referências as origens, respectivamente). Então, para cada origem, *VOL* e *CM*, são obtidas as distâncias mínima (*minDistMaos*) e máxima (*maxDistMaos*) das duas mãos, as distâncias (*distME* e *distMD*) e suas componentes (*xLH*, *yLH*, *zLH* e *xRH*, *yRH*, *zRH*).

Tabela 3 – Todas as características candidatas para o monitoramento do motorista

Características	Abreviações
Para as duas mãos: distância mínima, distância máxima.	<i>minDistMaos</i> , <i>maxDistMaos</i> .
Mão esquerda: distância, x, y, z.	<i>distME</i> , <i>xME</i> , <i>yME</i> , <i>zME</i> .
Mão direita: distância, x, y, z.	<i>distMD</i> , <i>xMD</i> , <i>yMD</i> , <i>zMD</i> .
Para (origens): Volante (<i>VOL</i>) e Cabeça do Motorista (<i>CM</i>)	
Total: 10 características para cada origem (<i>VOL</i> e <i>CM</i>), total de 20 características	

Todas as características são normalizadas entre -1 e $+1$ como mostrado pela Equação 4.1, onde, $feature_v$ é o valor calculado da característica, $shoulder_l$ é o comprimento dos ombros do motorista (distância 3D entre o ombro esquerdo e direito na esqueletização), um fator φ e i define cada característica. O fator φ é definido experimentalmente usando um conjunto de dados e descreve o quão representativo é o valor da característica em relação ao corpo do motorista.

$$feature_n(i) = feature_v(i) * shoulder_l * \varphi(i) \quad (4.1)$$

4.1.2 Qualidade de Direção

As principais fontes de dados para a mensuração da qualidade da direção são: o rastreamento da posição do veículo na pista de rodagem (LEE *et al.*, 2006; BERGASA *et al.*, 2014) e sensores de movimentação e orientação (JOHNSON; TRIVEDI, 2011; BERGASA *et al.*, 2014; PINILLA; QUINTERO; PREMACHANDRA, 2014). Portanto, para obtenção das características, utilizam-se sensores inerciais acoplados ao veículo (acelerômetro e giroscópio) e um sistema de detecção da rodovia e faixa de rodagem, ainda são utilizados dados provenientes diretamente da telemetria do veículo. Esses dados estão usualmente disponíveis em um veículo inteligente e/ou autônomo. Na Tabela 4, todas as características obtidas são mostradas.

Tabela 4 – Todas as características candidatas na qualidade de direção

From	Características	Abreviações
Telemetria do Veículo	Aceleração	<i>acel</i>
	Velocidade	<i>velocidade</i>
	Nível do pedal do acelerador	<i>PAlevel</i>
	Nível do pedal de freio	<i>PFlevel</i>
	Ângulo do volante	<i>VOLangulo</i>
Sensores Inerciais	Acelerômetro X	<i>aX</i>
	Acelerômetro Y	<i>aY</i>
	Acelerômetro Z	<i>aZ</i>
	Giroscópio X	<i>gX</i>
	Giroscópio Y	<i>gY</i>
	Giroscópio Z	<i>gZ</i>
Pista e Rodovia	Distância para o Centro da Faixa	<i>distCF</i>
	Ângulo da faixa em relação ao veículo	<i>fAngulo</i>
	Rodovia reta	<i>rReta</i>

O sistema de detecção de faixa e rodovia é considera um sistema capaz de detectar a rodovia, suas faixas, e a posição do veículo em relação a faixa, como por exemplo, os sistemas propostos por Paula e Jung (2013), Hillel *et al.* (2014), Kaur e Kumar (2015), Kumar e Simon (2015) e Mendes, Frémont e Wolf (2016). A distância para o centro da faixa é calculada usando a posição atual do veículo (ponto central do veículo) em relação ao centro da faixa atual. A orientação do veículo na faixa atual é usada para calcula o ângulo da faixa em relação ao veículo.

Em termos de sensores, o GPS é um sensor empregado na estimativa de posição (rota) e para detecção da velocidade de deslocamento do veículo em alguns sistemas apresentados na literatura (DAI *et al.*, 2010; HAILE, 1992; SHIRAZI; RAD, 2012), porém, sua utilização foi evitada pelo fato do GPS possuir dificuldade de operação em ambientes fechados como em túneis, ou mesmo, ser dependente das condições do tempo. A medição da velocidade, aceleração e frenagem, é obtida diretamente da telemetria do veículo. O acelerômetro e giroscópio são sensores que podem auxiliar na detecção de guinadas ou ziguezague do veículo.

A qualidade da direção também é um bom indicativo para distinguir um motorista embriagado, sendo empregada nos sistemas propostos por Dai *et al.* (2010), Haile (1992) e (SHIRAZI; RAD, 2012). A pista de rodagem pode funcionar, portanto, como um circuito onde o motorista precisa passar e de modo satisfatório para ser considerado apto a dirigir.

É importante, ainda, detectar o contexto do movimento em que o veículo se encontra. O *rReta* é booleano e é a característica responsável pelo contexto da direção, indicando se o atual segmento (região) da rodovia é reto (1) ou curvado (0). Dependendo do contexto pode-se esperar comportamentos diferenciados do motorista.

4.2 Sistema de Curto e Longo Prazos para Detecção de Riscos

O fluxo do Sistema de Curto e Longo Prazos (SCLP) para detecção de riscos é mostrado na Figura 15. Neste sistema, duas análises temporais são efetuadas sobre as características, uma de curto prazo (poucos dados) e outra de longo prazo (maior quantidade de informação utilizada) para diferenciar um motorista em condições normais frente a um distraído. O objetivo de se utilizar a análise de curto prazo é o de detectar rapidamente padrões sobre as características que indiquem riscos, permitindo assim, que o motorista seja alertado de sua postura ou direção perigosa o mais rápido possível. Já a análise de longo prazo tem como objetivo ser a ratificadora do risco, sendo capaz de avaliar um período maior de dados, ou seja, mais informação, determinando com maior acurácia o estado atual de perigo da condução do veículo.

O módulo de aquisição obtém novos dados brutos dos sensores e subsistemas do veículo. O módulo de extração de características converte os dados brutos da aquisição nas características normalizadas (ver Seção 4.1). Nas próximas subseções, o restante do SCLP é descrito.

4.2.1 Reconhecimento de Curto Prazo

O subsistema de Curto Prazo é responsável pelo reconhecimento do risco em um curto período de tempo e com menos dados, ou em outras palavras, em *nCP frames*.

No fim do período, a média e desvio padrão - os dados estatísticos - são calculados para cada características (veja na Figura 15) e estes valores são usados como entrada para o

Figura 15 – Fluxo do SCLP.

Fonte: Elaborada pelo autor.

classificador.

O classificador (*classificadorCP*) é previamente treinado com as características escolhidas (média e desvio padrão), assim, pode-se prever dentro de um período de nCP frames situações de “risco” ou “normais” para um motorista.

A última etapa do Reconhecimento de Curto Prazo é a de Avaliação das últimas $npCP$ previsões (períodos anteriores) e, sendo observado pelo menos 80% das últimas previsões como “risco”, a saída do subsistema de Curto Prazo é risco, de modo contrário, a saída do sistema é “normal”. Enquanto o final de um novo período não é alcançado, a última Avaliação (dos últimos $npCP$ períodos preditos) é adotada como saída do subsistema de Curto Prazo. Esta avaliação utilizando as últimas previsões é usada para evitar falsos alertas provenientes de previsões equivocadas do classificador de períodos curtos. Assim o sistema verifica algumas previsões de risco de curto prazo para observar risco na saída do sistema. Isso pode melhorar o conforto do motorista ao dirigir, sem comprometer a precisão e performance do sistema como um todo.

4.2.2 Reconhecimento de Longo Prazo

O subsistema de longo prazo usa uma longa sequência de dados dos últimos nLP frames (janela de dados) para prever risco para um motorista. Quando a janela está cheia (há nLP frames anteriores de dados), os Dados Estatísticos (média e desvio padrão) são calculados e podem ser selecionados como características de entrada para a janela atual de Longo Prazo.

Um classificador previamente treinado (*classificadorLP*) em janelas de nLP frames, prediz os dados estatísticos das características entre “risco” ou “normal”, gerando assim, a saída para o Reconhecimento de Longo Prazo. Nos frames iniciais, quando a janela ainda não está cheia, a saída do sistema utilizada é “normal”, e nenhuma classificação dos dados é efetuada. A

predição de longo prazo é um forte indicativo para o sistema de que o motorista está em risco ou em uma situação normal de condução do veículo.

4.2.3 Controle do Alerta

A saída do sistema para cada *frame* é um nível de alerta para o motorista, contendo três possibilidades de intensidades: “Sem alerta”, “Alerta baixo” e “Alerta alto”.

Nenhum alerta é emitido para o nível “Sem alerta”, sendo este empregado como saída para o sistema quando o Reconhecimento de Curto Prazo prediz uma situação normal (uma boa pose para a direção e estilo de direção considerado normal).

Quando risco é detectado pelo subsistema de Curto Prazo e o subsistema de Longo Prazo detecta uma situação normal, o nível de alerta empregado no sistema é “Alerta baixo” (aviso leve). Este alerta é utilizado como aviso ao motorista de que precisa voltar a ter uma pose e/ou direção que não acarrete risco.

Finalmente, o nível “Alerta alto” (aviso forte) ocorre quando ambos os subsistemas, Curto e Longo prazos, predizem “risco”. Este é o nível mais evidente de inaptidão do motorista na condução de um veículo, pode-se portanto ao observar esta saída, além de um aviso sonoro, levar a tomada de ações que visem assumir o controle do veículo a fim de evitar um acidente mais grave (por exemplo: realizar uma manobra automática de parada do veículo, impedindo que o motorista continue a conduzir o veículo).

4.3 Sistema Bufferizado de Detecção de Riscos

A proposta de Sistema Bufferizado de Detecção de Riscos (SB) é mostrada pela Figura 16. A SB utiliza uma análise temporal das características na detecção da direção perigosa, ou seja, para manter as boas práticas comportamentais do motorista tanto na direção quanto na sua postura.

Figura 16 – Fluxo do SB.

Fonte: Elaborada pelo autor.

Os dados brutos dos sensores e subsistemas do veículo são obtidos pelo módulo de aquisição. As características já normalizadas (ver Seção 4.1) são saídas do módulo de extração de características, obtidas dos dados brutos da aquisição.

O Subsistema de Reconhecimento é responsável por reconhecer “risco” na postura ou estilo de direção do motorista usando os dados extraídos (características) de um período de nf frames. Os módulos de Dados Estatísticos e Classificação são executados apenas no final de cada período, sendo estes responsáveis pela obtenção da média e desvio padrão das características no período e da predição destes dados usando um classificador (*classificadorSB*), respectivamente. *classificadorSB* é previamente treinado em períodos de nf frames e é capaz de prever os dados estatísticos das características em “risco” ou “normal”.

O módulo de *Buffer* para Avaliação mantém as *BS* (*Buffer Size*) últimas predições dos períodos, ou seja, *BS* é o tamanho do *buffer* de predições empregado. Se ao menos 65% (definido empiricamente) das predições dos últimos períodos são “risco”, então, “risco” (situação de perigo) é a saída do SB, caso contrário, o retorno do SB é “normal”. A saída do módulo de *Buffer* para Avaliação é mantida para todos os frames posteriores enquanto o fim de um novo período não é alcançado. Usando a análise temporal das últimas *BS* predições é possível evitar erros de predições de períodos pelo *classificadorSB*, assim, pode-se reduzir o número de falsas predições (falsos positivos/falsos negativos) na saída do SB. O módulo de *Buffer* para Avaliação funciona como um filtro na saída do sistema, provendo predições mais estáveis.

4.4 Considerações

Neste capítulo, foram apresentadas duas arquiteturas de sistemas para detecção de momentos de risco do condutor de veículos por distração pelo uso de celular, sonolência ou embriaguez. Outras distrações e perturbações podem vir a ser detectadas pelos sistemas já que estão baseados na detecção de padrões relacionados as características de monitoramento do motorista (comportamento e postura do motorista) e da qualidade de direção (comportamento observado pelo veículo), ou seja, atitudes distantes do padrão normal de dirigibilidade podem vir a serem detectadas pelo sistema. No que tange ao sistema como um todo, algumas definições empíricas serão apresentadas no Capítulo 6, na experimentação do sistema.

As características que podem ser extraídas de cada *frame* de informação bruta possuem duas fontes, o monitoramento do motorista e o estilo de direção do veículo. O monitoramento do motorista permite observar a postura do motorista, enquanto, o estilo de direção do veículo procura mostrar como o motorista está conduzindo o veículo, permitindo assim, que a análise do motorista seja mais completa. Ambos os sistemas, SCLP e SB, podem fazer uso das características apresentadas para detectar as mais diversas distrações e perturbações que permeiam a área de atuação do motorista no veículo.

O SCLP utiliza duas análises temporais, ou seja, um período de dados pequeno e uma

janela de informações maior. O tamanho do período e janela são estabelecidos por experimentação (Capítulo 6). Nas duas análises as médias e desvios padrões das características são analisadas para a detecção de risco. Essas duas análises permitem ao sistema funcionar como um “educador do motorista”. A análise de curto prazo permite, com dinamismo, saber a real situação de risco da condução instantânea. Quando um risco é detectado no curto prazo o motorista pode ser avisado para que sua postura e direção retorne a padrões aceitáveis, assim evitando, que a análise de longo prazo detecte risco. Quando o risco é detectado no curto e longo prazo um sinal evidente de risco pode ser emitido pelo SCLP, permitindo a atuação de sistemas autônomos dos veículos, de modo que estes possam assumir o controle do mesmo para si ou que possam estacionar o veículo em um local seguro. Este sistema é, portanto, capaz de emitir três níveis de alerta: sem risco, retome as boas práticas de direção e motorista está inapto a dirigir o veículo. Este sistema pode ser conveniente para distrações como uso de celular no trânsito, ao qual, é uma distração que permite a reeducação ou conscientização do motorista pelo uso de sinais sonoros, ou seja, o sistema pode expandir a sua utilidade se tornando um tutor de boas práticas do motorista. A sonolência é uma perturbação que pode também ser avaliada pelo SCLP com ganhos, já que sinais de curto prazo poderiam alertar o motorista de que está se tornando sonolento, podendo fazer com que o mesmo, por conta própria, dirija o veículo para um local onde possa descansar e se recuperar antes de seguir viagem.

O SB utiliza apenas uma análise temporal das características extraídas, o que permite apenas dois níveis de alertas, “risco” e “sem risco”. Portanto, esse tipo de sistema não pode operar na reeducação do motorista. Para algumas distrações, como embriaguez, a reeducação do motorista não é factível, já que não se espera que o motorista possa retomar as suas condições normais fisiológicas em um curto espaço de tempo. Por essa razão, a perturbação pelo uso de álcool ou outras drogas lícitas e ilícitas que prejudicam a condução do veículo podem ser detectadas com o uso da arquitetura SB. SB pode, portanto, apenas ligar um outro sistema autônomo no veículo, ou mesmo, avisar as autoridades do problema de condução do veículo. Sinais sonoros não deveriam ser considerados como o principal objetivo de saída do SB.

A detecção de distrações e perturbações é de difícil experimentação com dados reais, já que não se pode evitar de expor pessoas a riscos envolvendo a direção embriagada, sonolenta, ou mesmo, falando ao celular. Por essas razões, buscou-se utilizar um ambiente simulado na experimentação do sistema. Mais informações sobre o ambiente de simulação são mostradas no Capítulo 5. Espera-se ainda evitar a aquisição de dados de sensores ou sistemas de difícil correspondência entre o ambiente simulado e o real, já que, tornaria o resultado deste trabalho mais distante de sua utilização em um veículo real. Apesar do sistema simular o veículo, a maioria dos dados, como os dados do Kinect, provêm de sensores reais, fornecendo uma análise bastante realista do comportamento do condutor. Os dados do veículo simulado também são bastante realistas. Apesar de serem usados dados simulados, foi considerada a ampla experiência do grupo de pesquisa (LRM-ICMC/USP) na aquisição de dados reais e no desenvolvimento e implementação de aplicações reais, com veículos autônomos e sistemas inteligentes/ADAS para

veículos reais.

SIMULADOR E A COLETA DE DADOS PARA EXPERIMENTOS

A existência e o uso de simuladores nos remete a 1934 quando Miles e Vincent desenvolveram um equipamento que produzia a ilusão de dirigir para atender a necessidade de testes obrigatórios para motoristas, exigidos pelo governo do Reino Unido, como tentativa de diminuir o grande número de mortes e feridos do trânsito da época (FISHER *et al.*, 2011). Simuladores de direção, propriamente ditos, foram desenvolvidos para treinamento de forças militares durante a Segunda Guerra Mundial (BLANA, 1996).

Atualmente, a comunidade científica vem criando e melhorando simuladores de direção por diversas razões, dentre elas pode-se citar, o teste e treinamento de Sistemas Avançados de Assistência ao Condutor (ADAS, ver Seção 2.1) (HASSAN *et al.*, 2013), no teste e melhoramento de sistemas para veículos autônomos (GÓMEZ *et al.*, 2014) e para obtenção da simulação completa da dinâmica veicular (HUANG; CHEN, 2003), enfim, constituindo um ambiente virtual mais próximo ao mundo real (JIA; LU; WANG, 2011). De maneira geral, estes sistemas vem sendo aplicados por pesquisadores em estudos sobre o comportamento do motorista, como por exemplo, efeitos do cansaço e das drogas (JAMSON, 2007), possibilitando assim, o estudo do comportamento do motorista em condições, que de outra maneira, seriam consideradas ilegais ou não éticas (SLOB, 2008). Portanto, a constituição de um simulador de direção veicular é uma ferramenta de pesquisa atraente porque é segura, com ambiente conveniente e controlado oferecendo flexibilidade no projeto experimental (KAPTEIN; THEEUWES; HORST, 1996).

Jogos comerciais, os quais são desenvolvidos essencialmente para proporcionar lazer as pessoas, têm sido empregados em pesquisas científicas para adquirir dados com alta fidelidade, sendo usados para o treinamento e validação de sistemas de percepção (RICHTER *et al.*, 2016; RICHTER; HAYDER; KOLTUN, 2017). Dentre os motores de jogos que vêm sendo empregados pela comunidade científica na constituição de simuladores veiculares pode-se citar o Unity (TAHERI; MATSUSHITA; SASAKI, 2017; RAIMUNDO, 2017) e o Unreal Engine

4 (DOSOVITSKIY *et al.*, 2017; SHAH *et al.*, 2018). Em termos de características, ambos os motores de jogos possuem características similares, mas o Unreal Engine 4 possui uma qualidade gráfica melhor, no entanto, o Unity é mais popular entre os desenvolvedores, possuindo ainda, uma maior quantidade de modelos, sons e efeitos disponíveis na plataforma (MAYDEN, 2014).

No restante deste capítulo, descreve-se o simulador, criado para este trabalho, e o processo de obtenção dos dados para a experimentação da solução proposta neste trabalho.

5.1 Simulador de direção para pesquisas sobre distrações e perturbações do motorista no trânsito

A coleta de dados reais em pesquisas sobre distrações e perturbações dos motoristas envolvem riscos aos participantes e a eventuais pessoas que possam cruzar com um participante. Para evitar ilegalidades e riscos em submeter pessoas embriagadas, falando ao celular ou sonolentas na direção veículos, pode-se utilizar um ambiente simulado. Simuladores já vem sendo empregados em pesquisas científicas para este fim, como exemplos, os trabalhos de Shirazi e Rad (2014) e Ebrahim *et al.* (2014) podem ser citados. Para a obtenção dos dados desta pesquisa criou-se o *Simulator for Distracted Driving Research* (SDDR) ou o Simulador para Pesquisa de Direção Distraída¹, ao qual, é composto por: tela de projeção frontal (e projetor); cockpit para acomodar o motorista; *G27 Racing Wheel*² com pedais de aceleração e freio, com câmbio de troca de marchas manual; um Kinect v2 (Seção 3.1.5); e, por fim, os *softwares* de simulação e controle da simulação. Os equipamentos (físicos) do SDDR e um exemplo de tela de simulação são mostrados pela Figura 17.

Figura 17 – Equipamentos (a) e um exemplo de tela (b) do *Simulator for Distracted Driving Research* (SDDR).

Fonte: Elaborada pelo autor.

O sensor Kinect v2, no SDDR, foi instalado frontalmente ao motorista³ possibilitando a

¹ A chamada para a participação na coleta de dados utilizando o SDDR pode ser vista no link <<https://www.youtube.com/watch?v=rvH0fmGrAVM>>

² Para mais informações sobre Logitech G27, veja o link <<https://goo.gl/p76eDw>>.

³ Foram realizados testes de posicionamento do Kinect v2 em um caminhão Scania que culminou

captura de imagens e 3D do motorista. A distância entre o Kinect e o motorista foi estabelecida em 1,5 metros, assim, o motorista permanece posicionado na região mais precisa do Kinect para a aquisição da profundidade (entre 0,5 e 3 metros) (YANG *et al.*, 2015).

