

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

MAURÍCIO BRAGA DE PAULA

**Visão computacional para veículos
inteligentes usando câmeras embarcadas**

Tese apresentada como requisito parcial para a
obtenção do grau de Doutor em Ciência da
Computação

Orientador: Prof. Dr. Cláudio Rosito Jung

Porto Alegre
2015

CIP — CATALOGAÇÃO NA PUBLICAÇÃO

de Paula, Maurício Braga

Visão computacional para veículos inteligentes usando câmeras embarcadas / Maurício Braga de Paula. – Porto Alegre: PPGC da UFRGS, 2015.

102 f.: il.

Tese (doutorado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2015. Orientador: Cláudio Rosito Jung.

1. Câmeras veiculares embarcadas. 2. Detecção de linhas de bordo da pista. 3. Calibração de câmera. 4. Sinalização horizontal e vertical. 5. Sistemas de assistência ao motorista. 6. Realidade aumentada. I. Jung, Cláudio Rosito. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitor de Pós-Graduação: Prof. Vladimir Pinheiro do Nascimento

Diretor do Instituto de Informática: Prof. Luis da Cunha Lamb

Coordenador do PPGC: Prof. Luigi Carro

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

AGRADECIMENTOS

Após uma longa jornada de trabalho na tese de doutorado, agradeço primeiramente ao meu orientador Prof. Dr. Cláudio Rosito Jung pela orientação, confiança, incentivo e suporte durante a realização do meu doutorado. Suas contribuições e ensinamentos foram fundamentais para o enriquecimento da tese e que os objetivos traçados em 2011 fossem concretizados. Agradeço também pela amizade e compreensão em vários momentos.

Ao Prof. Dr. Marcelo Walter e ao Prof. Dr. Luiz Gonzaga da Silveira Jr. pelo apoio e contribuição. Aos membros da banca examinadora: Prof. Dr. Anderson Maciel (UFRGS), Prof. Dr. David Menotti (UFOP) e Prof. Dr. Fernando Santos Osório (USP - São Carlos) pela avaliação, comentários e contribuições para o aperfeiçoamento deste trabalho.

Aos meus pais Ivanor e Neusa que me ensinaram os verdadeiros caminhos a serem trilhados; à minha amiga e irmã Jaqueline, companheira durante esses anos, pelo incentivo e acompanhamento da minha vida acadêmica. À minha esposa Angelisa e ao meu filho Diogo pela paciência, carinho e compreensão. Aos demais familiares pelo apoio.

Aos colegas do PPGC pelo apoio e pelas incansáveis discussões e contribuições. Aos colegas da UFPel pelo suporte e incentivo. À UFPel e ao Programa de Pós-Graduação da UFRGS. Enfim, a todos aqueles que de alguma forma contribuíram para a realização deste trabalho.

RESUMO

O uso de sistemas de assistência ao motorista (DAS) baseados em visão tem contribuído consideravelmente na redução de acidentes e conseqüentemente no auxílio de uma melhor condução. Estes sistemas utilizam basicamente uma câmera de vídeo embarcada (normalmente fixada no para-brisa) com o propósito de extrair informações acerca da rodovia e ajudar o condutor num melhor processo de dirigibilidade. Pequenas distrações ou a perda de concentração podem ser suficientes para que um acidente ocorra. Este trabalho apresenta uma proposta para o desenvolvimento de algoritmos para extrair informações sobre a sinalização em rodovias. Mais precisamente, serão abordados algoritmos de calibração de câmera explorando a geometria da pista, de extração da marcação de pintura (sinalização horizontal) e detecção e identificação de placas de trânsito (sinalização vertical). Os resultados experimentais indicam que o método de calibração de câmera alcançou bons resultados na obtenção dos parâmetros extrínsecos com erros inferiores a 0.5° . O erro médio encontrado nos experimentos com relação a estimativa da altura da câmera foi em torno de 12 cm (erro relativo aproximado de 10%), permitindo explorar o uso da realidade aumentada como uma possível aplicação. A acurácia global para a detecção e reconhecimento da sinalização horizontal (marcas seccionadas, contínuas e mistas) foi acima de 96% perante uma diversidade de situações apresentadas, tais como: sombras, variação de iluminação, degradação do asfalto e pintura. O uso da câmera calibrada para a detecção da sinalização vertical contribui para delimitar o espaço de varredura da janela deslizante do detector, bem como realizar a procura por placas em uma única escala para cada região de busca, caracterizada pela distância ao veículo. Os resultados apresentados reportam uma taxa global de classificação de aproximadamente 99% para o sinal de proibido ultrapassar, considerando-se uma base de dados limitada a 962 amostras.

Palavras-chave: Câmeras veiculares embarcadas. Detecção de linhas de bordo da pista. Calibração de câmera. Sinalização horizontal e vertical. Sistemas de assistência ao motorista. Realidade aumentada.

Computer vision for intelligent vehicles using embedded cameras

ABSTRACT

The use of driver assistance systems (DAS) based on computer vision has helped considerably in reducing accidents and consequently aid in better driving. These systems primarily use an embedded video camera (usually fixed on the windshield) for the purpose of extracting information about the highway and assisting the driver in a better handling process. Small distractions or loss of concentration may be sufficient for an accident to occur. This work presents the development of algorithms to extract information about traffic signs on highways. More specifically, this work will tackle a camera calibration algorithm that exploits the geometry of the road track, algorithms for the extraction of road marking paint (lane markings) and detection and identification of vertical traffic signs. Experimental results indicate that the proposed method for obtaining the extrinsic parameters achieve good results with errors of less than 0.5° . The average error in our experiments, related to the camera height, were around 12 cm (relative error around 10%). Global accuracy for the detection and classification of road lane markings (dashed, solid, dashed-solid, solid-dashed or double solid) were over 96%. Finally, our camera calibration algorithm was used to reduce the search region and to define the scale of a sliding-window detector for vertical traffic signs. The use of the calibrated camera for the detection of traffic signs contributes to define the scanning area of the sliding window and perform a search for signs on a unique scale for each region of interest, determined by the distance to the vehicle. The results reported a global classification rate of approximately 99% for the no overtaking sign, considering a limited of 962 samples.

Keywords: Embedded vehicular cameras. Detection of road lane lines. Camera calibration. Horizontal and vertical traffic sign. Driver assistance systems. Augmented reality.

LISTA DE ABREVIATURAS E SIGLAS

ANTT	Agência Nacional de Transportes Terrestres
AR	<i>Augmented Reality</i>
BLM	<i>Basic Lane Marking</i>
CNM	Confederação Nacional de Municípios do Brasil
DAS	<i>Driver Assistance Systems</i>
DATASUS	Departamento de Informática do SUS
DNIT	Departamento Nacional de Infraestrutura de Transportes
FDP	Função de Distribuição de Probabilidade
FFT	<i>Fast Fourier Transform</i>
HOG	<i>Histograms of Oriented Gradients</i>
HSI	<i>Hue, Saturation, Intensity</i>
HMM	<i>Hidden Markov Model</i>
IPM	<i>Inverse Perspective Mapping</i>
IV	<i>Intelligent Vehicles</i>
IRTAD	<i>International Traffic Safety Data and Analysis Group</i>
LCA	<i>Lane Change Assistance</i>
LDW	<i>Lane Departure Warning</i>
LIDAR	<i>Light Detection And Ranging</i>
MAD	<i>Median Absolute Variation</i>
qps	quadros por segundo
RANSAC	<i>RANdom SAmple Consensus</i>
RBF	<i>Radial Basis Function</i>
ROI	<i>Region Of Interest</i>
SVM	<i>Support Vector Machine</i>

TSR *Traffic Sign Recognition*

WHO *World Health Organization*

LISTA DE SÍMBOLOS

$\mathbf{W} = (x_w, y_w, z_w)^T$	Ponto no sistema de coordenadas de mundo
$\mathbf{C} = (x_c, y_c, z_c)^T$	Ponto no sistema de coordenadas de câmera
$\mathbf{u} = (u, v)^T$	Projeção de \mathbf{C}
$\mathbf{I} = (u, v)^T$	Ponto na imagem
\mathbf{R}	Matriz de rotação de câmera
f	Distância focal da câmera
s_u, s_v	Dimensão do <i>pixel</i>
h	Altura da câmera com relação ao eixo x
α, β, γ	Ângulos de Euler para o sistema de coordenadas de câmera
W	Largura da pista
r_e	Região retangular externa em coordenadas de imagem
r_i	Região retangular interna em coordenadas de imagem
$m(t)$	Medida baseada na fração de <i>pixels</i> no interior de r_e
T	Período de $m(t)$
L	Comprimento entre o início de duas marcas longitudinais adjacentes
\mathbf{v}	Velocidade do veículo
$R_{mm}(\tau)$	Função de autocorrelação de $m(t)$
ℓ_e	Segmento de linha horizontal externo em coordenadas de mundo
ℓ_i	Segmento de linha horizontal interno em coordenadas de mundo
p_{max}	Número de segmentos de linha horizontais
$I_e^p(v)$	Imagem projetada a partir de r_e^p
$I_e'(v)$	Derivada de $I_e^p(v)$
$g_p(v)$	Filtro Gaussiano
δ_z	Espaçamento entre os segmentos as linhas de varredura

δ_x	Largura das linhas de varredura
δx_s	Distância entre as linhas de varredura
σ_w	Desvio padrão do filtro Gaussiano
ε	Limiar de rejeição para picos espúrios
$\epsilon(\rho)$	Limiar de rejeição para a distância de Mahalanobis
r_q	Região retangular de busca q por sinalização vertical
f_s	Fator de escala da região r_q
\mathcal{D}_q	Conjunto com detecções da sinalização vertical em coordenadas de imagem
J_c	Coefficiente de Jaccard

LISTA DE FIGURAS

Figura 1.1 Sinalização horizontal e vertical de proibido ultrapassar	16
Figura 1.2 Tipos de diferentes marcas longitudinais horizontais.....	18
Figura 2.1 Tipos de sinalização vertical.....	26
Figura 3.1 Sistema de coordenadas de câmera e de mundo 3D	30
Figura 3.2 Projeção	33
Figura 3.3 Quadro típico e as regiões retangulares usadas para estimar o comprimento entre as marcas longitudinais adjacentes ao longo das bordas de pista detectadas (em verde).	35
Figura 3.4 Gráfico das funções $m(t)$, $R_{mm}(\tau)$, $s(u)$, $s_f(u)$ e $s'_f(u)$	36
Figura 3.5 Marcas longitudinais adjacentes.....	37
Figura 3.6 Gráfico de $s'_f(u)$: tachão refletivo.....	38
Figura 3.7 Gráficos da sequência de vídeo <i>Cambará</i>	43
Figura 3.8 Quadro da sequência de vídeo <i>Germany</i>	44
Figura 4.1 Diagrama esquemático do classificador em cascata.	47
Figura 4.2 Gráficos da função $m(t)$ para diferentes tipos de marcações de pista.....	48
Figura 4.3 Gráfico da autocorrelação da função $R_{mm}^t(\tau)$ para diferentes tipos de marcações.....	49
Figura 4.4 Exemplo de evidências de marcações de pista $\rho_b^t(u)$	51
Figura 4.5 Distribuição das amostras do atributo f_2 ao longo das cinco classes ω_i	53
Figura 4.6 Exemplos de diferentes tipos de marcações contidas no <i>dataset</i> 1.	56
Figura 4.7 <i>Ground truth</i> e os resultados obtidos para as três sequências de vídeo usadas no conjunto de testes 2.....	57
Figura 4.8 Quadros da sequência de vídeo <i>RS-287 (2)</i>	59
Figura 4.9 Quadros da sequência de vídeo <i>RS-040 (1)</i>	60
Figura 4.10 Projeção das bordas de pista do sistema de coordenadas de imagem (ICS) para o sistema de coordenadas de mundo (WCS). Definição dos segmentos de linha em coordenadas de mundo. Reprojeção dos segmentos de linha para coordenadas de imagem.....	61
Figura 4.11 (a) Imagem sem marcação. (b)-(c) Imagem com uma e duas marcações, respectivamente. (d)-(f) Linha horizontal unidimensional (1D) na escala de cinza ao longo do centro da imagem. (g)-(i) Respectivas derivadas de cada linha horizontal da imagem.	62
Figura 4.12 Efeitos do uso de um filtro Gaussiano: (a) Segmento horizontal 1D com valor adequado de σ : original (azul) e filtrado (vermelho). (b) Segmento horizontal 1D com valor alto de σ , fundindo as bordas vizinhas em uma única borda.	63
Figura 4.13 Exemplo da extração de picos ao longo das bordas da pista. (a) imagem contendo a marcação contínua-dupla, (b) intensidades original (azul) e filtrada (vermelho), (c) derivada da intensidade da imagem filtrada e pares de picos detectados. (d)-(f) apresenta as imagens análogas, mas para uma imagem sem marcação.	65
Figura 4.14 Imagem com marcação de pista: (a) imagem RGB, (b) canal R, (c) canal G, (d) canal B.	66
Figura 4.15 Diagrama esquemático do classificador de dois estágios.	66
Figura 4.16 Distribuição das classes de f_1 para diferentes classes.....	68
Figura 4.17 Estimando as ordenadas do máximo e mínimo local.	69
Figura 4.18 Exemplos de quadros com erros de classificação devido a problemas com o rastreador de bordas de pista.....	74

Figura 4.19 Exemplos das limitações enfrentadas pelo método	74
Figura 4.20 Quadros da sequência clip_watec_DMA311013_HMS145905 sem atrasos na detecção durante as transições.	75
Figura 4.21 Quadros da sequência clip_iPad_DMA190414_1.....	76
Figura 4.22 Comparação das duas técnicas de detecção de marcação horizontal: sem e com calibração.	77
Figura 5.1 Posicionamento na via rural da sinalização vertical.	80
Figura 5.2 Definição das regiões de busca retangulares em coordenadas de mundo (esquerda). Reprojecção das regiões na imagem (direita).	81
Figura 5.3 Amostras do conjunto de dados de treinamento do classificador SVM. (a)-(d) Amostras positivas. (e)-(h) Amostras negativas.	82
Figura 5.4 Regiões de busca r_q (em amarelo).	83
Figura 5.5 (a) Falso negativo no quadro 680 do vídeo clip_i5s_0789 (nosso método). Região retangular de <i>ground truth</i> na cor verde. (b)-(c) Falso positivo no vídeo clip_i5s_0094 (HOG + SVM com varredura em toda imagem com múltiplas escalas e com fator de mudança de escala igual a 1.05).	86
Figura 5.6 Exemplos de placas de proibido ultrapassar detectadas no vídeo clip_i5s_0094. <i>Ground truth</i> na cor verde e a detecção na cor vermelha.....	87
Figura A.1 Visão geral de uma aplicação de AR	99
Figura A.2 Quadros ilustrando a inserção de sinalização horizontal (1)	100
Figura A.3 Quadros ilustrando a inserção de sinalização horizontal (2)	100
Figura A.4 Quadros ilustrando a inserção de sinalização vertical	101
Figura A.5 Quadros ilustrando a inserção de sinalização horizontal e vertical.	101

LISTA DE TABELAS

Tabela 1.1 Acidentes fatais em 2008.....	15
Tabela 1.2 Número de acidentes por tipo em rodovias federais.....	17
Tabela 3.1 Resultados da calibração	44
Tabela 4.1 Sequências de vídeo	54
Tabela 4.2 Trechos de vídeo.....	55
Tabela 4.3 Matriz de confusão para o conjunto de testes 1.....	55
Tabela 4.4 Matriz de confusão para o vídeo <i>RS-287 (2)</i>	58
Tabela 4.5 Matriz de confusão para o vídeo <i>Suburban (4)</i>	58
Tabela 4.6 Matriz de confusão para o vídeo <i>RS-040 (1)</i>	59
Tabela 4.7 Tempo médio de execução.....	59
Tabela 4.8 Clipes: parâmetros de calibração de câmera	72
Tabela 4.9 Sumário das variáveis e valores padrão	72
Tabela 4.10 Resolução, taxa de amostragem e acurácia global dos clipes usados.....	73
Tabela 4.11 Matriz de confusão para os clipes 3-12	76
Tabela 4.12 Tempo médio de execução.....	77
Tabela 5.1 Clipes: parâmetros de calibração de câmera	85
Tabela 5.2 Sumário das variáveis e valores padrão	85
Tabela 5.3 Resolução, taxa de amostragem e acurácia global dos clipes usados.....	86
Tabela 5.4 Matriz de confusão para os clipes 1-3	86
Tabela 5.5 Acurácia média e tempo de execução.....	88

SUMÁRIO

1 INTRODUÇÃO	14
1.1 Motivação	16
1.2 Objetivos	17
1.2.1 Objetivos gerais.....	17
1.2.2 Objetivos específicos.....	18
1.3 Contribuições	18
1.4 Organização da tese	19
2 REVISÃO BIBLIOGRÁFICA	20
2.1 Calibração da câmera	20
2.2 Detecção do tipo de marcação horizontal	21
2.3 Detecção do tipo de sinalização vertical	25
3 CALIBRAÇÃO DA CÂMERA	29
3.1 Formulação do problema	29
3.2 Obtendo o <i>Pitch</i> e o <i>Yaw</i>	31
3.3 Obtendo a altura da câmera: primeira abordagem	32
3.4 Obtendo a altura da câmera: segunda abordagem	37
3.5 Melhorando a estimativa com informação temporal	39
3.6 Análise de erros	40
3.7 Resultados	41
3.8 Conclusões	44
4 DETECÇÃO DO TIPO DE MARCAÇÃO HORIZONTAL	46
4.1 Análise da sinalização horizontal usando câmera não calibrada	46
4.1.1 Classificador \mathcal{C}_1	47
4.1.2 Classificador \mathcal{C}_2	50
4.1.3 Classificador \mathcal{C}_3	52
4.1.4 Classificador \mathcal{C}_4	52
4.1.5 Resultados	53
4.2 Análise da sinalização horizontal usando câmera calibrada	60
4.2.1 Classificação das marcações de pista.....	65
4.2.2 Classificador \mathcal{C}_1	67
4.2.3 Classificador \mathcal{C}_2	68
4.2.4 Melhorando os resultados da classificação	69
4.2.4.1 Detecção de <i>outliers</i>	69
4.2.4.2 Informação temporal.....	70
4.2.5 Resultados	71
4.3 Conclusões	78
5 DETECÇÃO DO TIPO DE MARCAÇÃO VERTICAL	79
5.1 Resultados	84
5.2 Conclusões	89
6 CONCLUSÕES E CONTINUIDADE DO TRABALHO	90
REFERÊNCIAS	92
APÊNDICE A — APLICAÇÃO: SINALIZAÇÃO VIRTUAL HORIZONTAL E VERTICAL	99
APÊNDICE B — LISTA DE PUBLICAÇÕES	102

1 INTRODUÇÃO

A segurança no trânsito é uma preocupação constante nos dias atuais, particularmente em países subdesenvolvidos e em desenvolvimento.