O *software* de simulação do SDDR foi desenvolvido baseando-se no motor de jogos Unity⁴ e efetua a simulação de um ambiente rodoviário com duas faixas de rodagem em cada sentido (rodovia duplicada). O mapa utilizado foi desenvolvido baseando-se no traçado da rodovia brasileira Anhanguera, trecho entre as cidades de São Paulo e Campinas. A rodovia virtual foi circundada por uma floresta, as árvores auxiliam o participante na percepção de velocidade. O mapa possui ainda, placas de sinalização de velocidade máxima permitida, aviso de curvas, etc. Além da utilização de um mapa próximo ao real, o motorista dirige um Volkswagen GolfR com potência do motor, relação entre marchas/diferencial e esterçamento do volante baseados em um veículo real. O campo de visão do motorista (participante da simulação) está situado na parte interna do veículo, analogamente a uma situação real de condução, sendo o interior do GolfR recriado pelo simulador. Além do veículo conduzido pelo participante, o simulador controla a adição de tráfego (nas duas direções da rodovia). O GolfR, outros 3 tipos de automóveis e um tipo de caminhão são utilizados pelo simulador. Ao inserir um novo veículo na simulação, o modelo (tipo), cor e velocidade máxima são escolhidos randomicamente. O caminhão conta ainda com 3 tipos diferentes de cargas transportadas. Exceto o veículo do participante, todos os outros são autônomos e controlados pela simulação baseado em *waypoints*. Os veículos autônomos são capazes de efetuar ultrapassagem (mudança de faixa) ao perceberem um veículo mais lento à frente.

A simulação possui ainda um controle remoto, que nada mais é do que um *software* que permite controlar a simulação remotamente. Exemplos de operações importantes possíveis de serem realizadas com controle remoto são o aumento e diminuição do número de veículos que compõem o tráfego e o “salvamento” do participante, ou seja, o reposicionamento do veículo do participante na faixa de rodagem (em caso de necessidade extrema). O controle remoto mostra duas imagens ao controlador, as imagens RGB vindas do Kinect e do simulador, possibilitando observar como o participante está se comportando e dirigindo o veículo na simulação.

Cerca de 19 participantes⁵ utilizaram o SDDR e o avaliaram. A maioria dos avaliadores considerou a direção na simulação similar a dirigir um veículo real, como pode ser observado no gráfico da Figura 18. Os pontos fortes principais do SDDR destacados pelos participantes foram: aceleração do GolfR; os sons do simulador; a pista de rolagem (rodovia, placas, etc); a tela de projeção; volante, pedais, câmbio e encosto do motorista; paisagem (a floresta que

no desenvolvimento de um suporte próprio para o sensor. Um vídeo demonstrando ser viável, em um veículo real, o posicionamento deste sensor em frente ao motorista pode ser visto em <<https://youtu.be/K1cIe6cWCPA/>>.

⁴ O motor de jogos Unity está disponível em <<https://unity3d.com/>>.

⁵ É possível ver o SDDR em funcionamento na chamada de participação desta pesquisa no link <<https://goo.gl/Bn49L6>>

circundou a rodovia, por exemplo); e reconstituição do painel do GolfR. Os pontos negativos principais observados foram: freio do GolfR e o tráfego (outros veículos que também estavam na rodovia, por exemplo). A sensação de resposta do freio do GolfR foi diferente da de um veículo real, dando a impressão para o motorista, que o freio estava demorando muito para parar o veículo (capacidade de frenagem inferior à de um veículo real). Sobre o tráfego, os participantes mencionaram alguns problemas como veículos muito lentos na faixa de rodagem em regiões de subida (escolha de marchas equivocadas na subida pelo sistema de controle autônomo dos veículos) e de veículos que demoram em mudar de faixa de rolagem ao finalizar uma ultrapassagem (veículos permanecem na faixa da esquerda da rodovia mais tempo do que usualmente é praticado por um motorista humano).

Figura 18 – Resultado da avaliação do SDDR comparando a uma situação real, feita por 19 participantes, após o utilizar por uma hora.

Fonte: Elaborada pelo autor.

5.2 Conjunto de dados

Utilizando o SDDR, pôde-se obter um conjunto de dados para os experimentos que denominou-se de *Naturalistic Driver Behavior Dataset* (NDBD) ou Conjunto de Dados Naturalista do Comportamento do Motorista. Com o emprego de simulação, os motoristas puderam ser submetidos a distrações e perturbações em situações similares a uma encontrada na realidade.

As informações coletadas do motorista envolvem o monitoramento do motorista (postura na condução do veículo), dados obtidos do veículo virtual (envolvendo sensores, telemetria e outros sistemas instalados no veículo) e dados pertinentes a coleta de dados. Todas as informações coletadas são mostradas na Tabela 5. Os dados do monitoramento do motorista são dados reais, ou seja, são obtidos por um Kinect v2 físico instalado em frente do motorista. Os dados coletados pelo NDBD do monitoramento do motorista contém imagem colorida (RGB, ver Figura 19a), infravermelha (IR, ver Figura 19b), mapa de profundidade (3D) e a esqueletização do motorista (ver ambos na Figura 19c). Toda a sincronização dos sensores reais e virtuais é feita pelo SDDR empregando uma frequência de 10 *frames* (capturas) de informação por segundo.

A coleta de dados para os experimentos do sistema de detecção de distrações e perturbações (NDBD) envolveu 14 participantes. Todos os participantes eram homens com idades entre

Figura 19 – Imagens obtidas do SDDR.

Fonte: Elaborada pelo autor.

22 e 33 anos (média de 30 anos de idade) e com permissão para dirigir veículos no Brasil de 8 anos, em média. Todos os participantes utilizaram transmissão manual para a troca de marchas do veículo.

Dois tipos de Seções de gravações foram realizados. A Seção 1 foi realizada por 13 participantes e sem a incidência de sonolência. A Seção 2 foi realizada por 6 participantes que ficaram, ao menos, 36 horas sem dormir antes das gravações, estando os participantes sonolentos na hora da coleta de dados. O longo período sem dormir do participante da Seção 2 se deve ao fato de que um dos principais fatores da deteriorização da performance de direção é a privação de sono do motorista (PHILIP *et al.*, 2003). Apenas um participante realizou somente a Seção 2. Antes de cada seção de gravação o motorista participante dirigia o SDDR por 2 minutos, auxiliando na sua aclimatação com o veículo virtual e ambiente simulado.

As seções eram compostas de três baterias de gravação, todas com duração de 15 minutos. A primeira gravação foi denominada de “Direção Normal”, ou seja, o motorista dirigia sem qualquer distração ou perturbação. A segunda bateria de gravações é a “Direção Embriagada”, ao qual, durante toda a gravação a visão do motorista esteve perturbada pela utilização dos óculos

Tabela 5 – Todas as informações disponíveis em cada *frame* do NDBD

Origem	Informações disponíveis
Monitoramento do Motorista (Kinect v2 - real)	Imagens frontais ao motorista disponíveis: - colorida (RGB) - infravermelha (IR) - mapa de profundidade (3D). Esqueletização (Seção 3.3) do motorista.
Telemetria do veículo virtual)	Velocidade Aceleração Ângulo do volante Nível de acionamento do pedal da aceleração Nível do pedal do freio
Sensores instalados no veículo virtual	GPS Acelerômetro Giroscópio Bússola Câmera RGB (focado no caminho à frente do veículo, ver exemplo na Figura 19d)
Contexto da direção (booleanos)	- Está em uma curva à direita? - Está em uma curva à esquerda? - Está em um trecho reto? - Está mudando de faixa? - Está ultrapassando? - Veículo parado? - Está freando? - Está na marcha neutra? - Está indo para frente? - Está em marcha ré?
Sistema de posicionamento do veículo virtual	- Distância do centro do veículo para a linha central da faixa - Ângulo formado entre a linha central do veículo e a linha central da faixa
Sistema de coleta de dados	- O veículo foi reposicionado na via?

*Drunk Busters Impairment Goggles*⁶ (KINDSCHUH, 2001) que simula a visão de um motorista embriagado a níveis entre 80 e 15 mg/100ml. Na Figura 20 é mostrado o *Drunk Busters*. A terceira gravação é a “Direção ao celular”, na qual o motorista recebe várias ligações telefônicas e estabelece uma conversa com um moderador, assim, foi distraído pelo uso do celular. Os participantes sempre utilizaram o celular ao ouvido nas chamadas telefônicas, foram obrigados portanto, a segurar o celular com uma das mãos e com a outra conduzir o veículo. Para balizar as conversas (ligações) entre o moderador e o motorista foram utilizadas questões retiradas do livro “*The Complete Book of Questions*” (POOLE, 2003), nos capítulos contendo questões “leves e fáceis” e sobre preferências pessoais. Exemplo de questões utilizadas foram: “Você aperta o

⁶ Veja mais informações sobre o *Drunk Buster* no link <https://goo.gl/pntvR4>.

tubo de pasta de dente ou enrola? Qual é a vantagem do seu método?” e “Qual é o seu utensílio favorito para cozinhar? Por quê?”. Todas as questões obrigaram ao participante pensar antes de responder. Na Tabela 6, são mostrados a quantidade de frames gravados para cada pessoa e seção de coleta de dados.

Figura 20 – *Drunk Busters Impairment Goggles* utilizado na simulação da visão embriagada.

Fonte: Elaborada pelo autor.

Tabela 6 – Frames do NDBD por pessoa e Seção.

Pessoa	Seção	Normal	Embriagada	Celular
1	1	9.111	10.010	9.102
1	2	9.028	9.064	9.195
2	1	9.066	9.026	9.642
3	1	9.078	9.101	9.069
4	1	9.088	9.078	9.162
5	1	9.058	8.939	9.460
6	1	9.105	9.046	9.053
7	1	9.076	9.061	9.048
7	2	9.276	9.171	9.108
8	1	9.061	9.093	9.164
8	2	9.054	9.059	9.059
9	1	9.066	9.116	9.304
10	2	9.096	9.060	9.094
11	1	9.063	9.065	9.114
11	2	9.064	9.086	9.303
12	1	9.059	9.065	9.171
13	1	9.056	9.153	9.079
14	1	9.152	9.057	9.062
14	2	9.064	9.249	9.104
Total	1	118.039	118.810	119.430
	1	196 minutos	198 minutos	199 minutos
Total	2	54.582	54.689	54.863
	2	90 minutos	91 minutos	91 minutos

Após a obtenção dos dados, cada *frame* das gravações foi rotulado entre dois tipos de situações “normal” ou “risco”. A situação normal indica que o motorista operou o veículo sem

distrações ou perturbações aparentes. A situação de risco indica que o motorista foi induzido a alguma forma de perturbação ou distração. Para a Seção 1, todos os dados da bateria de “Direção Normal” foi atribuído o rótulo “normal” e foi rotulada como risco todos os dados da “Direção Embriagada”. Na “Direção ao celular” da Seção 1, foi realizado a rotulação manual por uma pessoa, onde, os *frames* em que o motorista está somente dirigindo o veículo foram rotulados como “normal” e os dados em que o motorista fala ao celular foram marcados como de “risco”. Foi adotado o momento em que o motorista retira uma das mãos do volante para atender o telefone como o início do risco e o início de uma situação normal ocorre quando o motorista retorna com a mão ao volante. Todos os dados da Seção 2 (incluindo a bateria “Direção Normal”) foram rotulados como de risco, pelo fato do motorista estar sonolento no momento da gravação.

5.3 Considerações

Neste capítulo foram apresentados o simulador desenvolvido especialmente para este trabalho baseando no motor de jogos Unity e a descrição do processo de obtenção de dados utilizando o próprio sistema simulador. Denominou-se o simulador de SDDR e foi dotado de especificações que facilitam a coleta de dados para distrações e perturbações, deixam o motorista se sentindo próximo de uma direção real sem correr qualquer risco, mesmo enquanto dirige sendo distraído em algum aspecto. O SDDR faz a sincronização precisa entre os dados coletados no mundo virtual (dados do simulador) e os dados coletados de um sensor Kinect v2 que adquire dados reais do motorista durante a simulação, todos os dados são agrupados em um *frame* de dados. Portanto, cada *frame* de dado gravado pelo sistema simulador contém características extraídas diretamente do mundo virtual como dados de telemetria, sensores (inerciais, GPS, dentre outros) e uma câmera frontal ao veículo virtual; dados de posicionamento do veículo na pista de rodagem; contexto da direção (em uma curva, reta, etc) e dados provenientes do Kinect v2. Os dados obtidos do Kinect são gravados sem perda de dados, sendo a imagem RGB gravado do motorista com resolução de $1920 \times 1080 \text{ pixels}$ (Full HD), a resolução da imagem IR e de profundidade foi de $512 \times 424 \text{ pixels}$, e a esqueletização de todas as pessoas contidas na cena (real) (até o máximo de 6 pessoas) estão disponíveis no *frame* de dados provido pelo SDDR. Cada *frame* de dados do SDDR ocupa cerca de 9 MB, podendo ser compactado em memória sem perda de dados reduzindo o tamanho final do *frame* para valores entre 30-50% da ocupação original, possibilitando assim, uma diminuição significativa no tamanho de uma base de dados gravada utilizando o SDDR. O SDDR foi especialmente projetado com um programa “controle remoto”, que permite controlar toda a simulação à distância, ideal para alguns tipos experimentos com distrações e perturbações do motorista, fazendo com que o controlador da simulação deixe o motorista sozinho na sala da simulação. O “controle remoto” foi extremamente útil para coletar dados sobre o uso de celular no trânsito, onde, a própria pessoa que controlava a simulação foi capaz de efetuar ligações telefônicas ao motorista estando em outro ambiente, o que foi fundamental para o realismo da coleta. Por essas características, o SDDR pode vir a ser utilizado

para novas coletas de dados no futuro para as distrações e perturbações dentro e fora do foco deste trabalho.

Foi desenvolvida uma base de dados com a participação de 14 motoristas voluntários totalizando 865 minutos de gravação e foi denominada de NDBD. O NDBD é formado por dados coletados (10 Hz) em duas Seções, que não foram obrigatoriamente desempenhadas por todos os participantes. Cada seção é formada por dados referente à direção “normal” (sem qualquer tipo de distração ou perturbação), embriagada e de uso do celular. Na Seção 1 de gravação, os motoristas operaram o veículo virtual sem a incidência de sonolência, que é por sua vez o foco da Seção 2. Cada seção de gravação ocupa cerca de 90 GB (45 min de dados gravados) já compactados pelo SDDR e sendo cada *frame* rotulado por uma pessoa entre uma situação com e sem risco. Embora o NDBD tenha sido gerado visando este trabalho, pode-se utilizá-lo em outros experimentos dada a riqueza dos dados coletados. Novas características podem ser desenvolvidas e novas arquiteturas de sistemas podem ser experimentadas, já que contém todos os dados intactos do *frame* provido pelo SDDR. Os dados do NDBD também podem servir de base para outros tipos de estudos, como por exemplo, o desenvolvimento de novos algoritmos de rastreamento de pessoas, esqueletização, desenvolvimento de extratores de dados 3D, etc.

O SDBD e o NDBD são contribuições deste trabalho, além da contribuição principal que é a proposta, a especificação e validação de um sistema de detecção das anomalias na condução de veículos.

No Capítulo 6, os dados obtidos (NDBD) são empregados para a obtenção de parâmetro empíricos e validação do sistema proposto no Capítulo 4.

EXPERIMENTOS, RESULTADOS E DISCUSSÕES

Neste capítulo, os experimentos do sistema de detecção de distrações e perturbações do motorista são descritos. As discussões incluem detalhes sobre treinamento de classificadores, assim como, os resultados em um grande conjunto de dados. A técnica 9-fold cross-validation (KOHAVI, 1995) foi utilizada, como análise estatística em todos os treinamentos dos classificadores. Todos os dados utilizados nesta seção são provenientes do NDBD (Seção 5.2), sendo utilizados dados de 14 pessoas (direção normal, falando ao celular, embriagado e sonolento), com 7 para o treinamento de classificadores e 7 para validação. Os dados de treinamento e validação envolvem direção normal, uso de celular e embriaguez, porém, alguns experimentos aqui descritos usam parte destes dados coletados sobre as perturbações e distrações do motorista. Foi criado um grupo separado para avaliar sistemas para a sonolência contendo dados de 6 motoristas sonolentos, sendo então, estes dados formadores de um segundo grupo de validação.

Usando todos os dados do NDBD, pôde-se definir empiricamente o valor para ϕ (Equação 4.1) para cada característica do monitoramento do motorista (ver Seção 4.1.1). O valor mínimo de ϕ obtido para a característica em todos os *frames* e participantes foi adotado. Para o cálculo do ϕ foram empregados: o valor absoluto de cada característica no *frame* foi utilizado como $feature_v$; o valor para $feature_n$ ou valor da característica já normalizado entre -1 e 1 (valor de saída da equação); e o comprimento dos ombros ($shoulder_l$) utilizou-se o observado no *frame* do NDBD. Assim, o valor adotado para ϕ em cada uma das características nas origens VOL e CM são mostrados na Tabela 7.

Todas as características da Qualidade de Direção (ver Seção 4.1.2), ou seja, suas médias e desvio padrão utilizados em períodos e janelas de dados, para todos os *frames*, foram normalizados entre -1 e 1 .

Para avaliar a qualidade dos classificadores utiliza-se o $Findex_\eta = \frac{(\eta^2+1)PR}{\eta^2P+R}$, onde, η é 5 neste trabalho (definição empírica), P é a medida de Precisão (*Precision*) e R é cobertura (*Recall*)

Tabela 7 – Valor adotado para φ em cada uma das características nas origens VOL e CM.

Carac.	$minDistMaos_{VOL}$	$maxDistMaos_{VOL}$	$distME_{VOL}$	xME_{VOL}	yME_{VOL}
φ	1,69	1,62	1,66	8,11	4,09
Carac.	zME_{VOL}	$distMD_{VOL}$	xMD_{VOL}	yMD_{VOL}	zMD_{VOL}
φ	1,72	1,61	4,94	5,35	1,68
Carac.	$minDistMaos_{CM}$	$maxDistMaos_{CM}$	$distME_{CM}$	xME_{CM}	yME_{CM}
φ	2,86	2,62	2,62	7,89	3,32
Carac.	zME_{CM}	$distMD_{CM}$	xMD_{CM}	yMD_{CM}	zMD_{CM}
φ	2,76	2,75	5,12	3,32	2,97

das predições. Essa equação nada mais é do que uma variação do *Rand Index* (RAND, 1971), incluindo uma ponderação entre P e R . P e R são calculados levando-se em conta a situação “normal”, assim, o $Findex_5$ objetiva a diminuição de falsas detecção de risco, dando mais peso a cobertura. O índice $Findex_5$ possui valores entre 0 e 1, sendo, 1 o melhor valor.

No restante deste capítulo, os experimentos realizados no NDBD são mostrados, sendo na Seção 6.1 são descritos os experimentos utilizando as características da Qualidade de Direção e embriaguez; experimentações utilizando o Monitoramento do Motorista e o uso do celular no trânsito são apresentados na Seção 6.2; e por fim, na Seção 6.3 experimentos contendo todas as fontes de características (Monitoramento do Motorista e Qualidade de Direção) e todas as distrações e perturbações (uso de celular, embriaguez e sonolência) são mostrados.

6.1 Qualidade na direção

Foram realizados experimentos utilizando-se como foco a qualidade da direção (experimentação previamente publicada em Berri e Osório (2018b)), ou seja, empregando-se apenas as características de análise da direção (ver Seção 5.2) para determinar se um motorista se encontra embriagado (risco) ou sóbrio (normal), portanto, a embriaguez foi a perturbação objetivo destes experimentos e a arquitetura SB (Seção 4.3) foi a escolhida para o sistema.

Usando SVM Linear (ver Seção 3.5.1) na procura pela margem máxima entre as classes “risco” e “normal”, dados de treinamento (NDBD) já previamente separados em períodos de frequência variando entre 5 e 300 *frames*, combinando as 28 características candidatas (conforme Tabela 4 - Seção 4.1.2, sendo 14 de médias e 14 desvios padrões) e empregando o $Findex_5$, pôde-se obter um grupo de características para frequências com $Findex_5$ elevados (considerando os diferentes comprimentos de períodos).

Foi desenvolvido um método para a seleção das características mais relevantes, descrito a seguir, onde definiu-se o melhor conjunto de características em relação ao tamanho do período analisado (número de frames). Em cada frequência de período, a busca foi iniciada verificando todas as características individualmente e a que obteve melhor $Findex_5$ (mais próximo de 1)

foi incluída no grupo de características já testadas do período. Nos passos seguintes, todas as características ainda não incluídas no grupo são testadas, e individualmente combinadas com as características já incluídas, sendo a característica que obtiver o mais elevado $Findex_5$ inserida no grupo de testadas. O processo de verificação das características termina quando todas as características estiverem no grupo de testadas. De todos os agrupamentos testados para a frequência, o de maior $Findex_5$ é o considerado mais vantajoso. A Figura 21 mostra os melhores $Findex_5$ para cada comprimento de período verificado.

Figura 21 – Gráfico dos melhores $Findex_5$ para Qualidade na Direção obtidos em cada comprimento de período. O período com 220 *frames* possui o melhor $Findex_5$, 80 *frames* é o melhor $Findex_5$ para períodos curtos e 300 é a melhor opção para períodos longos.

Fonte: Elaborada pelo autor.