De acordo com o relatório sobre segurança rodoviária, publicado pela Organização Mundial da Saúde (WHO - *World Health Organization*) (WHO, 2013), ocorreram 1,24 milhões de mortes no trânsito em 2010, sendo que 80% dessas mortes são provenientes de países com baixa e média renda.

O Brasil está sendo considerado como um país emergente no âmbito mundial por possuir imensos recursos naturais e um forte potencial de desenvolvimento e produção industrial. Segundo a Agência Nacional de Transportes Terrestres (ANTT, 2010), o Brasil possui uma malha rodoviária de aproximadamente 1,7 milhões de quilômetros, sendo 146 mil quilômetros entre rodovias federais e estaduais asfaltadas. Com um dos maiores Produto Interno Bruto do mundo, com a quinta maior área territorial do planeta e sendo o quinto maior contingente populacional do mundo, usufrui das rodovias como principal meio de transporte de cargas e de passageiros.

Um estudo apresentado em 2009, por Vasconcellos e Sivak (2009), mostra que em 2007 mais de 33% das rodovias brasileiras foram consideradas ruins ou muito ruins com relação ao seu estado geral (sinalização, pavimentação e geometria). Ademais, que o número de mortes em acidentes de trânsito no Brasil em 2006 foi maior que 35.000. No ano de 2010, o número de mortes de trânsito passou para 43.869 (WHO, 2013).

O relatório anual de segurança de rodovias (IRTAD, 2011), apresentado em 2011 no *International Traffic Safety Data and Analysis Group* (IRTAD), mostra que as mortes no trânsito estão diminuindo em países que pertencem ao IRTAD (na sua maioria, países em desenvolvimento), levando-se em consideração reduções significativas no número de mortes nas estradas com relação ao levantamento feito em anos anteriores (2008 e 2009). Ao contrário de outros países, um estudo realizado pela Confederação Nacional de Municípios do Brasil (CNM, 2009) mostra a alta taxa de mortalidade nas estradas brasileiras. A taxa de mortes por acidente no Brasil é 2,5 vezes maior que nos Estados Unidos, e 3,7 vezes maior que na Europa. (Veja a Tabela 1.1.) Conforme citado pelo IRTAD, muitos países estão adotando estratégias de segurança nas estradas para reduzir colisões frontais, traseiras e laterais. Estes tipos de acidentes são a principal razão de acidentes de trânsito seguido com pelo menos uma morte.

Na comunidade de pesquisa em visão computacional, tem sido realizados esforços significativos nos últimos anos para o desenvolvimento de novas técnicas e sistemas baseados

¹Taxa de mortalidade por 100.000 habitantes.

Tabela 1.1 – Acidentes fatais em 2008.

País	Mortes	População (milhões)	Taxa de mortalidade ¹
Brasil	57.116	189,6	30,1
Estados Unidos	37.261	304,0	12,5
União Europeia	38.876	498,0	7,8

Fonte: CNM (2009).

em visão para estradas e veículos inteligentes (BERTOZZI et al., 2002; BUCH; VELASTIN; ORWELL, 2011). Algumas delas estão diretamente focadas na detecção de pedestres (GERÓNIMO et al., 2010) e ciclistas; outras, na detecção ou saída de pista (MCCALL; TRIVEDI, 2006).

Estes sistemas normalmente empregam uma câmera de vídeo (e/ou outros sensores) instalada no interior do veículo (comumente direcionada para frente), com o propósito de escanear o ambiente e prover informações úteis ao motorista, tal como: limite de velocidade, sinalização, obstáculos, etc. Para tanto, o uso de câmeras calibradas permite o mapeamento de coordenadas de imagem para coordenadas de mundo, tal que esta informação (como a distância entre as linhas de bordo² da pista, a distância com relação a outros veículos ou pedestres) possa ser obtida com a intenção de ajudar o condutor na execução de uma direção defensiva. Também, o uso de câmeras calibradas é necessário para cenários de Realidade Aumentada (AR - *Augmented Reality*), os quais proveem uma combinação de informações do mundo real e conteúdos virtuais pela sobreposição de objetos gráficos no mundo real (AZUMA, 1997). Um dos primeiros trabalhos que apresentou o uso de AR na área automotiva (simulador de direção veicular) foi proposto por Tonnis et al. (2005).

Em aplicações onde uma câmera portátil é fixada no para-brisa (como *smartphones*), um procedimento automático para a calibração de câmera é muito importante para sistemas de assistência ao motorista. Uma vez que os parâmetros intrínsecos podem ser obtidos da especificação da câmera ou por um método de calibração *offline*, como por exemplo o *Camera Calibration Toolbox for Matlab* (BOUGUET, 2008).

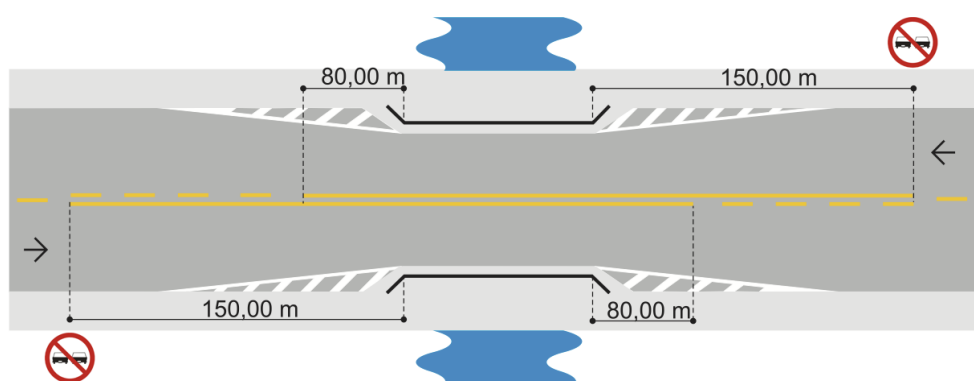
O desenvolvimento de sistemas baseados na detecção e de alerta de alteração de pista (HILLEL et al., 2014) também foram importantes para chamar a atenção do condutor que o veículo inicia um movimento de saída de pista. Salienta-se a importância destes sistemas, uma vez que esta manobra é comum em estradas com fluxos opostos e, conseqüentemente, potencialmente perigosas por haver a possibilidade de colisão frontal. Neste contexto, a detecção e reconhecimento de marcas longitudinais horizontais é um problema importante, visto que

²As linhas de bordo delimitam, através de linha contínua, a parte da pista destinada ao deslocamento dos veículos, estabelecendo seus limites laterais.

indicam partes da estrada onde a alteração lateral do veículo é ou não permitida.

Também, a detecção e reconhecimento de placas de trânsito (sinalização vertical) vem a complementar e reforçar a sinalização horizontal, advertindo os condutores sobre condições com potencial risco existente na via ou nas suas proximidades. Para o tipo de acidente colisão frontal, o sinal de proibido ultrapassar deve ser adotado, assinalando ao motorista que é proibido realizar o movimento de ultrapassagem no trecho regulamentado pelas faixas destinadas ao sentido oposto de circulação (Figura 1.1).

Figura 1.1 – Sinalização horizontal e vertical de proibido ultrapassar.



Fonte: CONTRAN (2007b).

1.1 Motivação

De acordo com as informações do Ministério da Saúde - DATASUS (SUS, 2015), o número de acidentes fatais no Brasil em 2013 foi de aproximadamente 43 mil óbitos. No âmbito das rodovias federais e conforme apresentado na Tabela 1.2, observa-se que aproximadamente 31,7% dos acidentes são provenientes de colisões frontais, abalroamento (no mesmo sentido e sentido oposto) e saída de pista. O acidente do tipo abalroamento se caracteriza por uma colisão entre veículos motorizados que trafegam no mesmo sentido, ou no sentido oposto, quando o impacto se dá entre as laterais dos veículos envolvidos. Ainda, esse percentual não inclui os acidentes do tipo saída de pista, que se dá pelo fato do veículo projetar-se para fora da área destinada ao tráfego, sem que tenha colidido, tombado ou capotado. O elevado percentual de acidentes dos três tipos supracitados reforça a necessidade e os motivos de prosseguirem com as pesquisas no campo de visão computacional em Sistemas de Assistência ao Motorista, sobretudo no que se refere a identificação da sinalização horizontal e vertical da pista. O emprego

Tabela 1.2 – Número de acidentes por tipo em rodovias federais

Tipo de acidente	Total
Choque com objeto fixo	19.222
Capotagem	4.513
Atropelamento	6.486
Atropelamento de animal	4.286
Choque com veículo estacionado	1.886
Colisão traseira	51.355
Abalroamento no mesmo sentido	23.393
Colisão frontal	5.312
Abalroamento em sentido oposto	4.609
Abalroamento transversal	19.805
Tombamento	4.887
Saída de pista	24.648
Outros tipos	5.381
Atropelamento e fuga	1.779
Queda de veículo	5.338
Total	182.900

Fonte: DNIT (2010).

desses sistemas pode contribuir para melhorar o processo de condução dos motoristas, sinalizando e alertando-os sobre potenciais problemas, evitando prováveis colisões e consequentemente aumentando a segurança nas rodovias.

1.2 Objetivos

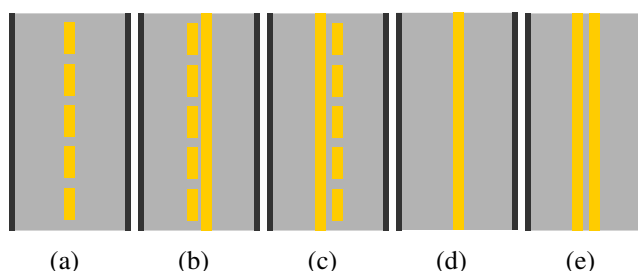
1.2.1 Objetivos gerais

O presente trabalho tem por objetivo geral desenvolver algoritmos de visão computacional para veículos inteligentes, com potencial aplicação em sistemas de assistência ao motorista (DAS - *Driver Assistance Systems*) ou inspeção de rodovias. O foco consiste na proposta de um novo método de calibração automática de câmera, da extração e classificação de características da rodovia (sinalização horizontal e vertical), visando verificar diferentes marcas longitudinais horizontais e sua conformidade e complementação com a sinalização vertical, contribuindo para que os condutores adotem comportamentos adequados, de forma a aumentar a segurança e consequentemente reduzir o número de acidentes.

1.2.2 Objetivos específicos

1. Calibração da câmera: desenvolver uma abordagem automática e dinâmica para estimar os parâmetros extrínsecos de uma câmera veicular embarcada, não necessitando de qualquer informação a priori com relação a pista (como o comprimento ou largura de marcas da pista) bem como algum padrão de calibração.
2. Detecção e identificação de marcas longitudinais horizontais: desenvolver um sistema que possa classificar cada parte da estrada, também por uma câmera veicular embarcada, em cinco tipos de marcas longitudinais horizontais, como apresentado na Figura 1.2.
3. Detecção e identificação de sinalização vertical: desenvolver um sistema que possa identificar e classificar sinais verticais (placas fixadas na posição vertical, ao lado da pista,), em específico o sinal R-7³ (proibido ultrapassar).

Figura 1.2 – Tipos de diferentes marcas longitudinais horizontais: (a) Seccionada. (b) Seccionada-contínua. (c) Contínua-seccionada. (d) Contínua simples. (e) Contínua dupla.



Fonte: do autor (2015).

1.3 Contribuições

A seguir, são apresentadas as principais contribuições da tese:

- Desenvolvimento de um método de calibração automática (parâmetros extrínsecos) de câmera sem informação prévia.
- Desenvolvimento de uma técnica que realiza a detecção e classificação da marcação horizontal em cinco tipos.

³Código definido pelo Conselho Nacional de Trânsito (CONTRAN, 2007b)

- Desenvolvimento de uma abordagem para explorar a relação de imagem \times mundo no contexto de detecção/reconhecimento da sinalização vertical, focando em reduzir o espaço de busca e evitar falsos-positivos que não possuam uma escala coerente.
- Criação e disponibilização de bases de dados com anotações manuais.

1.4 Organização da tese

A tese está organizada em 6 capítulos. O Capítulo 2 apresenta as principais abordagens referente de trabalhos relacionados ao procedimento de calibração de câmera, bem como à técnicas que realizam a detecção e reconhecimento da sinalização horizontal e vertical. O Capítulo 3 introduz umas das principais contribuições da tese: uma abordagem de calibração automática de câmera explorando a geometria da pista sem informação prévia, possibilitando estimar os parâmetros extrínsecos de câmera bem como a sua altura com relação ao solo. No Capítulo 4 são apresentadas duas abordagens que realizam a análise (detecção e classificação) da sinalização horizontal utilizando a câmera não calibrada e a câmera calibrada, respectivamente. O Capítulo 5 usufrui da informação de câmera calibrada para restringir a região de busca para realizar a detecção e reconhecimento da sinalização vertical bem como acelerar o processo dos detectores existentes. Por fim, o Capítulo 6 sumariza as principais contribuições e apresenta algumas linhas para pesquisas futuras. Ainda, no Apêndice A é apresentada uma aplicação de realidade aumentada usando a informação de câmera calibrada para a inclusão de sinalização virtual horizontal e vertical, e no Apêndice B é apresentada a lista de publicações.

2 REVISÃO BIBLIOGRÁFICA

Nas próximas seções serão apresentados alguns trabalhos relacionados ao processo de calibração de câmera, bem como técnicas de detecção e reconhecimento da sinalização horizontal e vertical. Esta revisão servirá de premissa para a definição das abordagens propostas neste trabalho.

2.1 Calibração da câmera

O problema geral referente a calibração de câmeras tem sido estudado de forma exaustiva para aplicações em visão computacional, realidade virtual e realidade aumentada. Um exemplo clássico é o algoritmo de Tsai (TSAI, 1987), o qual pode ser usado para obter a configuração completa dos parâmetros de uma dada câmera num ambiente controlado. Wang e Tsai (1991) exploraram as linhas de fuga (conhecidas como *vanishing lines*) em uma cena com o intuito de obter os parâmetros de câmera com base num padrão hexagonal fixado no plano do chão. Douchamps e Chihara (2009) apresentaram uma nova abordagem para calibração de câmera com maior acurácia, reportando erros na ordem de 0.002 *pixels* quando marcas planares são utilizadas. Uma desvantagem evidente nos métodos supracitados é que um padrão de calibração torna-se necessário, enquanto um dos objetivos do presente trabalho é realizar uma calibração *online* sem o uso de marcas artificiais.

Há algumas abordagens focadas na calibração de câmeras veiculares embarcadas. O método descrito em Bellino, Merenda e Kolski (2005) faz o uso de estruturas geométricas em veículos parados. Embora este procedimento de calibração seja realizado na linha de produção, o mesmo é bastante suscetível a pequenas alterações da posição da câmera, bem como modificações no veículo, como por exemplo, a troca de pneus. Portanto, a calibração deve ser realizada com uma certa periodicidade. Em Hold et al. (2009), os autores apresentam uma abordagem para obter todos os parâmetros extrínsecos da câmera com alta acurácia em sistemas de alerta de saída de pista, incluindo o tratamento de distorção radial. Contudo, Bellino, Merenda e Kolski (2005) e Hold et al. (2009) necessitam de um padrão de calibração artificial. Consequentemente, o processo é realizado de forma *offline*.

Tan et al. (2011) apresentaram uma abordagem interativa para calibrar os parâmetros extrínsecos de câmeras veiculares embarcadas. Eles exploraram a geometria de três retas (duas paralelas com relação ao veículo e uma perpendicular as outras duas) para derivar uma solução analítica com o propósito de obter os parâmetros extrínsecos. Embora os resultados obtidos

pelos autores sejam bons, a técnica sugerida necessita de uma intervenção manual.

Guiducci (2000) apresentou uma formulação matemática para a calibração de câmera de forma *online* com base nos parâmetros relacionados a pista (como a largura da estrada, comprimento e espaçamento entre duas marcas longitudinais seccionadas/tracejadas). No entanto, o autor assume que os parâmetros de pista sejam conhecidos *a priori*.

As técnicas descritas em Ribeiro, Dihl e Jung (2006) e Jung, Saldanha e Jr. (2009) também exploram a geometria da pista com o propósito de realizar a calibração automática da câmera. Em Ribeiro, Dihl e Jung (2006), um modelo projetivo simplificado foi utilizado, cujo objetivo principal foi o de alertar os condutores no caso de saída de pista. Em Jung, Saldanha e Jr. (2009), os parâmetros de câmera foram encontrados de forma explícita (assumindo algumas restrições com relação aos parâmetros extrínsecos), com o objetivo em aplicações de realidade aumentada. Percebe-se, entretanto, que ambas as abordagens necessitam de uma estimativa da largura de pista, que deve ser conhecida antecipadamente.

Kanhere e Birchfield (2010) propuseram uma taxonomia e análise de vários métodos de calibração de câmera para o monitoramento de tráfego. Embora eles não tenham avaliado câmeras embarcadas, a maioria dos métodos exploram a geometria da estrada para a calibração e, ainda, necessitam de alguma informação (parâmetro) relacionada a pista. Zhang et al. (2012) realizaram uma abordagem para a calibração de câmeras de monitoramento de tráfego com base no movimento dos veículos e pedestres. No entanto, a técnica também necessita de uma estimativa prévia da altura da câmera como entrada.

Em síntese, algumas abordagens empregam padrões de calibração de forma manual *offline* ou necessitam da interação com o usuário (BELLINO; MERENDA; KOLSKI, 2005; HOLD et al., 2009; TAN et al., 2011). Outros métodos exploram a geometria da estrada (GUIDUCCI, 1999; GUIDUCCI, 2000; RIBEIRO; DIHL; JUNG, 2006; JUNG; SALDANHA; JR., 2009), mas necessitam de alguma informação *a priori* de um ou mais parâmetros relacionados a pista, como citado anteriormente, a largura da pista ou o comprimento e a medida de espaçamento entre duas marcas longitudinais seccionadas.