Abaixo dos 150 *frames* de comprimento (período curto), o melhor $Findex_5$ obtido foi 0,968 usando 80 *frames* para períodos. As características selecionadas que maximizaram o $Findex_5$ para 80 *frames* de comprimento de período foram: $gZ@DP$, $aY@MED$, $distCF@MED$, $gZ@MED$, $acel@MED$, $VOLangulo@MED$, $aY@MED$, onde, $@MED$ representa uma característica de média e $@DP$ uma característica de desvio padrão para os períodos. Para períodos de 220 *frames*, o mais elevado $Findex_5$ obtido foi 0,972, empregando as características $gX@DP$, $acel@DP$, $VOLangulo@DP$, $distCF@MED$, $PFlevel@DP$, $gY@MED$, $gX@MED$, $aZ@DP$, $gZ@MED$, $aY@MED$, $VOLangulo@MED$, $aY@DP$, $aX@MED$ and $gZ@DP$. Já em períodos de 300 *frames*, o melhor $Findex_5$ obtido foi de 0,969 usando o grupo de características: $acel@DP$, $VOLangulo@DP$, $aZ@MED$, $gX@MED$, $aX@MED$, $acel@MED$, $distCF@MED$, $lAngle@MED$, $gX@DP$, $PFlevel@SD$, $PFlevel@AV$ e $gY@MED$.

Em todos os períodos, dados provenientes da telemetria do veículo (aceleração e ângulo do volante) e sensores inerciais (acelerômetro e giroscópio) estiveram entre as melhores características. Para períodos de 220 e 300 *frames*, características dos dados sobre a faixa de rodagem (distância para o centro da faixa) e o nível do pedal de freio (telemetria do veículo) foram incluídos no grupo de melhores características para detecção de um motorista embriagado.

Classificadores para períodos de 80, 220 e 300 *frames* foram obtidos usando os dados de

treinamento e empregando seus grupos de características, redes neurais MLP (ver Seção 3.5.2) como técnica de classificação e Rprop (RIEDMILLER; BRAUN, 1992) como meio de treinamento da rede. As funções de ativação Gaussiana e Sigmoid foram usadas nos experimentos. Para maximizar o $Findex_5$ e a acurácia do sistema classificador, usou-se Algoritmo Genético¹ (AG) com codificação binária na localização e otimização de parâmetros. A função objetivo do AG empregada foi $fitness = 0.6Findex_5 + 0.4A$, onde, A representa a acurácia do classificador. Os parâmetros otimizados pelo AG foram α , β , CO e $interac$. α e β são parâmetros das funções de ativação, CO contém o número de neurônios na camada oculta e $interac$ representa o número máximo de interações para o método de aprendizado Rprop. A faixa permitida para valores de α e β foram entre 0,001 e 10. Para $interac$ o limite de valores imposto foram entre 100 e 500. O limite de neurônios contido para CO foram para 80 [7-17], 220 [14-24] e 300 [12-22] (as faixas foram baseadas no número de características selecionadas).

Os parâmetros do AG foram definidos empiricamente com 400 gerações e 10 indivíduos por geração, 80% de índice de *crossover*, 5% de índice de mutação e como método de seleção foi empregado torneio. A população inicial foi gerada randomicamente. O comprimento do cromossomos dos indivíduos é de 66 *bits*. O indivíduo (classificador) com mais alto *fitness* foi adotado como resposta para cada teste. AG foi performado três vezes para cada teste. O mesmo número de características empregadas no teste determina o número de neurônios na camada de entrada da rede. Na camada de saída, somente um neurônio é utilizado (saída booleana de risco/sem risco) e CO (AG encontra a melhor opção para este parâmetro da rede) neurônios para a camada oculta da rede. Os melhores parâmetros encontrados pelo AG são mostrados na Tabela 8. Para todos os classificadores adotados nas funções de ativação Gaussiana (Gaussian) ou Sigmoid (Sigmoid) são estatisticamente similares para $Findex_5$, o que mostra um resultado satisfatório das execuções do AG pois foram obtidos parâmetros dos classificadores que resultam em poder classificatório similar usando os dados do NDBD. A Figura 22 mostra os resultados obtidos para os classificadores para os dados de treinamento e validação.

Tabela 8 – Parâmetros dos classificadores MLP para Qualidade na Direção.

Classificador	α	β	CO	$interac$
GAUSSIAN80	0,01	1,36	17	432
SIGMOID80	5,07	3,74	8	210
GAUSSIAN220	2,73	0,92	22	487
SIGMOID220	6,74	0,74	18	389
GAUSSIAN300	0,01	0,62	19	478
SIGMOID300	4,26	0,98	22	444

Usando o sistema proposto SB foi avaliado e o mais eficiente classificador foi selecionado. Nos experimentos realizados em todos os classificadores testados usando os dados disponíveis

¹ A biblioteca GALib versão 2.4.7 foi empregada (<http://lancet.mit.edu/ga>).

Figura 22 – Gráfico dos classificadores da Qualidade na Direção obtidos do AG, sendo os dados de treinamento (cores escuras) e dados de validação (cores claras), onde, a barra vermelha é o $Findex_5$, a barra verde é a acurácia e os limites do desvio padrão nos grupos de *cross-validation* de treinamento são mostrados em cinza.

Fonte: Elaborada pelo autor.

no NDBD. O tamanho do *Buffer* de Avaliação (*BS*) foi empiricamente definido como 3, então, enquanto ao menos 2 das últimas 3 predições do classificador são “risco”, a saída do sistema é “risco”, de outra forma, uma situação normal é detectada. Os resultados obtidos com SB usando cada um dos classificadores são mostrados na Figura 23 para os dados de treinamento e validação.

Figura 23 – Resultados dos seis sistemas para detecção de embriaguez considerando o buffer de avaliação (*BS*), onde, onde, as barras vermelhas são o índice $Findex_5$ e as verde a acurácia final. As cores escuras e claras indicam conjuntos de treinamento e validação, respectivamente.

Fonte: Elaborada pelo autor.

Todos os classificadores alcançaram um $Findex_5$ similar. O mais elevado $Findex_5$ obtido foi do sistema sSIGMOID220 que resultou 0,988 e 0,981 para dados de treinamento e validação, respectivamente. Considerando a acurácia dos sistemas, sSIGMOID220, sGAUSSIAN300 e sSIGMOID300 foram similares, sendo a mais alta acurácia atingida por sGAUSSIAN300, alcançando 0,986 para dados de treinamento e 0,974 para validação. É importante destacar que a abordagem proposta, alcançou um elevado grau de acertos e de confiabilidade (97,4% para os dados de validação), considerando o melhor classificador obtido em todo este processo de escolha de parâmetros e de configurações.

De todos os sistemas, o sSIGMOID220 foi considerado com melhor classificação, onde, os erros de classificação obtidos foram de 1,5% de todos os *frames* rotulados como “normal” (sem risco) do NDBD e atingiu uma acurácia final de 98%. Todos os dados do NDBD² foram processados pelo sSIGMOID220. O nf do sSIGMOID220 é 220, ou seja, o sistema necessita de informações de 220 *frames* consecutivo (período) para obter uma nova predição (22 segundos de dados). Assim, o sistema detecta um motorista embriagado usando entre 2 e 3 predições, o que resulta na necessidade de ao menos 44 segundos de dados para detectar um motorista embriagado.

Um primeiro aviso de “risco de embriaguez” pode ser obtido em um intervalo de tempo bem menor, porém com uma acurácia também inferior (com 80 *frames* a acurácia fica em torno de 75% de acerto). Por outro lado, ao considerar um intervalo de tempo menor, ainda assim praticamente não haveriam falsos alarmes, pois como é indicado pelo alto valor do $Findex_5$, a maioria dos erros neste caso seria devido aos Falsos Negativos, ou seja, o sistema estaria omitindo-se na identificação da embriaguez. É melhor ter um tempo de diagnóstico mais longo do que emitir falsos alertas (Falsos Positivos), que podem causar mais problemas do que ajudar.

6.2 Monitoramento do Motorista

O Monitoramento do Motorista foi foco de experimentos independentes (previamente publicados em Berri e Osório (2018a)). Nestes experimentos, portanto, apenas características sobre a análise do motorista e sua postura (ver Seção 5.2) foram empregadas para determinar se o motorista encontrava-se distraído (risco) ou com o foco na direção (normal). O uso de celular no trânsito foi a distração utilizada na validação e treinamento do sistema. A arquitetura empregada foi a de SCLP (Seção 4.3), ou seja, permitindo avisos leves para a reeducação do motorista (retorno a uma pose salutar para a condução do veículo).

A margem máxima de separação das classes, “risco” e “normal”, foram obtidas usando SVM linear (Seção 3.5.1). As características para frequências (nCP e nLP) foram obtidos usando NDBD com períodos entre 5 e 300 *frames*, 40 características candidatas (conforme Tabela 3 - Seção 4.1.1, sendo 20 médias e 20 desvios padrões) e $Findex_5$. As características e frequências

² Veja todos os vídeos resultantes do processamento do NDBD em <https://goo.gl/QLsXaV>.

que melhoram o $Findex_5$ foram selecionadas. Para cada período de teste, iniciou-se com todas as características incluídas no grupo de características candidatas e, em cada novo ciclo, a característica que ao ser excluída do grupo gere o maior $Findex_5$ foi removida. Assim, as características com menos influência no $Findex_5$ foram excluídas até que restasse apenas uma. A Figura 24 mostra o mais elevado $Findex_5$ dos melhores grupos de características testados para todas as frequências.

Figura 24 – Gráfico dos melhores $Findex_5$ obtidos em cada comprimento de períodos no Monitoramento do Motorista.

Fonte: Elaborada pelo autor.

Considerando o Longo Prazo (LP), os períodos de 80 e 210 frames obtiveram o mais alto $Findex_5$ de 0,995, sendo esses os nLP candidatos para o *classificadorLP*. As características selecionadas para períodos de 80 frames foram $distME@DP_{VOL}$, $xME@MED_{VOL}$, $xMD@MED_{VOL}$, $xMD@DP_{VOL}$, $yMD@DP_{VOL}$, $zMD@MED_{VOL}$, $minDistMaos@MED_{CM}$, $distME@DP_{CM}$, $xME@DP_{CM}$ e $yMD@MED_{CM}$, onde, @MED indica média e @DP desvio padrão do período. Para períodos de 210 frames as características com melhor $Findex_5$ foram: $xME@MED_{VOL}$, $distME@MED_{VOL}$, $distME@MED_{CM}$, $distMD@MED_{CM}$ e $yMD@MED_{CM}$.

Para o Curto Prazo (CP), adotou-se períodos de 5 frames, ou seja, a mínima escolha possível. A Figura 25 mostra o $Findex_5$ obtido para cada tamanho do grupo de características testados para 5 frames. O melhor $Findex_5$ foi 0,984 para o grupo de características chamado de 5@1, sendo composto por: $zMD@MED_{VOL}$, $minDistMaos@MED_{CM}$, $distMD@MED_{CM}$, e $yMD@MED_{CM}$. Outro grupo de características chamado de 5@2 foi selecionado por obter um $Findex_5$ similar. As características pertencentes ao 5@2 foram $minDistMaos@DP_{VOL}$, $maxDistMaos@MED_{VOL}$, $xME@DP_{VOL}$, $zME@DP_{VOL}$, $distMD@DP_{VOL}$, $xMD@MED_{VOL}$, $yMD@MED_{VOL}$, $zMD@MED_{VOL}$, $minDistMaos@MED_{CM}$, $maxDistMaos@MED_{CM}$, $maxDistMaos@DP_{CM}$, $xME@DP_{CM}$, $yME@DP_{CM}$, $zME@DP_{CM}$, $distMD@MED_{CM}$, $xMD@MED_{CM}$, $yMD@MED_{CM}$, $yMD@DP_{CM}$ e $zMD@MED_{CM}$.

Classificadores para os períodos selecionados foram obtidos usando a técnica de classificação MLP (ver Seção 3.5.2) e Rprop (RIEDMILLER; BRAUN, 1992) para treinamento

Figura 25 – Gráfico dos $Findex_5$ para cada tamanho de grupo de características usando período de 5 frames no Monitoramento do Motorista.

Fonte: Elaborada pelo autor.

da rede. Gaussiana (Gaussian) e Sigmoide (Sigmoid) foram as funções de ativação usadas nos experimentos. Os parâmetros dos sistemas de classificação foram maximizados ($Findex_5$ e acurácia) utilizando Algoritmo Genético³ (AG) (GOLDBERG, 1989). A função objetivo utilizada é $fitness = \kappa Findex_5 + (1 - \kappa)A$, onde, A é a acurácia do classificador e κ foi estabelecido 0,6 para classificadores de CP (5@1 e 5@2) e 0.99 para o LP (80 e 210 frames). A Tabela 9 mostra os limites para cada parâmetro e grupo de característica (frequência do período) utilizados, onde, α e β foram parâmetros da função de ativação, CO foi o número de neurônios na camada oculta e $interac$ foi usado para definir o número de interações do aprendizado da rede (Rprop).

Tabela 9 – Parâmetros da MLP empregados no AG para o Monitoramento do Motorista

Parâm.	Grupos →	5@1	5@2	80	210	Bits
α	Mínimo	10^{-3}	10^{-3}	10^{-3}	10^{-3}	25
	Máximo	20	20	20	20	
	Decimal	3	3	3	3	
β	Mínimo	10^{-3}	10^{-3}	10^{-3}	10^{-3}	25
	Máximo	20	20	20	20	
	Decimal	3	3	3	3	
CO	Mínimo	4	19	10	5	8
	Máximo	14	29	20	15	
$interac$	Mínimo	100	100	100	100	8
	Máximo	300	300	300	300	
Tamanho do código Genético						66

O AG contou com 10 indivíduos em 400 gerações, *crossover* de 80%, possibilidade de mutação em 5% e seleção por torneio. A população inicial foi gerada randomicamente. O tamanho do cromossomo do AG adotado foi de 66 bits. O mais elevado *fitness* foi adotado como resposta em cada teste de classificadores efetuados, sendo AG testado três vezes para

³ A biblioteca GALib versão 2.4.7 foi empregada (<http://lancet.mit.edu/ga>).

cada tipo de teste. A rede dos classificadores empregada possuía o número de neurônios na camada de entrada igual ao número de características utilizadas, um neurônio foi usado na camada de saída e, na camada oculta, são CO neurônios, ou seja, AG definiu a quantidade de neurônios nesta camada. Os resultados do processo de treinamento são mostrados pela Figura 26 e a Tabela 10 mostra os melhores parâmetros encontrados pelo AG. Todos os classificadores obtidos são estatisticamente similares para médias de $Findex_5$ e da acurácia, sendo portanto o AG capaz de obter bons parâmetros para todos os classificadores utilizados.

Figura 26 – Gráfico dos classificadores para o Monitoramento do Motorista obtidos usando AG, *cross-validation* e dados de teste, onde, a barra vermelha indica o $Findex_5$, a verde é indicativa de acurácia e os limites de um desvio padrão são mostrados em cinza.

Fonte: Elaborada pelo autor.

Tabela 10 – Parâmetros dos classificadores MLP para detecção de celular.

Classificadores	α	β	CO	<i>interac</i>
Gaussian5@1	5,68	2,63	11	171
Sigmoid5@1	6,97	1,96	12	193
Gaussian5@2	0,17	0,99	28	100
Sigmoid5@2	2,69	2,49	26	240
Gaussian80	0,11	3,17	10	101
Sigmoid80	3,49	1,3	10	218
Gaussian210	4,76	4,83	14	298
Sigmoid210	3,17	2	15	162

Usando o mais elevado $Findex_5$ dos grupos de características, formou-se duas opções de sistemas de CP e LP . Assim, as opções para CP são sigmoide (ss5@1) ao qual utiliza as características de 5@1 e $nCP = 5$; e gaussiana (sg5@2) como função de ativação, sendo características 5@2 e $nCP = 5$ para esta última opção.

Para LP, gaussiana (lg80) com $nLP = 80$ e as características selecionadas para 80 frames; e sigmoide (lg210) usando $nLP = 210$ e grupo de característica de 210 frames. Usando $npCP = 6$, ou em outras palavras, usando as seis últimas previsões (3 segundos de dados) no CP para detectar uma pose de risco, pode-se constituir quatro sistemas (SCLP) distintos pelas suas parametrizações individuais. A Figura 27 mostra os resultados obtidos para os sistemas no NDBD.

Figura 27 – Resultado dos quatro SCLP para detecção de risco do motorista pelo uso de celular enquanto dirige, onde, a barra vermelha indica o $Findex_5$, a azul a acurácia dos sistemas de alertas, em verde a acurácia do alerta baixo e amarelo é a acurácia para alerta alto.

Fonte: Elaborada pelo autor.

Todos os sistemas atingiram um valor para $Findex_5$ similar, ss5@1 foi de 0,989 e 0,989 para ss5@1. O *classificadorCP* determina o $Findex_5$ do sistema por ser responsável pela indicação da presença ou ausência de risco. Já o *classificadorLP* determina o nível de alerta de risco (máximo ou mínimo). 5@1 tem quatro características e todas estas estão contidas no 5@2, então, pode-se concluir que 5@1 contém as características mais relevantes para o monitoramento da pose motorista no tocante a detecção de celular para períodos de 5 frames. O valor para Cobertura (*Recall*) “normal” para o SCLP foi de 0,99 empregando ambos os grupos de características, assim, 1% de todos os frames rotulados como “normal” no NDBD foram preditos incorretamente (falso positivo).

Os sistemas LP têm somente uma característica em comum, o $yMD@MED_{CM}$. O ls210 foi a melhor opção para *classificadorLP* por possuir a melhor acurácia nos alertas (veja a linha azul na Figura 27). A acurácia do alerta foi vantajosa com o ls210 por possuir a acurácia de alerta superior ao lg80.

O ss5@1_ls210 foi o melhor sistema obtido nos experimentos do SCLP na detecção do uso de celular, alcançando acurácia de 95% dos alertas nos dados do NDBD. Todos os dados do

NDBD⁴ foram processados pelo ss5@1_ls210. É importante destacar que a abordagem proposta, alcançou um elevado grau de acertos e de confiabilidade (acurácia de 95% para os dados da base NDBD e $Findex_5$ de 0,99), considerando o melhor classificador obtido em todo este processo de escolha de parâmetros e de configurações. O desempenho de curto prazo (CP) também alcançou níveis de $Findex_5$ da ordem de 0,989, o que indica que o sistema pode ter um relevante papel e atuação na reeducação de motoristas em relação ao uso de celular enquanto dirige (sendo emitidos poucos falsos alertas).

6.3 Monitoramento do Motorista e Qualidade na Direção

Nos experimentos apresentados a seguir, todas as características apresentadas na Seção 5.2 foram candidatas a constituir um sistema classificador utilizando a arquitetura SCLP (Seção 4.3) que fosse capaz de detectar as distrações e perturbações do motorista foco deste trabalho, o uso de celular no trânsito, embriaguez e sonolência. O objetivo portanto é apresentar um sistema capaz de distinguir entre momentos de risco (motorista distraído ou perturbado) e momentos de direção normal. Os dados do NDBD sobre sonolência não fizeram parte do treinamento dos sistemas, portanto, contribuem apenas na validação dos sistemas e verificação da capacidade de generalização do sistema para outros casos perturbadores do motorista.

A primeira parte dos experimentos consistiu na identificação do hiperplano separador de máxima margem para as classes “risco” e “normal”, para isso utilizou-se SVM linear (ver Seção 3.5.1), NDBD já em períodos entre 5 e 300 *frames*, as 68 características candidatas (20 médias e 20 desvios padrões do monitoramento do motorista e 14 médias e 14 desvios padrões da qualidade da direção; ver Seção 4.1) combinadas e o avaliador $Findex_5$. Assim, pôde-se verificar o grupo de características e frequência (nCP e nLP) que maximizam o $Findex_5$. Em cada frequência de período, a obtenção do grupo de características de máximo $Findex_5$ foi iniciado com a escolha da característica com $Findex_5$ mais próximo a 1 (melhor). A característica com melhor $Findex_5$ é a primeira integrante do grupo das selecionadas. Dentre as ainda não selecionadas, a característica que combinada ao grupo das selecionadas maximize o $Findex_5$ é adicionada ao grupo. O processo perdura até que todas as 68 características estejam selecionadas (uma a uma) e, assim sendo, verificada a qualidade da separação das classes com o SVM Linear utilizando-se grupos de características selecionadas com tamanhos distintos. A Figura 28 mostra o melhor conjunto de características em cada frequência experimentada.

Dois períodos de tamanho longos, 140 e 280 *frames* de comprimento, obtiveram os $Findex_5$ mais altos de 0,951, assim, estes foram os candidatos para nLP . As características selecionadas para as frequências candidatas de longo prazo são mostradas na Tabela 11, onde, @MED indica uma característica de média, @DP de desvio padrão para os períodos, VOL indica uma característica proveniente da origem do volante e CM com origem na cabeça do motorista.

⁴ Todos os resultados do processamento do NDBD estão disponíveis em <https://goo.gl/E9UcFK>.

Figura 28 – Gráfico para Monitoramento do Motorista e Qualidade na Direção dos melhores $Findex_5$ obtidos para cada tamanho de período.

Fonte: Elaborada pelo autor.

Tabela 11 – Todas as características escolhidas para as frequências candidatas a nLP para Monitoramento do Motorista e Qualidade na Direção.