2.2 Detecção do tipo de marcação horizontal

Ultrapassagem ou deslocamento lateral (troca de pista) é uma das manobras mais perigosas a ser realizada no trânsito e tem sido extensivamente estudada para aplicações em visão computacional com o intuito de ajudar os motoristas durante a condução. Durante uma viagem, por exemplo, um motorista pode encontrar um número de diferentes tipos de marcas

longitudinais¹, incluindo as simples (seccionada/tracejada² ou contínua) ou duplas (seccionada e contínua, contínua e seccionada ou apenas contínua), conforme pode ser visto na Figura 1.2.

De acordo com o relatório do Laboratório de Pesquisa em Transporte (VISVIKIS et al., 2008), há dois tipos diferentes de sistemas de alerta de alteração de pista: (i) sistema de assistência de alteração de pista (LCA - *Lane Change Assistance*): é um tipo de sistema que alerta o motorista contra possíveis colisões que possam ocorrer devido a uma manobra de alteração de pista; e (ii) sistema de alerta de alteração de pista (LDW - *Lane Departure Warning*), que alerta o motorista quando uma alteração de pista não intencional está para ocorrer. Estes sistemas também monitoram os tipos de marcações e, conseqüentemente, um bom sistema de alteração de pista é aquele que consegue detectar e identificar diferentes tipos de marcações.

Embora existam várias abordagens e propostas para o rastreamento de pista (HILLEL et al., 2014), não há muitos trabalhos dedicados à identificação dos tipos de marcações.

A análise das marcações de pista baseada em visão computacional é um importante componente em Sistemas de Assistência ao Motorista (DAS) e na localização de veículos autônomos. O problema geral de detectar e seguir/rastrear as bordas da pista tem sido estudado extensivamente pela comunidade de visão, e uma vasta variedade de modelos de detecção e rastreamento de pista tem sido propostos nos últimos anos, como apresentado recentemente por Yenikaya, Yenikaya e Düven (2013) e Hillel et al. (2014). Entretanto, pouco esforço tem sido dedicado ao problema de classificação dos tipos de marcações horizontais de pista, principalmente pela diversidade das marcas. A seguir, são apresentadas algumas abordagens que de alguma forma procuram resolver a classificação das marcações de pista.

Lindner et al. (2010) apresentam uma abordagem multinível para extrair e classificar as marcações de pista. Para o sistema proposto, as imagens são capturadas pela câmera frontal do veículo conceito Carai³. Num primeiro momento, um detector de bordas pré-processa a imagem em escala de cinza e subsequentemente um conjunto de algoritmos busca por um grupo de quatro objetos: linhas, curvas, curvas paralelas e objetos fechados. Com base no conjunto de objetos fechados, um algoritmo de extração de características calcula algumas informações acerca da geometria e ajusta uma elipse aos pontos da borda. As características extraídas são o número de bordas, a direção do eixo maior, largura e comprimento. Uma árvore de decisão binária com quatro classes de saída realiza o processo de classificação. Os autores escolheram as seguintes marcações (todas seccionadas): marca longa grossa, marca longa fina, marca curta

¹As marcas longitudinais separam e ordenam as correntes de tráfego.

²Tracejada ou Seccionada: corresponde às linhas interrompidas, aplicadas em cadência, utilizando espaçamentos com extensão igual ou maior que o traço;

³<<http://www.carai.eu/>>

grossa e marca curta fina.

McCall et al. (2005) apresentaram uma abordagem para unificar a detecção de múltiplos tipos de marcações (originalmente sugerido por Gehrig et al. em 2002) incluindo marcas do tipo *bott's dots*⁴. Entretanto, as múltiplas marcas citadas pelos autores não levam em consideração a classificação, apenas a detecção e rastreamento.

Schubert, Schulze e Wanielik (2010) propuseram um sistema de avaliação de mudança de pista. A abordagem sugerida percebe o ambiente no entorno do veículo, avalia a situação do trânsito, reconhece algumas ocasiões apropriadas e recomenda a manobra de mudança de pista ao motorista. O algoritmo de avaliação, responsável em decidir se uma alteração de pista é possível, classifica o tipo de marcação de pista em seccionado ou contínuo. Com base no rastreamento das bordas da pista, um sinal amostrado no espaço discreto é derivado e usado como entrada para a Transformada Rápida de Fourier (FFT - *Fast Fourier Transform*). O espectro de potência é então calculado e o tipo de marcação de pista é classificado. Collado et al. (2005) também empregam uma abordagem baseada no espectro de potência, mas diferentemente de Schubert, Schulze e Wanielik (2010), distinguem um terceiro tipo de marcação de pista, conhecido como misto (seccionado-contínuo ou contínuo-seccionado, mas sem distinção entre elas).

Alguns métodos exploram o Mapeamento Perspectivo Inverso (IPM - *Inverse Perspective Mapping*) para melhor extrair as marcações de pista. Na área de Veículos Inteligentes (IV - *Intelligent Vehicles*), o IPM foi utilizado para detectar e rastrear as bordas de pista, como nos trabalhos de Muad et al. (2004) e Borkar, Hayes e Smith (2009). O IPM é uma técnica de transformação geométrica que projeta cada pixel da vista em perspectiva de um objeto 3D, mapeando cada *pixel* para uma nova posição, construindo uma nova imagem 2D removendo o efeito de perspectiva. O resultado produzido é conhecido por vista de cima (*bird-eye view*).

Em Collado et al. (2006), os autores dividem o problema em cinco etapas: realizam o IPM para obter uma vista de cima da pista, segmentam os *pixels* que pertencem as marcas longitudinais da rodovia, extraem as bordas da pista utilizando a transformada de Hough e realizam alguns ajustes no ângulo de arfagem (mais conhecido como *pitch*) da câmera. Após este processo, as bordas detectadas são classificadas em três tipos a partir do espectro de potência oriundo da FFT: em contínua (também conhecida por sólida), seccionada, ou mista (seccionada-contínua ou contínua-seccionada, sem distinção entre ambas). Este método necessita de uma calibração prévia da câmera para que o procedimento inicial possa gerar a vista de cima. Recentemente, Satzoda e Trivedi (2013) também propuseram um método para extrair as caracte-

⁴Tachões não refletivos do tipo calota.

rísticas das marcações de pista para soluções embarcadas, usando o IPM. Com base na vista de cima da pista, os autores usaram “bandas de varredura” com largura configurável para predizer a existência de marcações de pista. Dentro destas bandas/regiões, os níveis de intensidade da imagem são filtrados por um filtro direcional de tamanho 5×5 com a finalidade de detectar variações horizontais. Logo após, uma função de limiarização é aplicada juntamente com um esquema de deslocamento e casamento, com a intenção de detectar evidências de marcações na pista. A técnica foi avaliada com a base de dados LISA-Q, composta por cinco diferentes tipos de estradas: *freeway*, *freeway* com superfície de concreto, vias urbanas com sombras, *freeway* com veículos e *freeway* com tachões circulares. Três tipos de marcações são apresentados nesta base dados: seccionada, contínua e com tachões circulares.

No trabalho proposto por Nedevschi et al. (2013), foi realizada a introdução de histogramas periódicos para determinar o tipo de marcação de pista com a finalidade de melhorar a localização global de veículos autônomos nas proximidades de cruzamentos. Embora o foco principal seja o posicionamento global do veículo, também foi desenvolvido um algoritmo que classifica as bordas da pista em quatro tipos: contínuo simples, contínuo duplo, seccionado e misto.

Ainda, Suchitra, Satzoda e Srikanthan (2013) apresentaram uma abordagem modular para detectar e distinguir três tipos principais de marcações: seccionada, contínua e zigzag. Para tanto, eles definiram uma unidade de processamento chamada Marcação de Pista Básica (BLM - *Basic Lane Marking*), que particularmente lida com as marcações seccionada e contínua com base em propriedades de continuidade. O processo se resume em dividir um quadro da imagem em blocos e a partir das bordas da pista e da existência de alguma marcação (oriunda do BLM), um mapa de bordas é criado.

A técnica proposta por Lopez et al. (2005) explora uma medida de pico para detectar os diferentes tipos de marcações de pista. A marca do tipo seccionada deve conter valores altos ao longo da linha de divisão central de uma dada região de interesse, e valores baixos próximos as bordas da pista. O artigo apresenta resultados com a presença de marcas contínuas e seccionadas. Entretanto, o mesmo não trata, ou seja, não diferencia marcações mistas, as quais são comuns em estradas com fluxo de veículos em direções opostas.

Outra abordagem para realizar a extração do tipo de marcas, fundamentada na técnica de segmentação por limiar, foi proposta por Li et al. (2012). Segundo apresentado, o modelo, inicialmente, também cria a vista de cima da pista, segmenta a imagem transformada através de um limiar adaptativo, e finalmente algumas restrições geométricas são aplicadas para remover prováveis respostas falsas. A saída gerada é uma imagem binária com potenciais marcações.

Entretanto, nenhuma característica em alto nível é extraída e classificada.