Período	Qtde	Características
140	25	$minDistMaos@MED_{CM}$, $xME@MED_{CM}$, $gZ@DP$, $PFlevel@MED$, $aY@MED$, $aY@DP$, $velocidade@DP$, $gX@MED$, $VOLangulo@MED$, $gZ@MED$, $gX@DP$, $aZ@MED$, $fAngulo@DP$, $zMD@DP_{CM}$, $zME@MED_{CM}$, $distCF@DP$, $zMD@MED_{VOL}$, $distME@MED_{CM}$, $yMD@DP_{CM}$, $gZ@MED$, $rReta@MED$, $PFlevel@DP$, $yMD@MED_{CM}$, $aZ@DP$ e $xMD@MED_{VOL}$
280	41	$maxDistMaos@MED_{CM}$, $xMD@MED_{VOL}$, $acel@DP$, $aZ@MED$, $maxDistMaos@DP_{CM}$, $fAngulo@MED$, $zME@MED_{CM}$, $zME@MED_{VOL}$, $distME@MED_{VOL}$, $minDistMaos@MED_{VOL}$, $yMD@MED_{CM}$, $gY@MED$, $rReta@DP$, $VOLangulo@DP$, $distME@MED_{CM}$, $xMD@DP_{VOL}$, $distMD@MED_{CM}$, $aX@MED$, $distCF@MED$, $aY@MED$, $aY@DP$, $fAngulo@DP$, $PFlevel@MED$, $PFlevel@DP$, $gZ@MED$, $gX@DP$, $gZ@DP$, $VOLangulo@MED$, $PAlevel@DP$, $rReta@MED$, $xMD@DP_{CM}$ e $gX@MED$

Nas Figuras 29 e 30, mostram-se os melhores $Findex_5$ obtidos em cada tamanho de conjunto de características selecionadas nas frequências de 140 e 280, respectivamente.

Como frequências candidatas para nCP (Curto Prazo), o menor período testado foi um dos escolhidos por permitir respostas mais rápidas, porém, com $Findex_5$ abaixo dos demais em 0,944. As características selecionadas para as frequências candidatas de 5 e 15 podem ser vistas na Tabela 12, onde, $@MED$ representa uma característica de média, um desvio padrão

Figura 29 – Gráfico para Monitoramento do Motorista e Qualidade na Direção dos melhores $Findex_5$ para o número de características selecionadas e em períodos de 140 frames.

Fonte: Elaborada pelo autor.

Figura 30 – Gráfico para Monitoramento do Motorista e Qualidade na Direção dos melhores $Findex_5$ conforme número de características selecionadas e utilizando períodos de 280 frames.

Fonte: Elaborada pelo autor.

foi marcado com @DP e as origens das características do monitoramento do motorista estão indicadas por VOL (volante) e CM (cabeça do motorista). O melhor conjunto entre os possíveis tamanho de 1 e 68 características, segundo o índice $Findex_5$, nas frequências de 140 e 280 são mostrados nas Figuras 31 e 32, respectivamente.

Classificadores foram obtidos usando os frames do NDBD com MLP ou *Multilayer Perceptron* (ver Seção 3.5.2) e Rprop (RIEDMILLER; BRAUN, 1992) para o treinamento da rede. As funções de ativação Gaussiana e Sigmoide foram usadas nos experimentos. Com o objetivo de maximizar o $Findex_5$ e a acurácia do sistema de classificação, foi utilizado Algoritmo

Tabela 12 – Todas as características escolhidas para as frequências candidatas a nCP para Monitoramento do Motorista e Qualidade na Direção.

Período	Qtde	Características
5	15	$xMD@MED_{CM}$, $zME@MED_{VOL}$, $PFlevel@MED$, $gY@MED$, $gY@DP$, $zMD@MED_{CM}$, $aX@MED$, $yMD@DP_{VOL}$, $rReta@DP$, $minDistMaos@DP_{CM}$, $gZ@MED$, $aX@DP$, $fAngulo@DP$, $zMD@MED_{VOL}$ e $xMD@DP_{CM}$
15	9	$PAlevel@DP$, $xME@MED_{CM}$, $yMD@MED_{VOL}$, $xME@MED_{VOL}$, $acel@MED$, $zMD@DP_{CM}$, $distME@MED_{CM}$, $distME@MED_{VOL}$ e $xMD@MED_{VOL}$

Figura 31 – Gráfico para Monitoramento do Motorista e Qualidade na Direção do melhor $Findex_5$ conforme número de características selecionadas e utilizando períodos de 5 frames.

Fonte: Elaborada pelo autor.

Genético⁵ (AG) (GOLDBERG, 1989) com codificação binária para localizar parâmetros do MLP. A equação da função objetivo (fitness) utilizada foi a seguinte: $fitness = \kappa Findex_5 + (1 - \kappa)A$, onde, A é a acurácia obtida pelo classificador e κ é a ponderação entre o $Findex_5$ e a acurácia da classificação. Neste trabalho, adotou-se κ de 0.6 para classificadores (*classificadorCP*) para frequências de Curto Prazo (5 e 15) e de 0.9 para Longo Prazo (140 e 280 frames) ou candidatos a *classificadorLP*. Esta diferenciação de κ , entre o *classificadorLP* e *classificadorCP*, se deu pelo fato de ser imprescindível que o $Findex_5$ seja priorizado (minimizar a ocorrência de falsos positivos para a detecção de risco na direção) para *classificadorLP* já que é este classificador responsável pela ativação do nível alto do alerta (maior certeza de resposta). Na obtenção de *classificadorCP* a ponderação entre o $Findex_5$ e a acurácia foi quase que equiparada, embora o $Findex_5$ permaneça o mais relevante. Na Tabela 13, são mostrados os limites de busca do AG para cada parâmetro e frequência, onde, α e β foram parâmetros das funções de ativação, CO conteve o número de neurônios na camada oculta da rede e *interac* foi usada para definir o

⁵ A biblioteca GALib versão 2.4.7 foi utilizada (está disponível em <http://lancet.mit.edu/ga>).

Figura 32 – Gráfico para Monitoramento do Motorista e Qualidade na Direção do melhor $Findex_5$ conforme número de características selecionadas e utilizando períodos de 15 frames.

Fonte: Elaborada pelo autor.

número máximo de iterações permitidas para o método de treinamento Rprop.

Tabela 13 – Limites do AG para os Parâmetros do MLP para Monitoramento do Motorista e Qualidade na Direção

Parâmetros	Frequências →	5	15	140	280	Bits
α	Mínimo	10^{-3}	10^{-3}	10^{-3}	10^{-3}	25
	Máximo	10	10	10	10	
	Casas Decimais	6	6	6	6	
β	Mínimo	10^{-3}	10^{-3}	10^{-3}	10^{-3}	25
	Máximo	10	10	10	10	
	Casas Decimais	6	6	6	6	
CO	Mínimo	20	21	35	43	8
	Máximo	30	31	45	53	
$interac$	Mínimo	100	100	100	100	8
	Máximo	500	500	500	500	
Tamanho do Código Genético						66

O AG foi parametrizado (definição empírica) usando 10 indivíduos e 400 gerações, o índice de *crossover* foi de 80%, o nível de mutação foi de 5% e foi utilizada seleção por torneio. A população inicial foi gerada randomicamente. O tamanho do código cromossômico adotado pelo AG foi de 66 *bits*. O mais elevado *fitness* (parâmetros individuais) escolhido em cada teste para cada classificador. AG foi empregado três vezes para cada teste. A rede de cada classificador tem o mesmo número de neurônios na camada de entrada que o número de características que foram usadas em cada teste, um neurônio foi usado na camada de saída e na camada oculta foram usados CO neurônios (AG obteve o número de neurônios nesta camada). Os resultados obtidos nos grupos de treinamento, validação e sonolência são mostrados na Figura 33 e na Tabela 14 os melhores parâmetros encontrados pelo AG. Pode-se perceber que todos os

classificadores são estatisticamente similares para $Findex_5$ e acurácia. Então, AG encontrou parâmetros estatisticamente similares para funções de ativação para as frequências testadas.

Figura 33 – Gráfico para Monitoramento do Motorista e Qualidade na Direção dos classificadores obtidos pelo AG usando os dados de treinamento (cores escuras), dados de validação (cores claras) e dados de sonolência (cores claras com um padrão branco), onde, a barra vermelha é do $Findex_5$, a verde é da acurácia, e os limites permitidos pelo desvio padrão nos grupos de *cross-validation* de treinamento são mostrados em cinza.

Fonte: Elaborada pelo autor.

Tabela 14 – Parâmetros para o Monitoramento do Motorista e Qualidade na Direção dos classificadores MLP providos pelo AG.

Classificador	α	β	CO	interac
Gaussiana5	0,002	1,241	30	441
Sigmoide5	3,484	4,209	30	479
Gaussiana15	0,030	2,144	21	347
Sigmoide15	3,861	2,672	31	472
Gaussiana140	0,878	0,680	44	372
Sigmoide140	2,769	0,863	38	462
Gaussiana280	0,014	5,509	53	393
Sigmoide280	1,482	1,060	50	394

Para definir o $npCP$ (número de predições anteriores) do subsistema de curto prazo experimentou-se todos os classificadores candidatos gerados para este fim, ou seja, frequências de 5 e 15 nas funções de ativação Gaussiana e Sigmoide, em todos os períodos obtidos dos dados provenientes do NDBD (treinamento, validação e sonolência). O resultado dos experimentos pode ser observado na Figura 34. Dentre os valores candidatos a $npCP$, entre 2 e 8, a partir de 5

o valor de $Findex_5$ começa a se estabilizar próximo ao máximo de cada classificador. Optou-se, portanto, por $npCP = 5$ neste trabalho por utilizar a menor dependência temporal (2,5 segundos para frequência 5 e 7,5 segundos para 15) para a obtenção da saída do subsistema de curto prazo e com reduzido número de falsa detecções de risco ($Findex_5$ elevado).

Figura 34 – Resultado para Monitoramento do Motorista e Qualidade na Direção dos classificadores de curto prazo para $npCP$ entre 2 e 8, onde, as linhas escuras indicam o $Findex_5$ obtido, as linhas claras a acurácia, vermelho é indicação de Gaussiana5, verde de Sigmoide5, azul de Gaussiana15 e amarelo de Sigmoide15.

Fonte: Elaborada pelo autor.

Utilizando-se $npCP = 5$ e combinando os 4 classificadores de curto prazo com os 4 classificadores de longo prazo, pôde-se obter 16 sistemas candidatos. As opções de curto prazo são Gaussiana5 (cg5), Sigmoide5 (cs5), Gaussiana15 (cg15) e Sigmoide15 (cs15). Já as opções de longo prazo são Gaussiana140 (lg140), Sigmoide140 (ls140), Gaussiana280 (lg280) e Sigmoide280 (ls280). A Figura 35 mostra os resultados obtidos com os sistemas candidatos usando os dados do NDBD, onde emprega-se “#” na descrição do sistema para separar o *classificadorCP* do *classificadorLP*.

O *classificadorCP* determina o $Findex_5$ (este índice é somente para situações normais, ou seja, sem risco), assim, o subsistema de curto prazo é responsável por indicar se a situação atual é de “risco” ou “normal” e o subsistema de longo prazo indica o nível de alarme (nível de risco). Dois candidatos a *classificadorCP* alcançaram os melhores $Findex_5$ (veja linha vermelha na Figura 35), o cg15 atingiu 0,946 e foi obtido 0,942 para o cg5. A cobertura para cg15 e cg5 para situações normais foi de 0,99 e 0,98, respectiva. Essa cobertura próxima ao máximo resulta

Figura 35 – Resultados para 16 sistemas de detecção de distrações e perturbações do motorista usando características do Monitoramento do Motorista e da Qualidade na Direção, onde, a linha vermelha indica o $Findex_5$, a linha azul é a acurácia do sistema de alerta, a verde indica acurácia para detecção de risco, a linha amarela indica o acerto para “sem alerta” (direção normal), marrom é o acerto do “alerta baixo” e roxo indica o acerto para o “alerta alto”.

Fonte: Elaborada pelo autor.

em 1% para cg15 e 2% cg5 de predições de situações normais erradas (falsos positivos) em todos os *frames* do NDBD.

O NDBD possui três conjuntos dados principais (ver Seção 5.2): teste de direção normal (sem qualquer distração ou perturbação), direção falando ao celular e direção embriagada. Nas Figuras 36, 37, 38 e 39 mostram-se os acertos dos alertas nos sistemas que utilizam cg5 e cg15, individualizados para os diferentes tipos de testes do NDBD, empregando como fonte de dados: todos os dados do NDBD, somente o conjunto de treinamento, com o conjunto de validação e para o conjunto de teste da sonolência, respectivamente.

Os conjuntos de treinamento e validação foram adquiridos utilizando o mesmo procedimento de coleta de dados, diferindo-se apenas, por conterem dados de pessoas diferentes. O sistema apresentou acertos similares para os dois conjuntos, especialmente para o teste de direção normal ao qual atingindo acertos entre 98 e 99%, acertos entre 78 e 89% para o teste do uso do celular e entre 33 e 57% para o teste da embriaguez. O alto acerto para direção normal demonstra que o uso do AG e o $Findex_5$ possibilitaram a obtenção de classificadores especializados na detecção da postura e do modo de direção dos motoristas avaliados sob boas condições de dirigibilidade (sem distrações ou perturbações aparentes). No teste de direção falando ao celular

Figura 36 – Acerto dos alertas em todo o NDBD para Monitoramento do Motorista e Qualidade na Direção nos 8 sistemas utilizando cg5 e cg15 para o teste da direção normal (verde), teste do uso do celular (vermelho), teste de embriaguez (amarelo) e acurácia geral (azul).

Fonte: Elaborada pelo autor.

Figura 37 – Acerto dos alertas para o conjunto de treinamento do Monitoramento do Motorista e Qualidade na Direção nos 8 sistemas utilizando cg5 e cg15 para o teste da direção normal (verde), teste do uso do celular (vermelho), teste de embriaguez (amarelo) e acurácia geral (azul).

Fonte: Elaborada pelo autor.

foram efetuadas várias ligações telefônicas ao motorista durante a coleta de dados, portanto, há alguns *frames* rotulados como “normal” neste teste (entre as ligações), o que ajudou a explicar a acurácia mais elevada do que o teste de embriaguez. Outra vantagem para a detecção do uso do celular é o fato de ser mais perceptível no curto prazo, gestos rápidos podem ser captura-

Figura 38 – Acerto dos alertas para o conjunto de validação do Monitoramento do Motorista e Qualidade na Direção nos 8 sistemas utilizando cg5 e cg15 para o teste da direção normal (verde), teste do uso do celular (vermelho), teste de embriaguez (amarelo) e acurácia geral (azul).

Fonte: Elaborada pelo autor.

Figura 39 – Acerto dos alertas para a sonolência no Monitoramento do Motorista e na Qualidade na Direção nos 8 sistemas utilizando cg5 e cg15 para o teste da direção normal (verde), teste do uso do celular (vermelho), teste de embriaguez (amarelo) e acurácia geral (azul).

Fonte: Elaborada pelo autor.

dos em um curto espaço de tempo (como pode ser visto pela linha vermelha na Figura 40), enquanto na embriaguez o monitoramento do motorista tem relevância menor (motorista ao estar embriagado não altera relevantemente o modo como utiliza suas mãos). A proximidade dos resultados obtidos entre o grupo de treinamento e de validação indica que não houve um sobre treinamento (*overtraining*) dos classificadores e a boa capacidade de generalização existente nos classificadores.

O conjunto de sonolência do NDBD foi também um dos conjuntos de validação do sistema, ou seja, nenhum de seus dados foi empregado no treinamento dos classificadores. Os motoristas participantes ficaram mais de 36 horas sem dormir antes das coletas dos dados. Sendo realizadas as coletas em direção normal, falando ao celular e embriagado, onde, todos os dados coletados em um motorista sonolento considerados de risco (incluindo a direção normal). Os acertos para o teste embriagado foram entre 16 e 45%, mais baixos que o conjunto de validação e treinamento. O teste do uso do celular também apresentou piora em relação aos outros conjuntos de dados, atingindo valores entre 56 e 65%. Mas, no teste de direção normal, pode-se perceber a informação mais relevante, o acerto do sistema variou entre 1 e 9%, assim, podemos perceber que os classificadores treinados nas características escolhidas, não foram capazes de perceber significativas diferenças entre a direção de uma pessoa sonolenta de uma não sonolenta. As detecções mais elevadas observadas nos testes do uso do celular e da embriaguez se dão pelo uso do celular e pela utilização dos óculos de simulação de embriaguez.

Os sistemas *cg5#ls140* e *cg15#ls140* obtiveram a maior acurácia de 66% na detecção do nível de alerta em todos os dados providos pelo NDBD. Na Figura 40, mostram-se a acurácia geral e outras acurácias de alertas para o sistema *cg5#ls140* nos conjuntos de treinamento e validação do NDBD (sem dados de sonolência). Sem os dados de sonolência a acurácia geral do sistema para os alertas sobe para 79,6%. Não há alertas baixo ou altos para os dados do NDBD Normal, porque não há rotulação de risco nesses quadros. Não há rotulação de qualquer quadro “normal” para Embriaguez, por essa razão não há acurácia “sem alerta” para esse tipo de dado. Outro detalhe observado na Figura 40 é da baixa acurácia do alerta baixo, alcançando 39% para Celular e 26,6% para Embriaguez, o que mostra que a classificação de *classificadorCP* foi um dos principais motivos da acurácia geral do sistema não ter sido mais elevada.

Apesar das similaridades dos resultados dos sistemas, o *cg5#ls140* é a melhor opção por possuir uma vantagem no subsistema de curto prazo, ele é capaz de indicar um “alerta baixo” (risco) ao motorista utilizando 2,5 segundos de dados, ou seja, $\frac{nCP \times npCP}{frequenciaNDBD}$ (a frequência do NDBD é de 10 Hz). Sendo assim, os parâmetros mais vantajosos obtidos nestes experimentos foram: *classificadorCP* = *cg5*, *nCP* = 5, *npCP* = 5, *classificadorLP* = *ls140* e *nLP* = 140. O acionamento do “alerta alto” utilizou 14 segundos de dados ($\frac{nLP}{frequenciaNDBD}$). Todos os *frames* do NDBD⁶ e alguns exemplos de *frames* de saída são mostrados na Figura 41, onde, as últimas 5 predições de *frames* com risco são mostradas acima e a direita a indicação da saída do alerta é mostrada no centro-esquerdo da imagem de saída (o texto em preto indica nenhum nível de alerta, amarelo indica nível baixo e vermelho nível alto do alerta).

Os principais problemas de classificação⁷ do sistema *cg5#ls140* (momentos em que a rotulação efetuada de risco difere da classificação atribuída pelo sistema) são mostrados na

⁶ Todos os vídeos resultantes do processamento do NDBD estão disponíveis no link <https://goo.gl/yEexGd>.

⁷ Todos os problemas do processamento do NDBD pelo *cg5#ls140* estão disponíveis no link <https://goo.gl/aHZo11>.

Figura 40 – Acerto dos alertas do sistema cg5#ls140 agrupando os dados dos grupos de treinamento e validação nos diferentes conjuntos de dados do NDBD, sendo acerto para “sem alerta” indicado pela barra verde, alerta baixo indicado pela vermelha, alerta alto mostrado pela barra amarela e a acurácia geral dos alertas no grupo de treinamento apresentada pela barra azul.

Fonte: Elaborada pelo autor.

Figura 41 – Exemplo de saída do Sistema de Detecção de Distrações e Perturbações.

Fonte: Elaborada pelo autor.

Figura 42. O primeiro deles é um problema bom, o sistema detecta risco quando uma das mãos do motorista permanece próxima a sua face/cabeça por algum tempo, como pode ser visto na Figura 42a. As razões que levam o motorista retirar a mão do volante e a colocar próxima de sua cabeça podem ser as mais diversas, dentre elas, a razão mais comum observada é o ato de coçar a cabeça. Este problema pode ser dito “bom” pelo fato de permitir alertar (reeducar) o motorista

a permanecer com suas duas mãos no volante, algo que o ajude a reagir melhor na condução do veículo. Por outro lado, quando a mão está a uma distância similar da cabeça e do volante, como é mostrado pela Figura 42b, o sistema detecta como uma postura aceitável (sem risco) para o motorista.

Figura 42 – Exemplo de problemas de predição de risco para o cg5#ls140, onde o círculo azul indica a posição da mão direita e o vermelho a posição da mão esquerda.

Fonte: Elaborada pelo autor.

Outra característica apresentado pelo sistema provém de problemas da esqueletização e pode ser observado um exemplo na Figura 42c. Infelizmente uma esqueletização errada pode indicar a posição das mãos em um local equivocado, como por exemplo, as duas mãos no volante enquanto o motorista utiliza uma das mãos para falar ao celular. No entanto, nem sempre problemas de detecção das partes do corpo do motorista são problemas da esqueletização por si só, algumas vezes as mãos do motorista podem estar em algum local com oclusão das mãos, por

exemplo, atrás do volante como mostrado pela Figura 42d. Mesmo assim, a esqueletização infere a localização das mãos, sendo inequivocadamente a tarefa de rastrear as mãos nessa situação uma tarefa bem mais difícil e com um grau de confiabilidade menor do que em situação de mão visível. Porém, como é mostrado na Figura 42e, o modelo de predição de distrações não adota apenas o posicionamento das mãos para detectar ou não risco, várias características fazem parte da averiguação e há alguns casos em que as mãos do motorista são declaradas como posição “normal” mesmo estando segurando o celular junto a cabeça. O classificador aprendeu também que próximo ao câmbio do veículo (troca de marchas manual) não é uma posição de risco, assim manter e operar o celular próximo dessa região passa despercebido pelo sistema, desde que, o modo como o motorista conduz o veículo permaneça “normal”, como pode ser observado na Figura 42f. Até mesmo o ato de permanecer com as mãos no câmbio não é uma situação de risco para o sistema, por conta das muitas trocas de marchas observadas durante o treinamento, essa pose é observada como normal (Figura 42g).