Chira, Chibulcutean e Danescu (2010) apresentam um sistema em tempo real para detectar, medir e classificar sinais horizontais (marcas longitudinais e inscrições no pavimento) usando a detecção de bordas e o casamento de padrões geométricos. O foco do referido trabalho é, na sua maioria, na inscrição do pavimento (setas indicativas de posicionamento na pista para a execução de movimentos: siga em frente, siga em frente ou vire à esquerda, siga em frente ou vire à direita), e uma única marca longitudinal (seccionada).

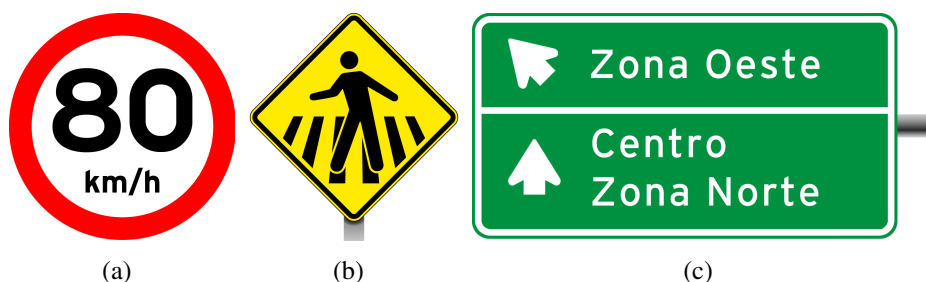
Também focado na inscrição do pavimento, Danescu e Nedevschi (2010) iniciam o processo de detecção das marcas e inscrições horizontais detectando as bordas da imagem e, posteriormente, realizam a segmentação utilizando mistura de Gaussianas. Por fim, a classificação é realizada com uma técnica de extração de cubóides combinado com o RANSAC (*RANdom SAmple Consensus*) (FISCHLER; BOLLES, 1981).

2.3 Detecção do tipo de sinalização vertical

Segundo o manual brasileiro de sinalização de trânsito (CONTRAN, 2007a), a sinalização vertical utiliza sinais apostos sobre placas fixadas na posição vertical, ao lado ou suspensas sobre a pista, com o intuito de transmitir mensagens de caráter permanente, ou variável, através de uma notação ou simbologia preestabelecida e aprovada por lei. A finalidade da sinalização vertical está em fornecer informações (orientação) aos condutores que os permitam adotar comportamentos adequados de modo a aumentar a segurança e ordenar os fluxos de tráfego. Esta padronização varia entre países. O Brasil possui diretrizes próprias e a sinalização vertical está subdividida em sinalização vertical de regulamentação (obrigações, limitações, proibições ou restrições que governam o uso da via), de advertência (advertir os condutores sobre condições com potencial risco existentes na via ou nas suas proximidades, tais como escolas e passagens de pedestres) ou de indicação (indicar direções, localizações) (Veja a Figura 2.1).

Os sistemas de reconhecimento de sinais de trânsito (*TSR - Traffic Sign Recognition*) fazem parte dos sistemas de auxílio ao motorista, possibilitando reconhecer placas de trânsito (como por exemplo, velocidade máxima permitida, proibido ultrapassar, parada obrigatória, etc.), e avisar o condutor acerca da sinalização presente na via. Segundo Mogelmoose, Trivedi e Moeslund (2012), nem todos os sinais de trânsito possuem a mesma capacidade de chamar a atenção do motorista. Ainda, o estudo reporta que normalmente os motoristas, de forma

Figura 2.1 – Tipos de sinalização vertical: (a) Regulamentação (b) Advertência (c) Indicação



Fonte: CONTRAN (2007a).

inconsciente, prestam mais atenção em sinais de limite de velocidade e são menos propensos a perceber os sinais de advertência de pedestres, tais como: trânsito de ciclistas, área escolar.

De acordo com Eichner e Breckon (2008), os sistemas de reconhecimento de sinais verticais de trânsito podem ser subdivididos em dois estágios: (i) detecção inicial de possíveis candidatos a sinais; (ii) reconhecimento e classificação do tipo de placa detectada. A detecção consiste em localizar os sinais nas imagens de entrada, enquanto que o reconhecimento e classificação incide em determinar qual é o tipo de sinalização. Além disso, muitos sistemas de reconhecimento de sinais de trânsito foram desenvolvidos para trabalhar com sequências de vídeo (na maior parte dos casos em tempo real). Por este motivo, os sinais devem ser rastreados ao longo do tempo e incluir alguma informação preditiva com o propósito de poder tratar oclusões, por exemplo, ao longo do rastreamento.

Para cada um dos estágios supracitados, várias abordagens foram propostas. Para o primeiro estágio (detecção), geralmente se divide a tarefa em dois tipos: detecção com base na cor e com base na forma. De acordo com Mogelmoose, Trivedi e Moeslund (2012), os métodos baseados em cor usufruem do fato de que os sinais de trânsito são projetados para serem facilmente distinguidos com o seu entorno, e sua identidade visual define cores fortes e contrastantes. Além disso, assim como possuem cores específicas, também possuem formas bem definidas, o que proporciona uma busca por figuras geométricas. Alguns trabalhos, como em Gomez-Moreno et al. (2010), subdividem a tarefa de detecção em segmentação (uso de informação de cor) e detecção (uso de formas geométricas). A notação varia muito entre os autores, mas o propósito final é o mesmo, e geralmente usam-se as palavras detecção e segmentação (com a mesma finalidade) como notação.

Vários métodos exploram a segmentação baseada na cor, particionando a imagem em regiões. Geralmente, a técnica consiste na limiarização num determinado espaço de cor. Um dos modelos de cor mais usados é o HSI (*hue, saturation, intensity* – matiz, saturação, inten-

sidade), pois separa o componente intensidade das informações de cores (matiz e saturação) em uma imagem colorida. Segundo Gonzalez e Woods (2007), o espaço de cores RGB é ideal para a geração de imagens coloridas (aquisição por câmera ou exibição em monitor), mas a sua utilização na descrição de cores é limitada. O modelo de cores HSI é usado como método de segmentação de sinais de trânsito em vários artigos, tais como, Kuo e Lin (2007), Nguwi e Kouzani (2008), Ren et al. (2009), Chiang et al. (2010), Qingsong, Juan e Tiantian (2010). De acordo com Mogelmoose, Trivedi e Moeslund (2012), outros trabalhos como, Timofte, Zimmermann e Gool (2009) e Prisacariu et al. (2010), utilizam o espaço RGB, mas adotam um limiar adaptativo para lidar com as variações de iluminação.

Alguns pesquisadores escolheram usar a segmentação com base em formas e, inicialmente, utilizam um detector de bordas para encontrar variações abruptas (locais) de intensidade. O detector de Canny (1986) é um dos mais usados, visto que possui como objetivo básico ter uma baixa taxa de erros, e está presente em vários sistemas de detecção de placas de trânsito, como em Houben (2011), Timofte, Zimmermann e Gool (2009), Ruta, Li e Liu (2010), Deguchi et al. (2011). Observa-se que os trabalhos de Timofte, Zimmermann e Gool (2009) e Gao et al. (2006) combinam a detecção com base na cor e na forma. Outra escolha para a detecção de bordas é o uso de Histogramas de Gradientes Orientados (HOG - *Histograms of Oriented Gradients*). O trabalho de Dalal e Triggs (2005) fez o uso do HOG para detectar pedestres em imagens estáticas. Posteriormente, Gao et al. (2006), Xie et al. (2009), Creusen et al. (2010), dentre outros, usaram a mesma técnica para detectar sinais verticais. Outros trabalhos, com Classificadores Neuraistambém foram usados por Nguwi e Kouzani (2008), Fistrek e Loncaric (2011), Zhang, Sheng e Li (2012).

Recentemente, Mogelmoose, Liu e Trivedi (2015) realizaram um estudo acerca da sinalização vertical definida no Estados Unidos, uma vez que a maioria dos trabalhos relacionados está focada em sinais verticais Europeus. Como há diferenças na definição da cor, geometria, posicionamento dos sinais de trânsito de um país para outro, o foco da pesquisa foi empregar a técnica de *Integral Channel Features* e *Aggregate Channel Features*, ambas propostas por Dollar et al. (2009) e (2014), com o intuito de avaliar o desempenho de ambas as técnicas com trabalhos relacionados.

Em suma, o processo de reconhecimento da sinalização vertical de trânsito envolve duas principais etapas: a detecção do sinal em uma imagem ou sequência de vídeo e o reconhecimento/classificação do sinal detectado. Nos trabalhos supracitados, várias abordagens são apresentadas para a detecção e classificação das placas de trânsito, mas dependendo da aplicação desejada, há uma classe de técnicas que em conjunto realizam a detecção e classificação

da placa de forma simultânea. Nestes casos, a ideia geral é treinar uma classificador que busque uma placa em específico, de modo que apenas aquele tipo de placa seja detectada; consequentemente, a detecção resulta no reconhecimento. Os trabalhos propostos por Viola, Jones e Snow (2005) e Dalal e Triggs (2005), que realizam a detecção e reconhecimento de pedestres servem de referência para a classe de técnicas que detectam e ao mesmo tempo reconhecem determinado objeto.

3 CALIBRAÇÃO DA CÂMERA

Neste capítulo serão apresentadas as técnicas que foram desenvolvidas para estimar os parâmetros extrínsecos de câmera bem como a sua altura com relação ao solo.

3.1 Formulação do problema

Embora o problema de calibração de câmeras tenha sido bastante estudado por vários pesquisadores, o escopo de aplicação do presente trabalho impõe algumas restrições com relação ao problema geral. Mais precisamente, foca-se em uma solução de baixo custo para câmeras veiculares embarcadas, na qual o usuário fixa a câmera na parte superior central de para-brisa¹.

Dado um ponto 3D $\mathbf{W} = (x_w, y_w, z_w)^T$ no sistema de coordenadas de mundo, o ponto correspondente $\mathbf{C} = (x_c, y_c, z_c)^T$ no sistema 3D de coordenadas de câmera é dado por uma transformação rígida da forma:

$$\mathbf{C} = \mathbf{R}(\mathbf{W} - \mathbf{W}_0), \quad (3.1)$$

onde \mathbf{W}_0 é a posição do sistema de coordenadas de câmera, e \mathbf{R} é a matriz de rotação.

A projeção correspondente \mathbf{C} em *pixels* de imagem $\mathbf{u} = (u, v)^T$, assumindo uma câmera *pinhole* é dada por:

$$u = \frac{f s_u x_c}{z_c}, \quad v = \frac{f s_v y_c}{z_c}, \quad (3.2)$$

onde f é a distância focal da câmera, s_u e s_v referem-se a dimensão do *pixel*, e $(u, v)^T$ são as coordenadas relativas ao eixo ótico da câmera (geralmente, o *pixel* central da imagem).

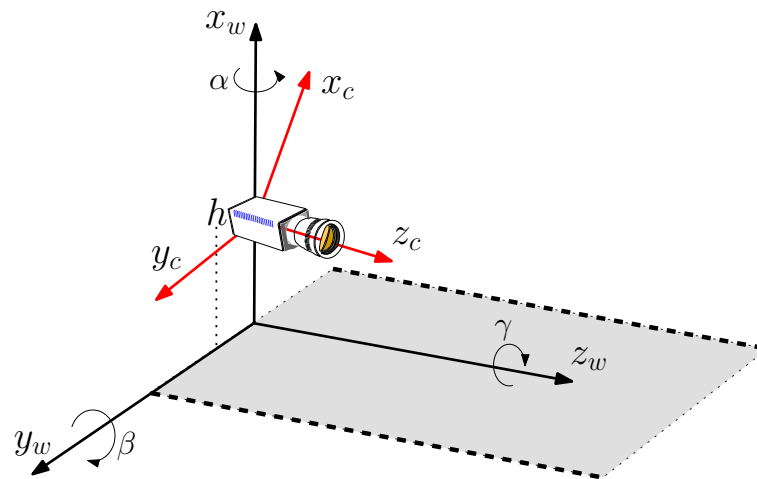
Os parâmetros f , s_u e s_v estão relacionados com o sistema de digitalização da câmera, e são os parâmetros intrínsecos (que podem ser computados previamente, e que permanecerão os mesmos). Por outro lado, \mathbf{R} e \mathbf{W}_0 dependem da configuração de câmera e são chamados de parâmetros extrínsecos da câmera. No escopo deste trabalho a câmera veicular embarcada pode ser retirada do para-brisa e recolocada noutro momento. Por este motivo, a cada colocação/fixação da câmera, geram-se novos valores para \mathbf{R} e \mathbf{W}_0 , os quais precisam ser recalculados.

A Figura 3.1 apresenta uma típica configuração para câmeras veiculares embarcadas. A câmera é posicionada em $(h, 0, 0)^T$ e os ângulos de Euler para o sistema de coordenadas de câmera, referente aos eixos x , y e z , são dados por α , β e γ , respectivamente. Estes ângulos são conhecidos por *yaw* (guinada), *pitch* (arfagem ou tangagem) e *roll* (rolamento, inclinação

¹Os efeitos de lente e prováveis distorções do para-brisa são desprezados.

lateral), respectivamente.

Figura 3.1 – Sistema de coordenadas de câmera e de mundo 3D.



Fonte: do autor (2015).

Na configuração ideal de câmera, não deve haver *yaw* e *roll* (apenas *pitch*). Entretanto, a instalação da câmera, realizada de forma manual, não é ideal. Para tanto, neste trabalho, considera-se que não exista *roll* efetivamente (uma vez que o movimento é normalmente evitado pelo para-brisa). A matriz de rotação correspondente é dada por:

$$\mathbf{R} = \begin{bmatrix} \cos \beta & -\sin \beta \sin \alpha & -\sin \beta \cos \alpha \\ 0 & \cos \alpha & -\sin \alpha \\ \sin \beta & \cos \beta \sin \alpha & \cos \beta \cos \alpha \end{bmatrix}. \quad (3.3)$$

A próxima seção apresenta uma abordagem para computar os parâmetros extrínsecos da câmera α , β e h com base nos parâmetros intrínsecos (assume-se que estes sejam conhecidos) e de algumas relações entre mundo-*pixels*, extraídas de marcas da pista e da geometria esperada da estrada.

Observa-se que a geometria da pista foi explorada para calibrar câmeras veiculares embarcadas em (RIBEIRO; DIHL; JUNG, 2006; JUNG; SALDANHA; JR., 2009). Entretanto, ambos os artigos assumem que a largura da pista seja previamente conhecida. Por outro lado, o único requisito da presente abordagem é que, durante o processo de calibração, o veículo deve se mover ao longo de um trecho plano da estrada em que as marcas longitudinais sejam seccionadas (com o mesmo espaçamento durante o processo de calibração) e que a velocidade do veículo seja conhecida.

3.2 Obtendo o *Pitch* e o *Yaw*

Assumindo que o veículo esteja se movendo paralelamente as bordas da pista, a equação que descreve as bordas esquerda e direita da pista em coordenadas de mundo são $y_w = y_0$ e $y_w = y_0 + W$, onde y_0 é a distância ortogonal da borda esquerda da pista a câmera e W é a largura da pista, como pode ser visto na Figura 3.2. As retas paralelas de borda da pista (em coordenadas de mundo) são mapeadas em retas que se interseccionam em coordenadas de câmera e o ponto de intersecção $\mathbf{v} = (v_u, v_v)^T$ é o ponto de fuga da cena. De fato, aplicando as Equações (3.1) e (3.2) em $y_w = y_0$ e $y_w = y_0 + W$, conduz às equações paramétricas das bordas da pista em coordenadas de imagem em função de z . Calculando o limite com $z \rightarrow \infty$ para cada uma das bordas da pista projetadas, tem-se as coordenadas do ponto de fuga desejado, como segue abaixo:

$$\mathbf{v} = \left(-f_u \tan \beta, -f_v \frac{\tan \alpha}{\cos \beta} \right)^T, \quad (3.4)$$

onde $f_u = f s_u$ e $f_v = f s_v$, os quais são os parâmetros intrínsecos da câmera.

Se $\mathbf{v} = (v_u, v_v)^T$ são conhecidos, α e β podem ser obtidos através das seguintes expressões:

$$\beta = -\tan^{-1} \left(\frac{v_u}{f_u} \right), \quad (3.5)$$

e α é dado por:

$$\alpha = -\tan^{-1} \left(\frac{v_v}{f_v} \cos \beta \right). \quad (3.6)$$

Uma questão importante é como obter o ponto de intersecção \mathbf{v} das retas que representam as bordas de pista de forma automática e de uma maneira robusta. Há várias maneiras para se detectar as bordas da pista, e uma escolha adequada para o propósito em questão é o modelo linear-parabólico usado em (JUNG; KELBER, 2005). No modelo linear-parabólico, as bordas de pista são modeladas como um função linear referente ao campo de visão próximo, de modo que \mathbf{v} possa ser calculado diretamente pela intersecção da parte linear das bordas da esquerda e da direita da pista. Além disso, a parte parabólica modela o campo de visão distante, sendo flexível a partes em curva da estrada. Por isso, como indicado em (JUNG; KELBER, 2005), o modelo mencionado é utilizado para identificar de forma automática partes da pista em linha reta ou curva. Ainda, pode-se selecionar automaticamente um conjunto de quadros de treinamento no procedimento de calibração. Para obter o ponto de intersecção \mathbf{v} foi utilizado apenas a parte linear do modelo.