Um motorista bêbado pode vir a ser considerado um motorista em condições normais caso consiga se manter dirigindo em uma linha reta (mantendo o esterçamento do volante inalterado), como pode ser observado na Figura 42h. Os testes do NDBD não são realizados com os motoristas realmente embriagados, apenas a visão do motorista está alterada e similar a um embriagado. Em outras palavras, o fato de sua mente não estar entorpecida nos leva a concluir que um motorista embriagado, de fato, teria maiores dificuldades de se manter dirigindo similarmente a um motorista normal. Outro fator que auxilia o motorista embriagado a não ser detectado é dirigir em baixa velocidade, pois, quanto maior a velocidade maior é a intensidade de movimentação observada pelo acelerômetro do veículo para uma mesma mudança do esterçamento do volante. De maneira semelhante a embriaguez, o celular mantido ao ouvido sem o auxílio de uma das mãos é passado despercebido pelo sistema caso o motorista permaneça dirigindo o veículo de maneira normal, como pode ser observado na Figura 42i.

O sistema é capaz ainda de perceber manobras não usuais do motorista, por exemplo, zigzagues em alta velocidade são detectados como risco mesmo para um motorista em perfeitas condições como é mostrado na Figura 42j. Dirigir um veículo entre as faixas de rodagem por um longo período, como pode ser observado na Figura 42k, é outro motivo para um motorista “normal” ser classificado como distraído ou perturbado, o que de fato é uma atitude temerária.

Duas principais características puderam ser notadas na experiência virtual de direção embriagada do NDBD. Ao utilizar os óculos *Drunk Busters*, os motoristas movem suas cabeças de maneira a encontrar uma melhor posição para ver a rodovia, como pode ser visto na Figura 43a. Uma segunda característica observada nos testes de embriaguez é o fato dos motoristas tentarem se basear nas marcas da rodovia como pontos de referência para conseguir conduzir melhor o seu veículo, como pode ser observado na Figura 43b.

Figura 43 – Características da experiência virtual de embriaguez.

Fonte: Elaborada pelo autor.

6.4 Considerações

Neste capítulo, foram apresentados os experimentos das propostas dos sistemas de detecção de momentos de risco do condutor de veículos por distração ou perturbação (ver Capítulo 4): o Sistema de Curto e Longo Prazos (SCLP) e o Sistema Bufferizado (SB).

O SB foi experimentado somente utilizando características focadas na qualidade de direção e validado utilizando a embriaguez. Esse sistema só analisa os dados em uma frequência de períodos, impossibilitando avisos sonoros rápidos para que o motorista volte a uma condução segura (reeducação). No entanto, a possibilidade de reeducação não é um problema para um motorista perturbado pela embriaguez já que a sua volta a um estado sóbrio carece de um tempo para que o álcool ingerido seja processado pelo organismo do motorista. Portanto, o sistema deve, ao perceber um motorista embriagado, acionar algum outro sistema capaz de levar o veículo a um lugar seguro (estacionar) e, mantendo assim, a rua e o motorista em segurança. A acurácia para o SB usando uma rede MLP com função de ativação Sigmoide e períodos com 220 *frames* foi de 98% para os dados do NDBD. A possibilidade de um sistema tomar o controle do veículo de maneira equivocada foi detectada em cerca de 1,5%, já que esse foi o percentual de erros de classificação do sistema para um usuário sóbrio.

O SCLP possui duas frequências de análise dos dados, de curto (CP) e longo (LP) prazos, e foi utilizado em dois conjuntos de experimentos. O primeiro deles foi o SCLP para detecção do uso de celular pelo motorista enquanto dirige o veículo. Esse experimento utilizou características focadas no Monitoramento do Motorista, mas especificamente nas posições das mãos do motorista. A grande vantagem do SCLP para o SB é que o SCLP permite a reeducação do motorista por possuir três níveis de alerta: sem alerta, alerta baixo e alerta alto. O alerta baixo ocorre quando apenas no CP é detectado risco na direção. Sendo reportado pelo sistema um alerta baixo, o motorista poderia ser avisado por um aviso sonoro para que volte a sua concentração exclusivamente ao ato de dirigir, reeducando-o. Caso o motorista persista em manter-se em risco e um alerta alto seja observado pelo SCLP, um sistema autônomo poderia então assumir o controle do veículo desempenhando alguma ação defensiva interessante como encostar o veículo no acostamento da rodovia. A acurácia dos alertas do SCLP nos dados do NDBD foi de 95%,

usando períodos de 5 *frames* para CP e uma janela de 210 *frames* para LP, sendo em ambos os casos, utilizados classificadores MLP com função de ativação sigmoidal. A incidência de falsas detecções de risco do sistema não foi superior a 1% dos dados do NDBD rotulados como sem risco, o que nos mostra que um possível usuário do sistema não seria muito incomodado com alertas indevidos.

Com o SCLP buscou-se ainda, utilizar características do Monitoramento do Motorista em conjunto com as da Qualidade de Direção, ou seja, unindo dados coletados da postura do motorista ao resultado efetivo da condução do veículo para identificar todas as perturbações e distrações objetivadas neste trabalho, sendo elas, o uso de celular, embriaguez e sonolência. O sistema é capaz também de reeducar o motorista, por ser um SCLP, possui os três níveis de alertas do motorista. A acurácia dos alertas do SCLP em todos os dados do NDBD foram de 66% usando períodos de 5 *frames* e MLP/Gaussiano para CP e MLP/Sigmoidal em janelas de dados com 140 *frames* para LP. O sistema emite 2% de falsas de detecções de risco nos dados do NDBD. Quando não são utilizados os dados de sonolência do NDBD o sistema atinge a acurácia de 79,6%. A sonolência foi verificada só a título de validação do sistema, ao qual, não se demonstrou muito acurada quando foram usados estes seus dados. Houveram dois motivos principais por se optar pela utilização da sonolência somente para validação. O primeiro deles é o de verificar a similaridade entre o uso do celular e a embriaguez relacionando-os a sonolência, o que não se confirmou para as características utilizadas, ou seja, a sonolência necessita de um treinamento específico para a obtenção de níveis de acertos próximos as outras distrações experimentadas. O segundo motivo foi a quantidade de dados reduzida para sonolência provida pelo NDBD.

Todos os classificadores de todos os sistemas utilizados nos experimentos obtiveram uma boa capacidade de generalização em dados não treinados (acurácia da validação próxima a de treinamento), assim, pode-se concluir que, com a otimização dos parâmetros feitas pelo AG e o treinamento usando o método de validação cruzada, foi possível evitar o treinamento excessivo. A função objetivo do AG na busca por melhores parâmetros de treinamentos empregou-se uma ponderação entre o $Findex_5$ e a acurácia final do classificador. $Findex_5$ por si só é uma ponderação onde o acerto do sistema para uma situação normal (sem risco) é priorizado em detrimento do acerto para uma situação de risco. Em outras palavras, o $Findex_5$ permitiu ao AG encontrar classificadores que predizem menos situações rotuladas em “normal” como “risco”. Assim, pôde-se aumentar o conforto da utilização dos sistemas em um veículo real, pela menor probabilidade de ocorrência de falsas detecções de risco. Porém, o $Findex_5$ sozinho não garante uma acurácia elevada do sistema, inclusive ele pode resultar em um classificador com predição de todos os quadros como “sem risco” (resultando em um $Findex_5$ próximo ao valor máximo, 1). Por essa razão utilizou-se também a acurácia na função objetivo do AG, permitindo que a acurácia do sistema também fosse um fator determinante na escolha do melhor classificador. O fator de ponderação entre o $Findex_5$ e a acurácia não foi sempre o mesmo em todos os experimentos. Para o SCLP, a acurácia recebe uma importância maior na classificação de curto prazo já que seu

sinal é utilizado somente na reeducação (avisos) ao motorista, já em classificadores de longo prazo, o fator de ponderação da acurácia é bem mais reduzido objetivando classificadores com uma menor taxa de falsas detecções de risco.

No que tange ao comportamento do motorista, os SCLP podem auxiliar o motorista a manter uma boa pose baseados na posição das mãos e sua distância para o volante e a cabeça do motorista (monitoramento do motorista). Foi demonstrado que utilizando um sensor de captura 3D (neste caso o Kinect v2) os sistemas foram capazes de perceber comportamentos perigosos e boas práticas do motorista. Somente utilizando o NDBD observou-se algumas atitudes de risco do motorista: mão(s) no rosto, mão(s) permanecendo entre a cabeça e o volante, mão(s) fora do campo de visão do sensor e manter a mão por um longo período no câmbio do veículo. Todos os classificadores do monitoramento do motorista aprenderam que manter as suas duas mãos no volante é uma boa atitude. Embora somente a detecção do uso do celular no trânsito tenha sido o foco da utilização do monitoramento do motorista, um sistema SCLP poderia ajudar a um motorista, mesmo que sem a utilização do celular, a permanecer em pose adequada por mais tempo. Percebeu-se, ainda, maneiras de se burlar a correta classificação do sistema, como por exemplo, deixar o celular no ouvido sem o auxílio das mãos. Outra contribuição interessante deste trabalho no monitoramento do motorista é o fato de se aplicar uma ponderação das características em relação ao ombro dos motoristas. Com os dados do NDBD e a dimensão dos ombros obtidas em todos os *frames*, foi possível na experimentação determinar um fator ϕ que normalize as característica entre -1 e 1 , levando-se em conta o tamanho relativo de cada motorista. Assim, dados provenientes de motoristas de dimensões diversas puderam ser utilizados no treinamento e validação (foram utilizados motoristas distintos na validação) detectando situações de risco e normais de maneira similar em cada um deles.

Algumas observações comportamentais sobre o modo de condução do veículo também puderam ser observadas nos sistemas SB e SCLP. Um classificador baseado na embriaguez precisou modelar em seu treinamento, por exemplo, entre os comportamentos problemáticos de permanecer por um longo período sem efetuar correções na direção (em linha reta, por exemplo) ao de efetuar ziguezagues excessivos na pista. Os dois comportamentos podem indicar risco ao motorista, porém os sistemas tiveram maiores dificuldades de observar um motorista embriagado (em risco) quando estes conseguiram manter o esterçamento do volante inalterado. Muito por conta de que os motoristas não estavam realmente embriagados e sim em uma experiência virtual de embriaguez (somente visual), portanto suas faculdades mentais estavam intactas. Outro comportamento perigoso observado pelos sistemas classificadores foi o de dirigir o veículo entre as faixas de rodagem. Os classificadores aprenderam a utilizar as faixas de rodagem como meio de mensurar a qualidade de direção, algo que poderia ser aplicado à outras distrações ou perturbações que alteram o modo usual de direção do veículo.

Analisando os três grupos de experimentos, pôde-se verificar um decréscimo significativo na acurácia dos sistemas ao buscar unir os dados de Monitoramento do Motorista com os de

Qualidade da Direção na mesma classificação. Uma parte deste problema é causado pela ruptura entre o que a pose (monitoramento do motorista) reflete no modo como o motorista dirige o veículo e uma relação inversa nem sempre é observada também. Em outras palavras, o ato de um motorista estar embriagado não altera, necessariamente, a sua postura em frente ao volante, ele pode manter suas mãos ao volante, por exemplo. Todas as características do monitoramento do motorista tem uma relação forte entre si, enquanto o mesmo ocorre na qualidade da direção. Ao se unir os dois ambientes, dificulta-se a constituição de um modelo de classificação de situações entre risco e sem risco. Um caminho futuro interessante seria criação de uma abordagem híbrida de detecção de distrações e perturbações, empregando-se classificadores especialistas em cada uma das fontes de dados, unindo-se então suas respostas. Outra ideia seria a de adotar técnicas de análise temporal classificatórias, como por exemplo deep-learning com recorrência, necessitando no entanto de uma ampliação dos dados de treinamento e validação utilizados (NDBD). Assim, espera-se que o alto grau de acurácia dos sistemas em separados sejam mantidos ou melhorados também para um sistema que una os dois focos de dados.

Um ponto importante que pode permitir uma acurácia maior do sistema seria a de obter uma base de dados maior. A ampliação do NDBD poderia permitir a utilização de redes profundas (*deep-learning*) permitindo assim a melhora na acurácia final do sistema. Um caminho importante seria obter dados de direção em outros cenários de direção, como por exemplo, em cidades. Todos os dados do NDBD foram obtidos em um ambiente rodoviário e este novo ambiente poderia, além de contribuir com novos dados, dar uma maior robustez aos sistemas validando o funcionamento para estes outros ambientes.

Além de se ampliar o NDBD, seria interessante utilizar diferentes níveis de embriaguez nas coletas com simulação da visão embriagada, assim poderíamos observar motoristas poucos embriagados e motoristas muito embriagados, por exemplo. O desenvolvimento de um sistema capaz de detectar a embriaguez, mesmo em situações de pouca embriaguez, é interessante já que em alguns países a embriaguez é tolerada até certos níveis, porém em outros, nenhum nível é aceito.

CONSIDERAÇÕES FINAIS

Neste capítulo são apresentadas as conclusões, contribuições e propostas para trabalhos futuros.

7.1 Conclusões

A identificação de distrações e perturbações do motorista na condução de veículos é uma tarefa essencial quando se objetiva diminuir o número de acidentes por erros humanos. Em veículos autônomos, esse sistema possibilita ao veículo identificar requisições de controle manual por motoristas incapacitados na condução de veículos. Este trabalho investigou e desenvolveu técnicas de detecção de motoristas distraídos ou perturbados, principalmente, por motivo de uso de celular no trânsito, embriaguez ou sonolência. Este trabalho apresentou, portanto, uma solução original ao qual permite a detecção de momentos de direção perigosa de um condutor de um veículo utilizando-se de dados provenientes do monitoramento e da qualidade de direção do motorista. Outras soluções intermediárias para detecção de embriaguez e do uso do celular foram criadas também e unidas as ideias em uma solução resultante deste trabalho.

A solução final desenvolvida utiliza dois níveis de análise, os subsistemas de Curto Prazo e de Longo Prazo, assim, pode-se detectar os momentos de risco aos quais motoristas se submetem e emitir diferentes níveis de alerta como saída para o sistema. O sistema tem 3 níveis de alerta: “Sem alerta”, “Alerta baixo” e “Alerta alto”. O subsistema de curto prazo é responsável pela detecção de situações de risco ou não risco, em outras palavras, entre “Sem alerta” e algum nível de alerta. As últimas 5 predições de curto prazo (2,5 s) são empregadas na identificação dos riscos para reduzir o número de falsos negativos (risco) e falsos positivos (não risco). Menos de 2% dos *frames* rotulados no conjunto de dados do NDBD como sem risco obteve uma predição equivocada. Assim, foi possível melhorar o conforto do motorista ao utilizar o sistema, conseguindo evitar as situações desagradáveis de falsos alertas de riscos emitidos. O subsistema de longo prazo é responsável por determinar o nível do alerta de risco: baixo ou alto.

O alerta alto é pensado para ser um alerta silencioso, podendo ser utilizado para habilitar um sistema autônomo do veículo para assumir o controle do mesmo e efetuar manobras defensivas de segurança, como por exemplo, estacionar o veículo em algum local seguro. Por outro lado, o alerta baixo é um aviso ao motorista de que a sua postura ou modo como conduz o veículo não estão satisfatórios, antes que o sistema autônomo assuma o controle do veículo, permitindo ao condutor retornar a sua atenção para a tarefa de dirigir o veículo (boa pose para direção e boas práticas na condução). Este sistema pode ser uma ferramenta educacional, condicionando o motorista a evitar distrações e perturbações enquanto dirige. O experimento cg5#ls140 (veja Capítulo 6) foi o que alcançou os melhores resultados, atingindo 79,5% de acurácia para os alertas (sendo 98,5% para motorista “normal”) no conjunto de dados de treinamento e validação do NDBD (embriaguez e detecção do uso de celular, apenas). Incluindo-se os dados provenientes da seção de teste de sonolência do NDBD, a acurácia geral para o sistema de alertas foi de 66%, assim, pode-se perceber o sistema não foi acuras para detecção de sonolência, ou seja, não foi capaz de perceber significativas diferenças entre a direção de uma pessoa sonolenta de uma não sonolenta. Um dos motivos que levou a este baixo acerto é o fato da sonolência ter pertencido ao conjunto de validação do sistema apenas, ou seja, os classificadores não foram treinados em seus dados. Considerando a detecção individualizada dos problemas de condução, no caso específico da embriaguez (usados dados de embriaguez e direção normal) o sistema obteve 98% de acurácia, e especificamente para o uso de celular obteve 95% de acurácia.

O sistema proposto pode ser usado para detecção de quando o motorista está com comportamento e/ou direção perigosos. Então, pode-se detectar outros tipos de distrações ou desorientações em veículos providos de sensores e dados constantes na Seção 5.2. o monitoramento da pose e comportamento da direção do motorista são um importante caminho para melhorar a segurança das estradas, já que nem todos os veículos serão completamente autônomos nos próximos anos devido ao alto custos destes sistemas.

Na Tabela 15, os trabalhos relacionados (Seção 2.2) são comparados a este trabalho. O sistema proposto evita alguns pontos restritivos na utilização dos trabalhos relacionados, como por exemplo, impedir passageiros ao lado do motorista como no caso de [Veeraraghavan et al. \(2007\)](#), onde, a restrição é ocasionada pela posição da câmera focada no monitoramento do motorista. Em algumas soluções ([CIDADES, 2012](#); [YANG et al., 2011](#); [PARK et al., 2018](#); [DESHMUKH; DEHZANGI, 2017](#); [DAI et al., 2010](#); [HAILE, 1992](#); [DKHIL et al., 2015](#); [EBRAHIM et al., 2014](#); [JOHNSON; TRIVEDI, 2011](#); [BERGASA et al., 2014](#)) precisa-se do consentimento do motorista para operar, ou seja, o motorista necessita participar de alguma forma para permitir seu monitoramento, como por exemplo, instalar eletrodos no seu corpo ou deixar operante um programa em seu próprio celular. Outros trabalhos ([YANG et al., 2011](#); [VEERARAGHAVAN et al., 2007](#); [BERRI et al., 2016](#); [CRAYE; KARRAY, 2015](#); [SESHADRI et al., 2015](#); [AKROUT; MAHDI, 2013](#); [LENSKIY; LEE, 2012](#); [KUMAR et al., 2012](#); [BERRI et al., 2013](#); [LEE et al., 2006](#); [JOHNSON; TRIVEDI, 2011](#); [BERGASA et al., 2014](#)) necessitam de iluminação para operar por utilizarem algum tipo de sensor passivo. Dois trabalhos não

permitem a possibilidade de reeducação do motorista distraído enquanto dirige (ARTAN *et al.*, 2014; SESHADRI *et al.*, 2015) por serem baseados em câmeras instaladas na rodovia, sendo desenvolvido com o objetivo de aplicação de punições posteriores (multas de trânsito). Não garantem que o motorista está sendo avaliado pelo sistema os trabalhos de Carroll, Bellehumeur e Carroll (2013), Cidades (2012) e Deshmukh e Dehzangi (2017), onde, um passageiro pode estar sendo avaliado no lugar do motorista, ou seja, motoristas que deveriam ser impossibilitados de dirigir podem conduzir veículos sem serem incomodados. Grande tempo ao volante pode prejudicar a avaliação do motorista pelo sistema de Murata *et al.* (2011). A utilização de modelos personalizados (SHIRAZI; RAD, 2014), um para cada motorista, é outro ponto restritivo já que obriga a criação de um modelo antes de permitir que um motorista dirija o veículo, modelos generalistas são mais viáveis e, depois de constituídos, funcionam para qualquer motorista.

A atuação do sistema em ambiente real ainda passa por alguns desafios, como por exemplo, o teste de resistência a trepidação (“quase-estático”) por parte do Kinect. A solução de esqueletização utiliza apenas um quadro para a obtenção das partes do motorista, por essa razão, pode operar mesmo com algum nível de vibração.

7.2 Contribuições Científicas

As principais contribuições acadêmico-científicas deste trabalho são:

- Proposta e desenvolvimento de algoritmos para a obtenção de características relevantes ao monitoramento do motorista pelo uso de informações espaciais (3D);
- A normalização utilizando as dimensões dos ombros (φ) do motorista pode contribuir para a invariância de outros sistemas de parâmetros;
- A utilização do *Findex*₅ para avaliar classificadores com o objetivo de diminuir a incidência de falsos alertas (falsas predições de risco) para sistemas de reconhecimento de padrões;
- SVM Linear foi empregado para obter o grupo de características relevantes para o problema tratado pelo sistema proposto;
- O uso do Algoritmo Genético para encontrar bons parâmetros de treinamento (estatisticamente similar) para todos os classificadores testados neste trabalho. A função objetivo (avaliação da melhor solução) do Algoritmo Genético empregou ainda uma ponderação entre *Findex*₅ e a acurácia resultante do treinamento dos classificadores;
- Aperfeiçoamento de técnicas para o monitoramento do motorista e a qualificação da direção do veículo, unindo informações espaciais extraídas de dentro da cabine do veículo do comportamento empregado pelo veículo. Deste trabalho resulta um sistema que identifica perturbações e distrações, porém, as técnicas empregadas poderão ser adaptadas para

mensurar o modo de condução do veículo por um motorista de maneira mais ampla, sugerindo a maneira ideal da postura e condução em determinadas situações;

- A criação de um simulador especialmente desenvolvido para a obtenção de dados sobre distrações e perturbações do motorista, poderá vir a ser relevante para outras aplicações futuras;
- A constituição de um conjunto de dados NBDB é uma importante contribuição pois permite teste e comparações da performance do sistema em tarefas relacionadas a análise natural do comportamento do motorista.

7.3 Trabalhos futuros

Como trabalhos futuros pode-se citar:

- Criação de uma arquitetura de sistema híbrido de detecção de distrações e perturbações;
- Adotar técnicas de análise temporal classificatórias, como por exemplo deep-learning com recorrência, como classificador dos dados;
- Ampliação dos dados de treinamento e validação utilizados (NBDB);
- Obter dados de motoristas embriagados em diferentes níveis de embriaguez;
- Novo algoritmo de esqueletização e rastreamento utilizado utilizando análise temporal (alguns *frames* consecutivos) e adaptando o resultado da esqueletização do motorista obtida a profundidade observada na cena;
- Novas características podem ser obtidas e empregadas na solução proposta para melhorar a detecção de perturbações e distrações do motorista (ex: orientação da cabeça);
- Explorar a atenção espacial e análise dinâmica da cena para obter melhores resultados (exemplo de técnicas candidatas: fluxo ótico, análise de densidade e variação da ocupação espacial);
- Aperfeiçoar os algoritmos de subtração de fundo em mapas de profundidade preparando-os para remoção de fundos “quase-estáticos”, ou seja, minimizando as implicações de vibrações do sensor ocasionadas pelo balanço do movimento do veículo;
- Implantação dos sistemas propostos em veículos reais.

7.4 Lista de Publicações do Autor Relacionadas a esta Tese

Foram publicados artigos e capítulos de livros baseados em experimentos e estudos pertinentes a este trabalho. A lista de publicações encontra-se abaixo:

- Berri, R. A.; Wolf, D.; Osorio, F. Telepresence robot with image-based face tracking and 3d perception with human gesture interface using kinect sensor. In: *Robotics: SBR-LARS Robotics Symposium and Robocontrol (SBR LARS Robocontrol), 2014 Joint Conference on*, 2014, p. 205–210.
- Berri, R. A.; Osorio, F. S. Simulação de Robôs Móveis e Articulados: Aplicações e Prática. In: *Anais do CSBC'2015 XXXIV Jornada de Atualização em Informática (JAI)*, cap. 6, Recife, PE: SBC, p. 273–322, 2015.
- Berri, R. A.; Wolf, D.; Osorio, F. From tele-operated robots to social robots with autonomous behaviors. In: *Robotics - Communications in Computer and Information Science (CCIS)*, Springer, 2015, p. 32-52.
- Berri, R.; Osório, F.; Parpinelli, R.; Silva, A. A hybrid vision system for detecting use of mobile phones while driving. In: *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, p. 4601-4610.
- Berri, R.; Osório, F. A 3D vision system for detecting use of mobile phones while driving. In: *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, p. 1-8.
- Berri, R.; Osório, F. A nonintrusive system for detecting drunk drivers in modern vehicles. In: *2018 Brazilian Conference on Intelligent Systems (BRACIS)*, 2018, p. 73–78.

Tabela 15 – Tabela de comparação entre os trabalhos relacionados e este trabalho.

Detecção	Embriaguez	X	X	X	X	X																					X
	Sonolência						X	X	X	X	X	X														X	X
	Celular ao ouvido													X	X	X	X	X	X	X	X	X				X	X
	Celular (viva voz)													X	X	X			X								X
	Direção insegura (geral)																						X	X	X	X	X
Abordagem	Segmentação de pele									X										X	X						
	Viola-Jones						X		X		X	X								X	X	X					
	Característica invariantes (SIFT, SURF, HOG, FV)									X							X	X									
	Transformada Circular de Hough								X	X																	
	Morfologia matemática										X	X															
	Mapeamento não linear entre duas séries temporais(DTW)																						X				
	Expressão facial																		X								
	Amplitude da boca										X											X					
	Comportamento dos olhos						X	X		X	X	X							X		X						
	Orientação da cabeça																		X		X						
	Detecção de movimento																				X	X	X	X	X	X	X
	Rastreamento da posição do veículo na pista de rodagem																				X		X	X	X	X	X
	Baseado em sensores de movimento/orientação			X												X							X	X	X	X	X
	Pose e esqueletização 3D															X			X								X
	Posição dos braços/mãos																			X		X					X
	Avaliação da direção				X	X	X																X	X	X	X	X
	Avaliação de sinais biológicos	X	X	X			X																				
	Utiliza dados da telemetria do veículo																									X	X
	Sons de alta frequência														X												
	Bloqueio do uso do celular por um determinado período													X													
Intrusiva							X	X											X								
Meio	Monitoramento do motorista	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X					X
	Qualidade da direção				X	X	X															X	X	X	X	X	
	Motorista é testado antes de ligar o veículo	X																									
Restrições	Impede passageiros ao lado do motorista																			X							
	Precisa do consentimento do motorista para operar				X	X		X	X					X	X	X				X			X	X			
	Precisa de iluminação (passiva)									X	X	X	X	X				X	X	X	X	X	X	X	X		
	Trabalha sem possibilidade de reeducação do motorista distraído																X	X									
	Não garante que o motorista está sendo avaliado	X												X						X							
	Grande tempo ao volante altera o resultado do sistema	X																									
	Utiliza modelos personalizados para cada motorista						X																				
Sensores	1T1P	1A#M1G				1C8E1C1C1C1C							1!1A1I1C1K6E1C1C2C1C1C1@1K			1G1B1!						1A1!1G1@1B1A1@1G			1A1G1@1B1A1@1G		

Legenda: X - Sim, C - Câmera (RGB), I - Câmera (IR), A - Acelerômetro, G - Giroscópio, @ - GPS, T - Transdérnico
 K - Kinect, M - Sensor de massa, B - Bússola, P - Medidor de Pulso, ! - Microfone, E - Eletrodos,
 V - Telemetria do veículo, # - vários.

REFERÊNCIAS

AKROUT, B.; MAHDI, W. A visual based approach for drowsiness detection. In: **Intelligent Vehicles Symposium (IV), 2013 IEEE**. [S.l.: s.n.], 2013. p. 1324–1329. ISSN 1931-0587. Citado nas páginas 39, 43, 44, 120 e 124.

AL-NAJI, A.; GIBSON, K.; LEE, S.-H.; CHAHL, J. Real time apnoea monitoring of children using the microsoft kinect sensor: A pilot study. **Sensors**, v. 17, n. 2, 2017. ISSN 1424-8220. Disponível em: <<http://www.mdpi.com/1424-8220/17/2/286>>. Citado nas páginas 19 e 52.

ALCANTARILLA, P.; BERGASA, L.; JIMENEZ, P.; PARRA, I.; LLORCA, D.; SOTELO, M.; MAYORAL, S. Automatic lightbeam controller for driver assistance. **Machine Vision and Applications**, Springer, v. 22, n. 5, p. 819–835, 2011. Citado na página 36.

ALLSTATE. **New Allstate Survey Shows Americans Think They Are Great Drivers - Habits Tell a Different Story**. 2011. Disponível em: <<http://factor-tech.com/feature/autonomous-cars-and-mans-future-the-road-ahead/>>. Citado na página 30.

AMDITIS, A.; BIMPAS, M.; THOMAIDIS, G.; TSOOGAS, M.; NETTO, M.; MAMMAR, S.; BEUTNER, A.; MÖHLER, N.; WIRTHGEN, T.; ZIPSER, S.; ETEMAD, A.; LIO, M. D.; CICILLONI, R. A situation-adaptive lane-keeping support system: Overview of the safelane approach. **Intelligent Transportation Systems, IEEE Transactions on**, v. 11, n. 3, p. 617–629, Sept 2010. ISSN 1524-9050. Citado na página 29.

ARAKI, S.; MATSUOKA, T.; YOKOYA, N.; TAKEMURA, H. Real-time tracking of multiple moving object contours in a moving camera image sequence. **IEICE TRANSACTIONS on Information and Systems**, The Institute of Electronics, Information and Communication Engineers, v. 83, n. 7, p. 1583–1591, 2000. Citado na página 58.

ARMINGOL, J. M.; ESCALERA, A. de la; HILARIO, C.; COLLADO, J. M.; CARRASCO, J. P.; FLORES, M. J.; PASTOR, J. M.; RODRÍGUEZ, F. J. Ivvi: Intelligent vehicle based on visual information. **Robotics and Autonomous Systems**, Elsevier, v. 55, n. 12, p. 904–916, 2007. Citado na página 35.

ARTAN, Y.; BULAN, O.; LOCE, R. P.; PAUL, P. Driver cell phone usage detection from hov/hot nir images. In: **Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on**. [S.l.: s.n.], 2014. p. 225–230. Citado nas páginas 42, 44, 46, 121 e 124.

BADGER, E. **5 confounding questions that hold the key to the future of driverless cars**. 2015. Disponível em: <<http://www.washingtonpost.com/blogs/wonkblog/wp/2015/01/15/5-confounding-questions-that-hold-the-key-to-the-future-of-driverless-cars/>>. Citado na página 31.

BAF, F. E.; BOUWMANS, T.; VACHON, B. A fuzzy approach for background subtraction. In: **Image Processing, 2008. ICIIP 2008. 15th IEEE International Conference on**. [S.l.: s.n.], 2008. p. 2648–2651. ISSN 1522-4880. Citado na página 55.

- BARNICH, O.; DROOGENBROECK, M. V. Vibe: A universal background subtraction algorithm for video sequences. **Image Processing, IEEE Transactions on**, v. 20, n. 6, p. 1709–1724, June 2011. ISSN 1057-7149. Citado na página [61](#).
- BAY, H.; ESS, A.; TUYTELAARS, T.; GOOL, L. V. Speeded-up robust features (surf). **Comput. Vis. Image Underst.**, Elsevier Science Inc., New York, NY, USA, v. 110, n. 3, p. 346–359, jun. 2008. ISSN 1077-3142. Citado na página [40](#).
- BERGASA, L. M.; ALMERIA, D.; ALMAZAN, J.; YEBES, J. J.; ARROYO, R. Drivesafe: An app for alerting inattentive drivers and scoring driving behaviors. In: **Intelligent Vehicles Symposium Proceedings, 2014 IEEE**. [S.l.: s.n.], 2014. p. 240–245. Citado nas páginas [42](#), [44](#), [45](#), [74](#), [120](#) e [124](#).
- BERRI, R.; OSÓRIO, F. A 3d vision system for detecting use of mobile phones while driving. In: **2018 International Joint Conference on Neural Networks (IJCNN)**. [S.l.: s.n.], 2018. p. 1–8. ISSN 2161-4407. Citado na página [96](#).
- _____. A nonintrusive system for detecting drunk drivers in modern vehicles. In: **2018 Brazilian Conference on Intelligent Systems**. [S.l.: s.n.], 2018. p. 1–6. Citado na página [92](#).
- BERRI, R.; OSÓRIO, F.; PARPINELLI, R.; SILVA, A. A hybrid vision system for detecting use of mobile phones while driving. In: **2016 International Joint Conference on Neural Networks (IJCNN)**. [S.l.: s.n.], 2016. p. 4601–4610. ISSN 2161-4407. Citado nas páginas [41](#), [43](#), [44](#), [45](#), [120](#) e [124](#).
- BERRI, R. A.; SILVA, A. G.; ARTHUR, R.; GIRARDI, E. Detecção automática de sonolência em condutores de veículos utilizando imagens amplas e de baixa resolução. **Computer on the Beach**, p. 21–30, 2013. Citado nas páginas [40](#), [43](#), [44](#), [120](#) e [124](#).
- BERRI, R. A.; SILVA, A. G.; PARPINELLI, R. S.; GIRARDI, E.; ARTHUR, R. A pattern recognition system for detecting use of mobile phones while driving. In: **In Proceedings of the 9th International Conference on Computer Vision Theory and Applications**. [S.l.]: SCITEPRESS Digital Library, 2014. (Conference on Computer Vision Theory and Applications), p. 411–418. Citado na página [45](#).
- BISHOP, C. M.; NASRABADI, N. M. **Pattern recognition and machine learning**. New York, NY 10013, USA: Springer Science+Business Media, LLC, 2006. Citado na página [65](#).
- BLANA, E. A survey of driving research simulators around the world. Institute of Transport Studies, University of Leeds, 1996. Citado na página [81](#).
- BLOISI, D. Background Modeling and Foreground Detection for Maritime Video Surveillance. In: BOUWMANS, T.; PORIKLI, F.; HÖFERLIN, B.; VACAVANT, A. (Ed.). **Background Modeling and Foreground Detection for Video Surveillance**. Boca Raton, USA: CRC Press, 2014. cap. 18. ISBN 9781482205374. Citado nas páginas [59](#) e [60](#).
- BO, L.; LAI, K.; REN, X.; FOX, D. Object recognition with hierarchical kernel descriptors. In: **Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on**. [S.l.: s.n.], 2011. p. 1729–1736. ISSN 1063-6919. Citado na página [64](#).
- BONI, R. D.; VASCONCELLOS, M. T. L.; HOLMER, B. P.; ROBIN, R.; BASTOS, F. I.; PECHANISKY, F. Beber e dirigir em uma amostra de condutores que frequentam bares de porto alegre. **Uso de bebidas alcoólicas e outras drogas nas rodovias brasileiras e outros estudos**,

Porto Alegre. Secretaria Nacional de Políticas sobre Drogas, p. 90–95, 2010. Citado na página 29.

BOUWMANS, T. Traditional and recent approaches in background modeling for foreground detection: An overview. **Computer Science Review**, v. 11, n. 0, p. 31–66, 2014. ISSN 1574-0137. Citado na página 55.

_____. Traditional and recent approaches in background modeling for foreground detection: An overview. **Computer Science Review**, Elsevier, v. 11, p. 31–66, 2014. Citado na página 56.

_____. Traditional Approaches in Background Modeling for Static Cameras. In: BOUWMANS, T.; PORIKLI, F.; HÖFERLIN, B.; VACAVANT, A. (Ed.). **Background Modeling and Foreground Detection for Video Surveillance**. Boca Raton, USA: CRC Press, 2014. cap. 1. ISBN 9781482205374. Citado nas páginas 54 e 55.

BRADSKI, G.; KAEHLER, A. **Learning OpenCV: Computer vision with the OpenCV library**. Cambridge, MA: O'Reilly Media, Inc., 2008. Citado na página 50.

BROOKHUIS, K.; WAARD, D. D.; JANSSEN, W. Behavioural impacts of advanced driver assistance systems—an overview. **European Journal of Transport and Infrastructure Research**, v. 1, n. 3, p. 245–253, 2001. Citado na página 35.

BROX, T.; MALIK, J. Object segmentation by long term analysis of point trajectories. In: DANIILIDIS, K.; MARAGOS, P.; PARAGIOS, N. (Ed.). **Computer Vision – ECCV 2010**. [S.l.]: Springer Berlin Heidelberg, 2010, (Lecture Notes in Computer Science, v. 6315). p. 282–295. ISBN 978-3-642-15554-3. Citado na página 58.

BUCKERIDGE, R. **With Autonomous, Self-driving Cars Likely to Be Commonplace By Around 2025, These Vehicles Will Change Our Roads, Our Relationship With Our Cars And Society At Large. Buckle up, a Revolution is Coming!** 2015. Disponível em: <<http://factor-tech.com/feature/autonomous-cars-and-mans-future-the-road-ahead/>>. Citado na página 29.

CAMPLANI, M.; SALGADO, L. RGB-D Cameras for Background-Foreground Segmentation. In: BOUWMANS, T.; PORIKLI, F.; HÖFERLIN, B.; VACAVANT, A. (Ed.). **Background Modeling and Foreground Detection for Video Surveillance**. Boca Raton, USA: CRC Press, 2014. cap. 18. ISBN 9781482205374. Citado na página 61.

CARDARELLI, E. Vision-based blind spot monitoring. In: ESKANDARIAN, A. (Ed.). **Handbook of Intelligent Vehicles**. [S.l.]: Springer London, 2012. p. 1071–1087. ISBN 978-0-85729-084-7. Citado na página 36.

CARROLL, J.; BELLEHUMEUR, D.; CARROLL, C. **System and method for detecting and measuring ethyl alcohol in the blood of a motorized vehicle driver transdermally and non-invasively in the presence of interferents**. [S.l.]: Google Patents, 2013. US Patent App. 2013/0027209. Citado nas páginas 38, 44, 46, 121 e 124.

CAUCHIE, J.; FIOLET, V.; VILLERS, D. Optimization of a hough transform algorithm for the search of a center. **Pattern Recognition**, v. 41, n. 2, p. 567–574, 2008. ISSN 0031-3203. Citado na página 39.

CDC, U. **Mobile Vehicle Safety-Impaired Driving**. 2013. Disponível em: <http://www.cdc.gov/MotorVehicleSafety/Impaired_Driving/impaired-driv_factsheet.html>. Citado na página 29.

- CHACON-MURGUIA, M. I.; GONZALEZ-DUARTE, S. An adaptive neural-fuzzy approach for object detection in dynamic backgrounds for surveillance systems. **Industrial Electronics, IEEE Transactions on**, v. 59, n. 8, p. 3286–3298, Aug 2012. ISSN 0278-0046. Citado na página 57.
- CHEUNG, S. S.; KAMATH, C. Robust techniques for background subtraction in urban traffic video. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **Electronic Imaging 2004**. [S.l.], 2004. v. 5308, p. 881–892. Citado na página 60.
- CHUNG, Y.; WANG, J.; CHEN, S. Progressive background images generation. In: **Proc. of 15th IPPR Conf. on Computer Vision, Graphics and Image Processing**. [S.l.: s.n.], 2002. p. 858–865. Citado na página 55.
- CIDADES, B. Ministério das. **Aplicativo Mãos no Volante**. 2012. Disponível em: <<http://www.paradapelavida.com.br/maos-no-volante/>>. Citado nas páginas 40, 44, 45, 46, 120, 121 e 124.
- CLAPÉS, A.; REYES, M.; ESCALERA, S. Multi-modal user identification and object recognition surveillance system. **Pattern Recognition Letters**, v. 34, n. 7, p. 799–808, 2013. ISSN 0167-8655. Scene Understanding and Behaviour Analysis. Citado na página 61.
- COLQUE, R. V. H. M.; CAMARA-CHAVEZ, G. Progressive background image generation of surveillance traffic videos based on a temporal histogram ruled by a reward/penalty function. In: **Graphics, Patterns and Images (Sibgrapi), 2011 24th SIBGRAPI Conference on**. [S.l.: s.n.], 2011. p. 297–304. Citado na página 55.
- CRAYE, C.; KARRAY, F. Driver distraction detection and recognition using rgb-d sensor. **arXiv preprint arXiv:1502.00250**, 2015. Citado nas páginas 42, 43, 44, 120 e 124.
- CRISTANI, M.; FARENZENA, M.; BLOISI, D.; MURINO, V. Background subtraction for automated multisensor surveillance: A comprehensive review. **EURASIP J. Adv. Signal Process**, Hindawi Publishing Corp., New York, NY, United States, v. 2010, p. 43:1–43:24, fev. 2010. ISSN 1110-8657. Citado na página 59.
- DAI, J.; TENG, J.; BAI, X.; SHEN, Z.; XUAN, D. Mobile phone based drunk driving detection. In: IEEE. **Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2010 4th International Conference on-NO PERMISSIONS**. [S.l.], 2010. p. 1–8. Citado nas páginas 39, 44, 45, 75, 120 e 124.
- DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: **Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on**. [S.l.: s.n.], 2005. v. 1, p. 886–893. ISSN 1063-6919. Citado nas páginas 42 e 64.
- DALLEY, G.; MIGDAL, J.; GRIMSON, W. E. L. Background subtraction for temporally irregular dynamic textures. In: **Applications of Computer Vision, 2008. WACV 2008. IEEE Workshop on**. [S.l.: s.n.], 2008. p. 1–7. ISSN 1550-5790. Citado na página 60.
- DESHMUKH, S. V.; DEHZANGI, O. Ecg-based driver distraction identification using wavelet packet transform and discriminative kernel-based features. In: **2017 IEEE International Conference on Smart Computing (SMARTCOMP)**. [S.l.: s.n.], 2017. p. 1–7. Citado nas páginas 41, 44, 45, 46, 120, 121 e 124.
- DKHIL, M. B.; NEJI, M.; WALI, A.; ALIM, A. M. A new approach for a safe car assistance system. In: **Advanced Logistics and Transport (ICALT), 2015 4th International Conference on**. [S.l.: s.n.], 2015. p. 217–222. Citado nas páginas 40, 43, 44, 45, 120 e 124.

- DOLLAR, P.; WOJEK, C.; SCHIELE, B.; PERONA, P. Pedestrian detection: A benchmark. In: **Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on**. [S.l.: s.n.], 2009. p. 304–311. ISSN 1063-6919. Citado na página [57](#).
- DOSOVITSKIY, A.; ROS, G.; CODEVILLA, F.; LOPEZ, A.; KOLTUN, V. Carla: An open urban driving simulator. **arXiv preprint arXiv:1711.03938**, 2017. Citado na página [82](#).
- EBRAHIM, P.; ABDELLAOUI, A.; STOLZMANN, W.; YANG, B. Eyelid-based driver state classification under simulated and real driving conditions. In: **Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on**. [S.l.: s.n.], 2014. p. 3190–3196. Citado nas páginas [39](#), [44](#), [45](#), [82](#), [120](#) e [124](#).
- EDMONDS, D. S.; HOPTA, J. W. **Driver alcohol ignition interlock**. [S.l.]: Google Patents, 2001. US Patent 6,229,908. Citado na página [38](#).
- EL-LAITHY, R. A.; HUANG, J.; YEH, M. Study on the use of microsoft kinect for robotics applications. In: **Position Location and Navigation Symposium (PLANS), 2012 IEEE/ION**. [S.l.: s.n.], 2012. p. 1280–1288. ISSN 2153-358X. Citado na página [51](#).
- ELGAMMAL, A.; DURAISWAMI, R.; HARWOOD, D.; DAVIS, L. S. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. **Proceedings of the IEEE**, v. 90, n. 7, p. 1151–1163, Jul 2002. ISSN 0018-9219. Citado na página [57](#).
- ELGAMMAL, A.; ELQURSH, A. Background Subtraction for Moving Cameras. In: BOUWMANS, T.; PORIKLI, F.; HÖFERLIN, B.; VACAVANT, A. (Ed.). **Background Modeling and Foreground Detection for Video Surveillance**. Boca Raton, USA: CRC Press, 2014. cap. 4. ISBN 9781482205374. Citado nas páginas [54](#), [57](#) e [58](#).
- ELHABIAN, S. Y.; EL-SAYED, K. M.; AHMED, S. H. Moving object detection in spatial domain using background removal techniques - state-of-art. **Recent Patents on Computer Science**, v. 1, n. 1, p. 32–54, 2008. Citado na página [60](#).
- ELQURSH, A.; ELGAMMAL, A. Online moving camera background subtraction. In: FITZGIBBON, A.; LAZEBNIK, S.; PERONA, P.; SATO, Y.; SCHMID, C. (Ed.). **Computer Vision – ECCV 2012**. [S.l.]: Springer Berlin Heidelberg, 2012, (Lecture Notes in Computer Science, v. 7577). p. 228–241. ISBN 978-3-642-33782-6. Citado na página [59](#).
- ERIKSSON, J.; GIROD, L.; HULL, B.; NEWTON, R.; MADDEN, S.; BALAKRISHNAN, H. The pothole patrol: using a mobile sensor network for road surface monitoring. In: ACM. **Proceedings of the 6th international conference on Mobile systems, applications, and services**. [S.l.], 2008. p. 29–39. Citado na página [37](#).
- FANKHAUSER, P.; BLOESCH, M.; RODRIGUEZ, D.; KAESTNER, R.; HUTTER, M.; SIGEWART, R. Kinect v2 for mobile robot navigation: Evaluation and modeling. **Computer Vision and Image Understanding**, 2015. Citado nas páginas [51](#) e [52](#).
- FERNANDES, L. C.; SOUZA, J. R.; PESSIN, G.; SHINZATO, P. Y.; SALES, D.; MENDES, C.; PRADO, M.; KLASER, R.; MAGALHÃES, A. C.; HATA, A.; PIGATTO, D.; BRANCO, K. C.; JR., V. G.; OSORIO, F. S.; WOLF, D. F. Carina intelligent robotic car: Architectural design and applications. **Journal of Systems Architecture**, Elsevier, v. 60, n. 4, p. 372–392, 2014. ISSN 1383-7621. Citado nas páginas [31](#) e [50](#).

FISCHLER, M. A.; BOLLES, R. C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. **Commun. ACM**, ACM, New York, NY, USA, v. 24, n. 6, p. 381–395, jun. 1981. ISSN 0001-0782. Citado na página 58.

FISHER, A. Inside google's quest to popularize self-driving cars. **Popular Science**. Retrieved from <http://www.popsci.com/cars/article/2013-09/google-self-driving-car>, 2014. Acessado em: 11.06.2015. Citado nas páginas 31 e 50.

FISHER, D. L.; LEE, J. D.; RIZZO, M.; CAIRD, J. **Handbook of driving simulation for engineering, medicine, and psychology**. [S.l.]: CRC Press, 2011. Citado na página 81.

FLOUDAS, N.; AMDITIS, A.; KEINATH, A.; BENGLER, K.; ENGELN, A. Review and taxonomy of ivis/adas applications. **Deliverable D2.1.2**, v. 1, 2005. Citado na página 35.

GERONIMO, D.; SAPPA, A.; LÓPEZ, A. Stereo-based candidate generation for pedestrian protection systems. 2009. Citado na página 37.

GOLDBERG, D. E. **Genetic Algorithms in Search, Optimization, and Machine Learning**. 1. ed. [S.l.]: Addison-Wesley Professional, 1989. Hardcover. ISBN 0201157675. Citado nas páginas 98 e 104.

GÓMEZ, A. E.; SANTOS, T. C. d.; FILHO, C. M.; GOMES, D.; PERAFAN, J. C.; WOLF, D. F. Simulation platform for cooperative vehicle systems. In: **Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on**. [S.l.: s.n.], 2014. p. 1347–1352. Citado na página 81.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. Upper Saddle River, N. J.: Pearson/Prentice Hall, 2008. ISBN 9780131687288. Citado na página 60.

GORDON, G.; DARRELL, T.; HARVILLE, M.; WOODFILL, J. Background estimation and removal based on range and color. In: **Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on**. [S.l.: s.n.], 1999. v. 2, p. 459–464. ISSN 1063-6919. Citado na página 61.

GRUBER, A.; WEISS, Y. Multibody factorization with uncertainty and missing data using the em algorithm. In: **Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on**. [S.l.: s.n.], 2004. v. 1, p. I-707–I-714 Vol.1. ISSN 1063-6919. Citado na página 58.

HAILE, E. **Drunk driver detection system**. [S.l.]: Google Patents, 1987. US Patent 4,716,413. Citado na página 38.

_____. **Drunk driver detection system**. [S.l.]: Google Patents, 1992. US Patent 5,096,329. Citado nas páginas 38, 44, 45, 75, 120 e 124.

HAMMERLE, M.; HOFLE, B.; FUCHS, J.; SCHRODER-RITZRAU, A.; VOLLWEILER, N.; FRANK, N. Comparison of kinect and terrestrial lidar capturing natural karst cave 3-d objects. **Geoscience and Remote Sensing Letters, IEEE**, v. 11, n. 11, p. 1896–1900, Nov 2014. ISSN 1545-598X. Citado na página 53.

HARTLEY, R. I.; ZISSERMAN, A. **Multiple View Geometry in Computer Vision**. Second. Cambridge: Cambridge University Press, 2004. ISBN 0-521-54051-8. Citado na página 58.

- HASSAN, B.; BERSSENBRÜGGE, J.; QAISI, I. A.; STÖCKLEIN, J. Reconfigurable driving simulator for testing and training of advanced driver assistance systems. In: **IEEE. Assembly and Manufacturing (ISAM), 2013 IEEE International Symposium on**. [S.l.], 2013. p. 337–339. Citado na página 81.
- HAYMAN, E.; EKLUNDH, J.-O. Statistical background subtraction for a mobile observer. In: **Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on**. [S.l.: s.n.], 2003. p. 67–74 vol.1. Citado na página 58.
- HEARST, M. A.; DUMAIS, S. T.; OSMAN, E.; PLATT, J.; SCHÖLKOPF, B. Support vector machines. **Intelligent Systems and their Applications, IEEE, IEEE**, v. 13, n. 4, p. 18–28, 1998. Citado na página 66.
- HILLEL, A. B.; LERNER, R.; LEVI, D.; RAZ, G. Recent progress in road and lane detection: a survey. **Machine vision and applications**, Springer, v. 25, n. 3, p. 727–745, 2014. Citado na página 74.
- HOJJATI-EMAMI, K.; DHILLON, B.; JENAB, K. Reliability prediction for the vehicles equipped with advanced driver assistance systems (adas) and passive safety systems (pss). **International Journal of Industrial Engineering**, v. 3, 2012. Citado na página 37.
- HSIAO, P.; YEH, C.; HUANG, S.; FU, L. A portable vision-based real-time lane departure warning system: day and night. **Vehicular Technology, IEEE Transactions on**, IEEE, v. 58, n. 4, p. 2089–2094, 2009. Citado na página 36.
- HU, M. K. Visual pattern recognition by moment invariants. **Information Theory, IRE Transactions on**, IEEE, v. 8, n. 2, p. 179–187, 1962. Citado na página 41.
- HUANG, A. R. W.; CHEN, C. A low-cost driving simulator for full vehicle dynamics simulation. **IEEE Transactions on Vehicular Technology**, IEEE, v. 52, n. 1, p. 162–172, 2003. Citado na página 81.
- ICHIMURA, N. Motion segmentation based on factorization method and discriminant criterion. In: **Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on**. [S.l.: s.n.], 1999. v. 1, p. 600–605 vol.1. Citado na página 58.
- IFIXIT. **Xbox One Kinect Teardown**. 2013. Acessado em 17/06/2015. Disponível em: <<https://www.ifixit.com/Teardown/Xbox+One+Kinect+Teardown/19725>>. Citado na página 52.
- ISHIGAMI, Y.; KLEIN, R. M. Is a hands-free phone safer than a handheld phone? **Journal of Safety Research**, Elsevier, v. 40, n. 2, p. 157–164, 2009. Citado na página 30.
- JAAKKOLA, T.; HAUSSLER, D. Exploiting generative models in discriminative classifiers. In: **Advances in neural information processing systems**. [S.l.: s.n.], 1999. p. 487–493. Citado na página 42.
- JAIN, A. K.; MAO, J.; MOHIUDDIN, K. M. Artificial neural networks: A tutorial. **IEEE computer**, v. 29, n. 3, p. 31–44, 1996. Citado na página 67.
- JAIN, V.; LEARNED-MILLER, E. Online domain adaptation of a pre-trained cascade of classifiers. In: **Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on**. [S.l.: s.n.], 2011. p. 577–584. ISSN 1063-6919. Citado na página 57.

- JAMSON, H. Driving me round the bend—behavioural studies using the new university of leeds driving simulator. In: **2nd Motion Simulator Conference, Braunschweig, Germany**. [S.l.: s.n.], 2007. Citado na página 81.
- JEPSON, A.; BLACK, M. J. Mixture models for optical flow computation. In: **Computer Vision and Pattern Recognition, 1993. Proceedings CVPR '93., 1993 IEEE Computer Society Conference on**. [S.l.: s.n.], 1993. p. 760–761. ISSN 1063-6919. Citado na página 58.
- JIA, L.; LU, M.; WANG, J. Using real-world data to calibrate a driving simulator measuring lateral driving behaviour. **IET Intelligent Transport Systems**, IET, v. 5, n. 1, p. 21–31, 2011. Citado na página 81.
- JIN, Y.; TAO, L.; DI, H.; RAO, N. I.; XU, G. Background modeling from a free-moving camera by multi-layer homography algorithm. In: **Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on**. [S.l.: s.n.], 2008. p. 1572–1575. ISSN 1522-4880. Citado na página 58.
- JOHNSON, D. A.; TRIVEDI, M. M. Driving style recognition using a smartphone as a sensor platform. In: **Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on**. [S.l.: s.n.], 2011. p. 1609–1615. ISSN 2153-0009. Citado nas páginas 42, 44, 45, 74, 120 e 124.
- KANATANI, K. Motion segmentation by subspace separation and model selection. In: **Proc. 8th Int. Conf. Comput. Vision**. [S.l.: s.n.], 2001. p. 586–591. Citado na página 58.
- KAPTEIN, N.; THEEUWES, J.; HORST, R. V. D. Driving simulator validity: Some considerations. **Transportation Research Record: Journal of the Transportation Research Board**, Transportation Research Board of the National Academies, n. 1550, p. 30–36, 1996. Citado na página 81.
- KATHAR, P. N.; BHUYAR, D. L. Design and implementation of driver drowsiness and alcohol intoxication detection using raspberry pi. **Intl. Journal of Innovative Research in Computer and Communication Engineering**, v. 4, n. 8, p. 14617–14625, Aug 2016. ISSN 2320-9798. Citado nas páginas 38, 44 e 124.
- KAUR, G.; KUMAR, D. Lane detection techniques: A review. **International Journal of Computer Applications**, Citeseer, v. 112, n. 10, 2015. Citado na página 74.
- KINDSCHUH, C. S. **Eyewear that simulates bodily impairments**. [S.l.]: Google Patents, 2001. US Patent 6,206,521. Citado na página 86.
- KISACANIN, B. Automotive vision for advanced driver assistance systems. In: **IEEE. VLSI Design, Automation and Test (VLSI-DAT), 2011 International Symposium on**. [S.l.], 2011. p. 1–2. Citado na página 38.
- KNIPLING, R. F.; WIERWILLE, W. W. Vehicle-based drowsy driver detection: Current status and future prospects. In: **Proceedings of the IVHS AMERICA Conference. Moving Toward Deployment**. [S.l.: s.n.], 1994. Citado na página 39.
- KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: **LAWRENCE ERLBAUM ASSOCIATES LTD. International joint Conference on artificial intelligence**. [S.l.], 1995. v. 14, p. 1137–1145. Citado na página 91.

KUMAR, A. M.; SIMON, P. Review of lane detection and tracking algorithms in advanced driver assistance system. **Int. J. Comput. Sci. Inf. Technol**, v. 7, n. 4, p. 65–78, 2015. Citado na página [74](#).

KUMAR, K.; ALKOFFASH, M.; DANGE, S.; IDARROU, A.; SRIDEVI, N.; SHEEBA, J.; SHAH, N.; SHARMA, S.; ELYASI, G.; SAREMI, H. Morphology based facial feature extraction and facial expression recognition for driver vigilance. **International Journal of Computer Applications**, Foundation of Computer Science (FCS), v. 51, n. 2, p. 17–24, 2012. Citado nas páginas [40](#), [43](#), [44](#), [120](#) e [124](#).

KUTILA, M. **Methods for machine vision based driver monitoring applications**. [S.l.]: Cite-seer, 2006. Citado na página [37](#).

KWAK, S.; LIM, T.; NAM, W.; HAN, B.; HAN, J. H. Generalized background subtraction based on hybrid inference by belief propagation and bayesian filtering. In: **Computer Vision (ICCV), 2011 IEEE International Conference on**. [S.l.: s.n.], 2011. p. 2174–2181. ISSN 1550-5499. Citado na página [59](#).

LAPTEV, I.; MARSZALEK, M.; SCHMID, C.; ROZENFELD, B. Learning realistic human actions from movies. In: **Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on**. [S.l.: s.n.], 2008. p. 1–8. ISSN 1063-6919. Citado na página [64](#).

LEE, J.; LI, J.; LIU, L.; CHEN, C. A novel driving pattern recognition and status monitoring system. In: CHANG, L.; LIE, W. (Ed.). **Advances in Image and Video Technology**. [S.l.]: Springer Berlin Heidelberg, 2006, (Lecture Notes in Computer Science, v. 4319). p. 504–512. ISBN 978-3-540-68297-4. Citado nas páginas [42](#), [43](#), [44](#), [74](#), [120](#) e [124](#).

LEENS, J.; PIÉRARD, S.; BARNICH, O.; DROOGENBROECK, M. V.; WAGNER, J. Combining color, depth, and motion for video segmentation. In: FRITZ, M.; SCHIELE, B.; PIATER, J. H. (Ed.). **Computer Vision Systems**. [S.l.]: Springer Berlin Heidelberg, 2009, (Lecture Notes in Computer Science, v. 5815). p. 104–113. ISBN 978-3-642-04666-7. Citado na página [61](#).

LEFÈVRE, S.; LAUGIER, C.; IBÁÑEZ-GUZMÁN, J. Exploiting map information for driver intention estimation at road intersections. In: IEEE. **Intelligent Vehicles Symposium (IV), 2011 IEEE**. [S.l.], 2011. p. 583–588. Citado na página [35](#).

LEFLOCH, D.; NAIR, R.; LENZEN, F.; SCHÄFER, H.; STREETER, L.; CREE, M. J.; KOCH, R.; KOLB, A. Technical foundation and calibration methods for time-of-flight cameras. In: GRZEGORZEK, M.; THEOBALT, C.; KOCH, R.; KOLB, A. (Ed.). **Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications**. [S.l.]: Springer Berlin Heidelberg, 2013, (Lecture Notes in Computer Science, v. 8200). p. 3–24. ISBN 978-3-642-44963-5. Citado na página [51](#).

LENSKIY, A.; LEE, J. Driver's eye blinking detection using novel color and texture segmentation algorithms. **International Journal of Control, Automation and Systems**, Springer, v. 10, n. 2, p. 317–327, 2012. Citado nas páginas [40](#), [43](#), [44](#), [120](#) e [124](#).

LINDSTRÖM, J.; LINDGREN, F.; ÅSTRÖM, K.; HOLST, J.; HOLST, U. Background and foreground modeling using an online em algorithm. In: JONES, G. (Ed.). **IEEE International Workshop on Visual Surveillance**. [S.l.]: Faculty of Computing, Information Systems and mathematics, Kingston University, Surrey, UK, 2006. VS2006, p. 9–16. Citado na página [56](#).

- LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, Kluwer Academic Publishers, v. 60, n. 2, p. 91–110, 2004. ISSN 0920-5691. Citado na página 64.
- LUN, R.; ZHAO, W. A survey of applications and human motion recognition with micro-soft kinect. **International Journal of Pattern Recognition and Artificial Intelligence**, World Scientific, 2015. Citado nas páginas 62 e 63.
- MADDALENA, L.; PETROSINO, A. A fuzzy spatial coherence-based approach to background/-foreground separation for moving object detection. **Neural Computing and Applications**, Springer-Verlag, v. 19, n. 2, p. 179–186, 2010. ISSN 0941-0643. Citado na página 56.
- MAGEE, D. R. Tracking multiple vehicles using foreground, background and motion models. **Image and Vision Computing**, v. 22, n. 2, p. 143–155, 2004. ISSN 0262-8856. Statistical Methods in Video Processing. Citado na página 56.
- MAHBUB, U.; IMTIAZ, H.; ROY, T.; RAHMAN, M. S.; AHAD, M. A. R. A template matching approach of one-shot-learning gesture recognition. **Pattern Recognition Letters**, v. 34, n. 15, p. 1780–1788, 2013. ISSN 0167-8655. Smart Approaches for Human Action Recognition. Citado na página 61.
- MANIKOTH, N.; LOCE, R.; BERNAL, E.; WU, W. Survey of computer vision in roadway transportation systems. **Proc. of SPIE-IS&T Electronic Imaging Vol.**, v. 8305, p. 83050W, 2012. Citado na página 36.
- MARENCO, D.; FONTANA, D.; GHISIO, G.; MONCHIERO, G.; CARDARELLI, E.; MEDICI, P.; VISLAB, P. A validation tool for traffic signs recognition systems. In: IEEE. **Intelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference on**. [S.l.], 2009. p. 1–6. Citado na página 36.
- MATARIC, M. **Introdução a Robótica**. 1ed(traduzido). ed. São Paulo, SP, Brasil: Blucher/UNESP Editora, 2014. ISBN 9788521208532. Citado na página 48.
- MAYDEN, A. **Digitaltutors: Unreal engine 4 vs. unity: Which game engine is best for you**. 2014. Disponível em: <<https://www.pluralsight.com/blog/film-games/unreal-engine-4-vs-unity-game-engine-best>>. Citado na página 82.
- MEHROTRA, K.; MOHAN, C. K.; RANKA, S. **Elements of artificial neural networks**. [S.l.]: MIT press, 1996. Citado na página 67.
- MENDES, C. C. T.; FRÉMONT, V.; WOLF, D. F. Exploiting fully convolutional neural networks for fast road detection. In: IEEE. **Robotics and Automation (ICRA), 2016 IEEE International Conference on**. [S.l.], 2016. p. 3174–3179. Citado na página 74.
- MEYER, D. Support vector machines. **Technische Universität Wien, Austria**, 2012. Citado na página 65.
- MITROVIC, D. Reliable method for driving events recognition. **Intelligent Transportation Systems, IEEE Transactions on**, v. 6, n. 2, p. 198–205, June 2005. ISSN 1524-9050. Citado na página 42.
- MOESLUND, T. B.; HILTON, A.; KRÜGER, V. A survey of advances in vision-based human motion capture and analysis. **Computer Vision and Image Understanding**, v. 104, n. 2, p. 90–126, 2006. ISSN 1077-3142. Citado na página 62.

- MOLINIER, M.; HÄME, T.; AHOLA, H. 3d-connected components analysis for traffic monitoring in image sequences acquired from a helicopter. In: KALVIAINEN, H.; PARKKINEN, J.; KAARNA, A. (Ed.). **Image Analysis**. [S.l.]: Springer Berlin Heidelberg, 2005, (Lecture Notes in Computer Science, v. 3540). p. 141–150. ISBN 978-3-540-26320-3. Citado na página 55.
- MURATA, K.; FUJITA, E.; KOJIMA, S.; MAEDA, S.; OGURA, Y.; KAMEI, T.; TSUJI, T.; KANEKO, S.; YOSHIKUMI, M.; SUZUKI, N. Noninvasive biological sensor system for detection of drunk driving. **Information Technology in Biomedicine, IEEE Transactions on**, v. 15, n. 1, p. 19–25, Jan 2011. ISSN 1089-7771. Citado nas páginas 29, 38, 44, 46, 121 e 124.
- MURPHY, R. R. **Introduction to AI Robotics**. 1st. ed. Cambridge, MA, USA: MIT Press, 2000. ISBN 0262133830. Citado na página 47.
- MURRAY, D.; BASU, A. Motion tracking with an active camera. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 16, n. 5, p. 449–459, May 1994. ISSN 0162-8828. Citado na página 58.
- NEUTEL, C. I. Risk of traffic accident injury after a prescription for a benzodiazepine. **Annals of Epidemiology**, v. 5, n. 3, p. 239–244, 1995. ISSN 1047-2797. Citado nas páginas 29 e 30.
- NHTSA. Driver electronic device use in 2010. **Traffic Safety Facts - December 2011**, p. 1–8, 2011. Citado na página 30.
- NTSB. Evaluation of us department of transportation efforts in the 1990s to address operator fatigue. DTIC Document, 1999. Citado na página 30.
- OCHS, P.; BROX, T. Higher order motion models and spectral clustering. In: **Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on**. [S.l.: s.n.], 2012. p. 614–621. ISSN 1063-6919. Citado na página 58.
- OFLI, F.; CHAUDHRY, R.; KURILLO, G.; VIDAL, R.; BAJCSY, R. Sequence of the most informative joints (smij): A new representation for human skeletal action recognition. **Journal of Visual Communication and Image Representation**, v. 25, n. 1, p. 24–38, 2014. ISSN 1047-3203. Visual Understanding and Applications with RGB-D Cameras. Citado na página 64.
- OGAMA, W. O. Art. 306 do código de trânsito brasileiro: do texto original às mudanças surgidas com o advento da lei n. 12.760/12. **UNOPAR Científica Ciências Jurídicas e Empresariais**, v. 15, n. 2, 2014. Citado na página 29.
- OTSU, N. A threshold selection method from gray-level histograms. **IEEE Transactions on Systems, Man and Cybernetics**, v. 9, n. 1, p. 62–66, 1979. Citado na página 61.
- PARK, H.; AHN, D.; PARK, T.; SHIN, K. G. Automatic identification of driver's smartphone exploiting common vehicle-riding actions. **IEEE Transactions on Mobile Computing**, IEEE, v. 17, n. 2, p. 265–278, Feb 2018. ISSN 1536-1233. Citado nas páginas 41, 44, 45, 120 e 124.
- PAULA, M. B. d.; JUNG, C. R. Real-time detection and classification of road lane markings. In: IEEE. **Graphics, Patterns and Images (SIBGRAPI), 2013 26th SIBGRAPI-Conference on**. [S.l.], 2013. p. 83–90. Citado na página 74.
- PECHANSKY, F.; BONI, R. D.; DIEMEN, L. V.; BUMAGUIN, D.; PINSKY, I.; ZALESKI, M.; CAETANO, R.; LARANJEIRA, R. Highly reported prevalence of drinking and driving in brazil: data from the first representative household study. **Revista Brasileira de Psiquiatria, SciELO Brasil**, v. 31, n. 2, p. 125–130, 2009. Citado na página 29.

- PEISSNER, M.; DOEBLER, V.; METZE, F. Can voice interaction help reducing the level of distraction and prevent accidents? 2011. Citado na página 30.
- PHILIP, P.; TAILLARD, J.; KLEIN, E.; SAGASPE, P.; CHARLES, A.; DAVIES, W.; GUILLEMINAULT, C.; BIOULAC, B. Effect of fatigue on performance measured by a driving simulator in automobile drivers. **Journal of Psychosomatic Research**, Elsevier, v. 55, n. 3, p. 197–200, 2003. Citado na página 85.
- PINILLA, A. C. C.; QUINTERO, M. C. G.; PREMACHANDRA, C. Intelligent driving diagnosis based on a fuzzy logic approach in a real environment implementation. In: **Intelligent Vehicles Symposium Proceedings, 2014 IEEE**. [S.l.: s.n.], 2014. p. 102–107. Citado nas páginas 43, 44, 74 e 124.
- PINTO, A. M.; COSTA, P.; MOREIRA, A. P.; ROCHA, L. F.; VEIGA, G.; MOREIRA, E. Evaluation of depth sensors for robotic applications. In: **Autonomous Robot Systems and Competitions (ICARSC), 2015 IEEE International Conference on**. [S.l.: s.n.], 2015. p. 139–143. Citado na página 52.
- POOLE, G. **The Complete Book of Questions: 1001 Conversation Starters for Any Occasion**. [S.l.]: Zondervan, 2003. ISBN 9780310244202. Citado na página 86.
- POPPE, R. Vision-based human motion analysis: An overview. **Computer Vision and Image Understanding**, v. 108, n. 1, p. 4–18, 2007. ISSN 1077-3142. Special Issue on Vision for Human-Computer Interaction. Citado na página 62.
- PORIKLI, F.; TUZEL, O. Human body tracking by adaptive background models and mean-shift analysis. In: CITESEER. **IEEE International Workshop on Performance Evaluation of Tracking and Surveillance**. [S.l.], 2003. p. 1–9. Citado na página 56.
- POURREZA, H. R. Fuzzy running average and fuzzy background subtraction: Concepts and application. **International Journal of Computer Science and Network Security**, v. 8, n. 2, p. 138–143, February 2008. ISSN 1738-7906. Citado na página 56.
- RADKE, R. J.; ANDRA, S.; AL-KOFAHI, O.; ROYSAM, B. Image change detection algorithms: a systematic survey. **Image Processing, IEEE Transactions on**, v. 14, n. 3, p. 294–307, March 2005. ISSN 1057-7149. Citado na página 57.
- RAIMUNDO, D. A. R. **Simulator for driving performance monitoring**. Tese (Doutorado) — Instituto Superior de Engenharia de Lisboa, 2017. Citado na página 81.
- RAND, W. M. Objective criteria for the evaluation of clustering methods. **Journal of the American Statistical Association**, Taylor & Francis, v. 66, n. 336, p. 846–850, 1971. Citado na página 92.
- RAY, W. A.; FOUGHT, R. L.; DECKER, M. D. Psychoactive drugs and the risk of injurious motor vehicle crashes in elderly drivers. **American journal of epidemiology**, Oxford Univ Press, v. 136, n. 7, p. 873–883, 1992. Citado nas páginas 29 e 30.
- REDELMEIER, D. A.; TIBSHIRANI, R. J. Association between cellular-telephone calls and motor vehicle collisions. **The New England Journal of Medicine**, v. 336, n. 7, p. 453–458, 1997. Citado na página 30.
- REGAN, M. A.; LEE, J. D.; YOUNG, K. L. **Driver distraction: Theory, effects, and mitigation**. [S.l.]: CRC, 2008. Citado na página 30.

- RICHTER, S. R.; HAYDER, Z.; KOLTUN, V. Playing for benchmarks. In: **International conference on computer vision (ICCV)**. [S.l.: s.n.], 2017. v. 2. Citado na página 81.
- RICHTER, S. R.; VINEET, V.; ROTH, S.; KOLTUN, V. Playing for data: Ground truth from computer games. In: SPRINGER. **European Conference on Computer Vision**. [S.l.], 2016. p. 102–118. Citado na página 81.
- RIEDMILLER, M.; BRAUN, H. Rprop-a fast adaptive learning algorithm. In: CITESEER. **Proc. of ISICIS VII, Universitat**. [S.l.], 1992. Citado nas páginas 68, 94, 97 e 103.
- ROMERO, R. A. F.; PRESTES, E.; OSÓRIO, F.; WOLF, D. Sensores e Atuadores. In: **Robótica Móvel**. Rio de Janeiro: Editora LTC, 2014. cap. 2, p. 13–25. ISBN 978-85-216-2303-8. Citado nas páginas 47, 48, 49 e 50.
- ROWE, S.; BLAKE, A. Statistical mosaics for tracking. **Image and Vision Computing**, v. 14, n. 8, p. 549–564, 1996. ISSN 0262-8856. 6th British Machine Vision Conference. Citado na página 58.
- SALVADOR, A. A culpa foi do celular. **Revista Veja**, 21/12/2011, 2011. Citado na página 30.
- SANTOS, D. B.; BITTENCOURT, L. G.; VIEGAS, C. A. d. A. de; GAIO, E. Sonolência diurna e atenção em motoristas de ônibus urbanos de 2 capitais do brasil. **Revista Portuguesa de Pneumologia**, v. 19, n. 4, p. 152–156, 2013. ISSN 0873-2159. Citado na página 30.
- SARBOLANDI, H.; LEFLOCH, D.; KOLB, A. Kinect range sensing: Structured-light versus time-of-flight kinect. **Computer Vision and Image Understanding**, 2015. ISSN 1077-3142. Citado nas páginas 51 e 52.
- SCHÖLKOPF, B. **Support Vector Learning**. Dissertação (Mestrado) — Technischen Universität Berlin, Berlin, 1997. Citado na página 67.
- SESHADRI, K.; JUEFEI-XU, F.; PAL, D. K.; SAVVIDES, M.; THOR, C. P. Driver cell phone usage detection on strategic highway research program (shrp2) face view videos. In: . [S.l.: s.n.], 2015. Citado nas páginas 42, 43, 44, 46, 120, 121 e 124.
- SHAH, S.; DEY, D.; LOVETT, C.; KAPOOR, A. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In: SPRINGER. **Field and service robotics**. [S.l.], 2018. p. 621–635. Citado na página 82.
- SHEIKH, Y.; JAVED, O.; KANADE, T. Background subtraction for freely moving cameras. In: **Computer Vision, 2009 IEEE 12th International Conference on**. [S.l.: s.n.], 2009. p. 1219–1225. ISSN 1550-5499. Citado na página 59.
- SHIRAZI, M. M.; RAD, A. B. Modeling the steering behavior of intoxicated drivers. In: **Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on**. [S.l.: s.n.], 2012. p. 648–653. ISSN 2153-0009. Citado na página 75.
- _____. Detection of intoxicated drivers using online system identification of steering behavior. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 15, n. 4, p. 1738–1747, 2014. Citado nas páginas 39, 44, 46, 82, 121 e 124.
- SHOTTON, J.; SHARP, T.; KIPMAN, A.; FITZGIBBON, A.; FINOCCHIO, M.; BLAKE, A.; COOK, M.; MOORE, R. Real-time human pose recognition in parts from single depth images. **Commun. ACM**, ACM, New York, NY, USA, v. 56, n. 1, p. 116–124, jan. 2013. ISSN 0001-0782. Citado nas páginas 62 e 72.

SIEGWART, R.; NOURBAKHS, I. R.; SCARAMUZZA, D. **Introduction to Autonomous Mobile Robots**. 2nd. ed. [S.l.]: The MIT Press, 2011. ISBN 0262015358, 9780262015356. Citado nas páginas 47, 48 e 49.

SIVABALAKRISHNAN, M.; MANJULA, D. Adaptive background subtraction in dynamic environments using fuzzy logic. **International Journal of Video & Image Processing & Network Security**, v. 10, n. 1, 2010. Citado na página 57.

SLOB, J. State-of-the-art driving simulators, a literature survey. **DCT report**, Eindhoven University of Technology, v. 107, 2008. Citado na página 81.

STANIMIROVA, I.; ÜSTÜN, B.; CAJKA, T.; RIDDELOVA, K.; HAJŠLOVA, J.; BUYDENS, L.; WALCZAK, B. Tracing the geographical origin of honeys based on volatile compounds profiles assessment using pattern recognition techniques. **Food Chemistry**, Elsevier, v. 118, n. 1, p. 171–176, 2010. Citado na página 66.

STAUFFER, C.; GRIMSON, W. E. L. Adaptive background mixture models for real-time tracking. In: **Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on**. [S.l.: s.n.], 1999. v. 2, p. 246–252. ISSN 1063-6919. Citado na página 61.

STONE, E.; SKUBIC, M. Evaluation of an inexpensive depth camera for in-home gait assessment. **Journal of Ambient Intelligence and Smart Environments**, IOS Press, Nieuwe Hemweg 6 B Amsterdam 1013 BG Netherlands, v. 3, n. 4, p. 349–361, 2011. Citado na página 61.

STÖRMER, A.; HOFMANN, M.; RIGOLL, G. Depth gradient based segmentation of overlapping foreground objects in range images. In: **Information Fusion (FUSION), 2010 13th Conference on**. [S.l.: s.n.], 2010. p. 1–4. Citado na página 61.

STOYANOV, T.; MOJTAHEDZADEH, R.; ANDREASSON, H.; LILIENTHAL, A. J. Comparative evaluation of range sensor accuracy for indoor mobile robotics and automated logistics applications. **Robotics and Autonomous Systems**, v. 61, n. 10, p. 1094–1105, 2013. ISSN 0921-8890. Selected Papers from the 5th European Conference on Mobile Robots (ECMR 2011). Citado na página 50.

STRAYER, D. L.; COOPER, J. M.; TURRILL, J.; COLEMAN, J.; MEDEIROS-WARD, N.; BIONDI, F. Measuring cognitive distraction in the automobile. **AAA Foundation for Traffic Safety - June 2013**, p. 1–34, 2013. Citado na página 30.

STRAYER, D. L.; DREWS, F. A. Profiles in driver distraction: effects of cell phone conversations on younger and older drivers. **The Journal of the Human Factors and Ergonomics**, v. 46, n. 4, p. 640–648, 2004. Citado na página 30.

STRAYER, D. L.; DREWS, F. A.; CROUCH, D. J. A comparison of the cell phone driver and the drunk driver. **Human factors: The journal of the human factors and ergonomics society**, Sage Publications, v. 48, n. 2, p. 381–391, 2006. Citado na página 30.

STRAYER, D. L.; DREWS, F. A.; JOHNSTON, W. A. Cell phone-induced failures of visual attention during simulated driving. **Journal of experimental psychology: Applied**, American Psychological Association, v. 9, n. 1, p. 23, 2003. Citado na página 30.

STRAYER, D. L.; WATSON, J. M.; DREWS, F. A. 2 cognitive distraction while multitasking in the automobile. **Psychology of Learning and Motivation-Advances in Research and Theory**, v. 54, p. 29, 2011. Citado na página 30.

STRINGA, E. Morphological change detection algorithms for surveillance applications. In: **British Machine Vision Conference**. [S.l.: s.n.], 2000. Citado na página 57.

TAHERI, S. M.; MATSUSHITA, K.; SASAKI, M. Virtual reality driving simulation for measuring driver behavior and characteristics. **Journal of transportation technologies**, Scientific Research Publishing, v. 7, n. 02, p. 123, 2017. Citado na página 81.

TANGO, F.; CARREA, P.; GOBETTO, E. The development of a smart pre-crash system—the chameleon project. In: **Proceedings of ITS 2000, 7th World Congress on Intelligent Transport Systems**. [S.l.: s.n.], 2000. Citado na página 37.

TAPANI, A. Analysis of system effects of driver assistance systems by traffic simulation. In: **Proceeding of the Young Researchers Seminar organised by ECTRI, FEHRL and FERSI**. Accessible through: <http://www.ectri.org/YRS07/Papiers/Session-5/Tapani.pdf>. [S.l.: s.n.], 2007. Citado na página 35.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition, Fourth Edition**. [S.l.]: Academic Press, 2009. Hardcover. ISBN 9781597492720. Citado na página 65.

TORR, P. H. S.; SZELISKI, R.; ANANDAN, P. An integrated bayesian approach to layer extraction from image sequences. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 23, n. 3, p. 297–303, Mar 2001. ISSN 0162-8828. Citado na página 58.

TORR, P. H. S.; ZISSERMAN, A. Mlesac: A new robust estimator with application to estimating image geometry. **Computer Vision and Image Understanding**, v. 78, n. 1, p. 138–156, 2000. ISSN 1077-3142. Citado na página 58.

TOTZKE, I.; JESSBERGER, S.; MUHLBACHER, D.; KRUGER, H. Semi-autonomous advanced parking assistants: Do they really have to be learned if steering is automated? **Intelligent Transport Systems, IET**, IET, v. 5, n. 2, p. 141–147, 2011. Citado na página 36.

VAPNIK, V. **The nature of statistical learning theory**. New York: Springer-Verlag, 1995. Citado na página 65.

VEERARAGHAVAN, H.; BIRD, N.; ATEV, S.; PAPANIKOLOPOULOS, N. Classifiers for driver activity monitoring. **Transportation Research Part C: Emerging Technologies**, Elsevier, v. 15, n. 1, p. 51–67, 2007. Citado nas páginas 41, 43, 44, 45, 120 e 124.

VIDAL, R.; HARTLEY, R. Motion segmentation with missing data using powerfactorization and gpca. In: **Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on**. [S.l.: s.n.], 2004. v. 2, p. II–310–II–316 Vol.2. ISSN 1063-6919. Citado na página 58.

VIOLA, P.; JONES, M. Robust real-time face detection. **International journal of computer vision**, Springer, v. 57, n. 2, p. 137–154, 2004. Citado nas páginas 39 e 40.

WANG, J.; LIU, Z.; WU, Y. Conclusion. In: **Human Action Recognition with Depth Cameras**. [S.l.]: Springer International Publishing, 2014, (SpringerBriefs in Computer Science). p. 57–58. ISBN 978-3-319-04560-3. Citado na página 65.

_____. Introduction. In: **Human Action Recognition with Depth Cameras**. [S.l.]: Springer International Publishing, 2014, (SpringerBriefs in Computer Science). p. 1–9. ISBN 978-3-319-04560-3. Citado nas páginas 63 e 64.

_____. Learning actionlet ensemble for 3d human action recognition. In: **Human Action Recognition with Depth Cameras**. [S.l.]: Springer International Publishing, 2014, (SpringerBriefs in Computer Science). p. 11–40. ISBN 978-3-319-04560-3. Citado nas páginas 53 e 64.

_____. Random occupancy patterns. In: **Human Action Recognition with Depth Cameras**. [S.l.]: Springer International Publishing, 2014, (SpringerBriefs in Computer Science). p. 41–55. ISBN 978-3-319-04560-3. Citado na página 64.

WANG, J. Y. A.; ADELSON, E. H. Representing moving images with layers. **Image Processing, IEEE Transactions on**, v. 3, n. 5, p. 625–638, Sep 1994. ISSN 1057-7149. Citado na página 58.

WANG, L. **Support Vector Machines: theory and applications**. Berlin, Germany: Springer, 2005. v. 177. Citado nas páginas 56 e 65.

WERBOS, P. J. **The roots of backpropagation: from ordered derivatives to neural networks and political forecasting**. [S.l.]: Wiley-Interscience, 1994. Citado na página 68.

WHITE, B.; SHAH, M. Automatically tuning background subtraction parameters using particle swarm optimization. In: **Multimedia and Expo, 2007 IEEE International Conference on**. [S.l.: s.n.], 2007. p. 1826–1829. Citado na página 56.

WILLIAMS, S. J. Our hard days' nights. **Contexts**, SAGE Publications, v. 10, n. 1, p. 26–31, 2011. Citado nas páginas 29 e 30.

WILSON, F. A.; STIMPSON, J. P. Trends in fatalities from distracted driving in the united states, 1999 to 2008. **American Journal of Public Health**, American Public Health Association, v. 100, n. 11, p. 2213–2219, 2010. Citado na página 29.

WOLF, D. F.; SIMÕES, E. V.; OSÓRIO, F. S.; JR., O. T. Robótica inteligente: Da simulação as aplicações no mundo real. In: **Anais do CSBC'09 XXIV Jornada de Atualização em Informática (JAI)**. Bento Gonçalves, RS: SBC, 2009. cap. 6, p. 279–330. Citado na página 47.

WREN, C. R.; AZARBAYEJANI, A.; DARRELL, T.; PENTLAND, A. P. Pfnder: real-time tracking of the human body. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 19, n. 7, p. 780–785, Jul 1997. ISSN 0162-8828. Citado na página 57.

YAN, J.; POLLEFEYS, M. A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate. In: LEONARDIS, A.; BISCHOF, H.; PINZ, A. (Ed.). **Computer Vision – ECCV 2006**. [S.l.]: Springer Berlin Heidelberg, 2006, (Lecture Notes in Computer Science, v. 3954). p. 94–106. ISBN 978-3-540-33838-3. Citado na página 58.

YANG, J.; SIDHOM, S.; CHANDRASEKARAN, G.; VU, T.; LIU, H.; CECAN, N.; CHEN, Y.; GRUTESER, M.; MARTIN, R. P. Detecting driver phone use leveraging car speakers. In: **ACM. Proceedings of the 17th annual international conference on Mobile computing and networking**. [S.l.], 2011. p. 97–108. Citado nas páginas 40, 44, 45, 120 e 124.

YANG, L.; ZHANG, L.; DONG, H.; ALELAIWI, A.; SADDIK, A. E. Evaluating and improving the depth accuracy of kinect for windows v2. **Sensors Journal, IEEE**, v. 15, n. 8, p. 4275–4285, Aug 2015. ISSN 1530-437X. Citado nas páginas 52, 53 e 83.

YANG, X.; TIAN, Y. Eigenjoints-based action recognition using naive-bayes-nearest-neighbor. In: **Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on**. [S.l.: s.n.], 2012. p. 14–19. ISSN 2160-7508. Citado na página [64](#).

ZANG, Q.; KLETTE, R. Evaluation of an adaptive composite gaussian model in video surveillance. In: PETKOV, N.; WESTENBERG, M. A. (Ed.). **Computer Analysis of Images and Patterns**. [S.l.]: Springer Berlin Heidelberg, 2003, (Lecture Notes in Computer Science, v. 2756). p. 165–172. ISBN 978-3-540-40730-0. Citado na página [56](#).

ZHANG, Y.; LIANG, Z.; HOU, Z.; WANG, H.; TAN, M. An adaptive mixture gaussian background model with online background reconstruction and adjustable foreground mergence time for motion segmentation. In: **Industrial Technology, 2005. ICIT 2005. IEEE International Conference on**. [S.l.: s.n.], 2005. p. 23–27. Citado na página [56](#).

ZHAO, Y.; LIU, Z.; YANG, L.; CHENG, H. Combing rgb and depth map features for human activity recognition. In: **Signal Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific**. [S.l.: s.n.], 2012. p. 1–4. Citado na página [64](#).