3.3 Obtendo a altura da câmera: primeira abordagem

Para se obter a altura da câmera com base na geometria da estrada, um parâmetro adicional que relaciona as coordenadas de mundo e de imagem é necessário. Considerando dois pontos ao longo de uma das bordas da pista (sem perder a generalidade, considera-se a borda da esquerda), dado $\mathbf{W}_1 = (0, y_0, z_0)$ e $\mathbf{W}_2 = (0, y_0, z_0 + L)$, como mostrado na Figura 3.2. A projeção destes pontos 3D em coordenadas de imagem é dado por:

$$\mathbf{I}_1 = \begin{pmatrix} u_1 \\ v_1 \end{pmatrix} = \begin{pmatrix} \frac{f_u (-h \cos \beta - y_0 \sin \beta \sin \alpha - z_0 \sin \beta \cos \alpha)}{-h \sin \beta + y_0 \cos \beta \sin \alpha + z_0 \cos \beta \cos \alpha} \\ \frac{f_v (y_0 \cos \alpha - z_0 \sin \alpha)}{-h \sin \beta + y_0 \cos \beta \sin \alpha + z_0 \cos \beta \cos \alpha} \end{pmatrix}, \quad (3.7)$$

$$\mathbf{I}_2 = \begin{pmatrix} u_2 \\ v_2 \end{pmatrix} = \begin{pmatrix} -\frac{f_u (h \cos \beta + y_0 \sin \beta \sin \alpha + z_0 \sin \beta \cos \alpha + L \sin \beta \cos \alpha)}{-h \sin \beta + y_0 \cos \beta \sin \alpha + z_0 \cos \beta \cos \alpha + L \cos \beta \cos \alpha} \\ -\frac{f_v (-y_0 \cos \alpha + z_0 \sin \alpha + L \sin \alpha)}{-h \sin \beta + y_0 \cos \beta \sin \alpha + z_0 \cos \beta \cos \alpha + L \cos \beta \cos \alpha} \end{pmatrix}. \quad (3.8)$$

Para se determinar de forma explícita a altura h , primeiramente isola-se z_0 usando a primeira coordenada das Equações (3.7) e (3.8), tal que se tem duas diferentes expressões para z_0 . Realizando a comparação entre ambas expressões e isolando h , tem-se:

$$h = \frac{L \cos \alpha \left((u_1 + u_2)(f_u \cos \beta \sin \beta) + f_u^2 + u_1 u_2 \cos^2 \beta - f_u^2 \cos^2 \beta \right)}{f_u (u_2 - u_1)} \quad (3.9)$$

Outrossim, o cálculo referente aos coeficientes angulares m_l e m_r , das equações lineares relativas as bordas da esquerda e da direita da pista em coordenadas de imagem, podem ser obtidos por:

$$m_l = \frac{f_v (h \sin \alpha \sin \beta - y_0 \cos \beta)}{f_u h \cos \alpha}, \quad (3.10)$$

$$m_r = \frac{f_v (h \sin \alpha \sin \beta - y_0 \cos \beta - W \cos \beta)}{h f_u \cos \alpha} \quad (3.11)$$

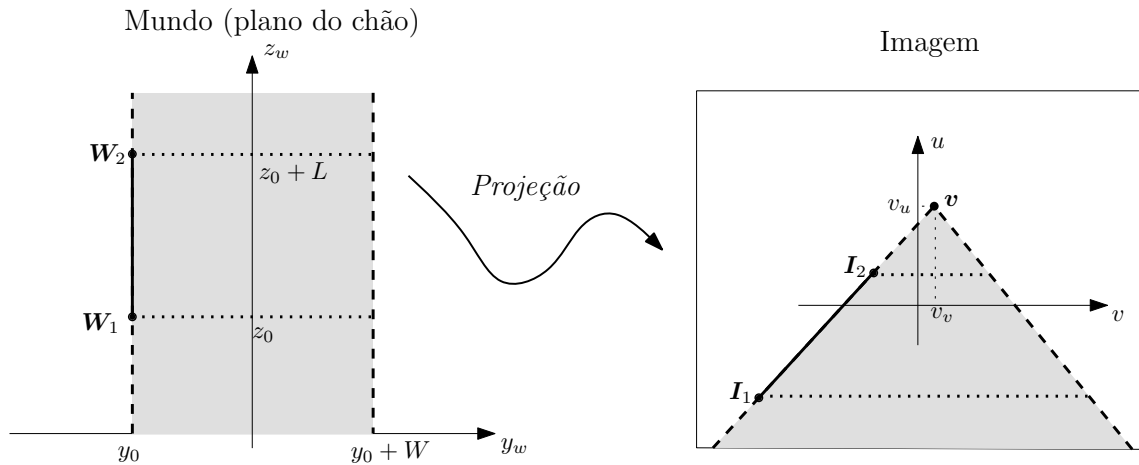
Isolando y_0 em ambas as Equações (3.10) e (3.11), igualando-as e isolando W na expressão resultante obtém-se:

$$W = h(m_l - m_r) \frac{f_u \cos \alpha}{f_v \cos \beta}, \quad (3.12)$$

que é uma estimativa da largura da pista que pode ser obtida de forma automática.

Na Equação (3.9), torna-se necessário encontrar um par de pontos na imagem $\mathbf{I}_1 = (u_1, v_1)$ e $\mathbf{I}_2 = (u_2, v_2)$ ao longo da borda da pista, tal que a distância dos correspondentes pontos em coordenadas de mundo seja L . Para encontrar estes pontos, assume-se que a borda

Figura 3.2 – Esquerda: estrada em coordenadas de mundo. Direita: estrada em coordenadas de imagem.



Fonte: do autor (2014).

da pista seja tracejada, e que I_1 , I_2 correspondem ao início das marcas tracejadas adjacentes. Por este motivo, L é o comprimento entre o início entre duas marcas longitudinais tracejadas adjacentes, que podem ser calculadas de forma automática se o veículo estiver se movendo em uma velocidade v conhecida e constante durante o período de calibração.

Para estimar L , primeiramente assume-se que a intensidade da luz refletida das marcas de sinalização da pista é maior que a intensidade do pavimento da pista (asfalto). Portanto, considera-se uma região retangular externa r_e centralizada no *pixel* localizado na parte inferior da borda de pista, com altura $0.025w_l$ e largura $0.15w_l$, onde w_l é a distância horizontal (em *pixels*) entre os *pixels* na parte inferior das bordas da pista. Também, utiliza-se outra região retangular r_i , com a mesma dimensão de r_e , mas deslocada para o interior da pista $0.3w_l$ unidades como apresentado na Figura 3.3. O retângulo interno r_i é utilizado para aprender os valores da média μ e do desvio padrão σ dos *pixels* referentes ao pavimento, os quais serão usados para avaliar a aderência de cada *pixel* no interior de r_e as estatísticas do pavimento. Mais precisamente, uma imagem binária $r'_e(u, v)$ é obtida através de:

$$r'_e(u, v) = \begin{cases} 1, & \text{if } r_e(u, v) > \mu + k\sigma \\ 0, & \text{caso contrário} \end{cases}, \quad (3.13)$$

onde $k = 5$ é a tolerância (número de desvios padrões) utilizados para distinguir os *pixels* de pavimento e de marcas longitudinais, obtidos de forma experimental. Para um dado quadro t , calcula-se a medida $m(t)$ baseada na fração de *pixels* no interior de r_e que são supostamente

relacionados com as marcações de pista:

$$m(t) = \frac{\sum r'_e(u, v)}{\#r'_e}, \quad (3.14)$$

onde $\#r'_e$ é o número total de *pixels* em r'_e . Enquanto o veículo se movimenta, os valores de $m(t)$ alternam entre partes de marcação (pintura) e não marcação (sem pintura) da estrada, como apresentado na Figura 3.5(a). Ademais, se o comprimento das marcações longitudinais tracejadas é o mesmo (pelo menos ao longo dos quadros utilizados na calibração da câmera), então $m(t)$ deve ser uma função periódica. Se o período (em segundos) é T , então o comprimento L entre o início das marcas adjacentes é $L = vT$. Neste trabalho, a velocidade do veículo foi medida de forma manual, mas considera-se importante mencionar que dispositivos eletrônicos embutidos nos modernos carros podem ser usados para se obter a velocidade de forma automática a qualquer momento (KOSCHER et al., 2010).

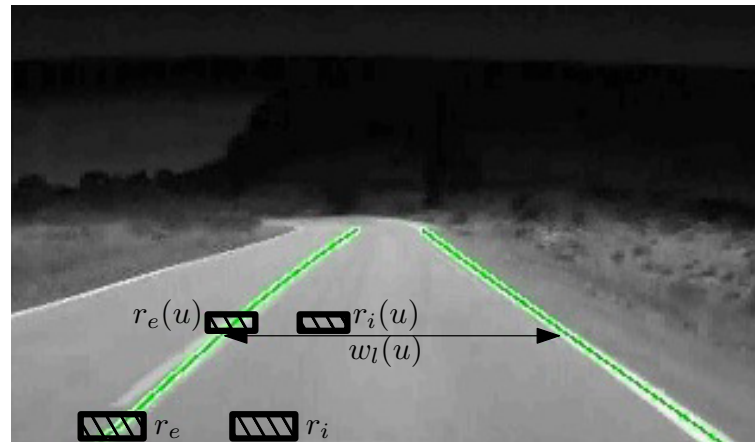
Na prática, entretanto, partes não homogêneas do pavimento da estrada e a baixa qualidade das marcações de pista (pintura desgastada, por exemplo) fazem com que a determinação do período T seja uma tarefa desafiadora. Portanto, decidiu-se por explorar a função de autocorrelação (GONZALEZ; WOODS, 2007) para se determinar T . Dada a função $m(t)$ no domínio discreto, a função de autocorrelação $R_{mm}(\tau)$ é expressa por:

$$R_{mm}(\tau) = \sum_t m(t)m(t + \tau), \quad (3.15)$$

e os valores de T_τ que produzem um máximo local para $R_{mm}(\tau)$ devem indicar os períodos da função. Para estimar o período fundamental de $m(t)$, seleciona-se T como o primeiro valor positivo de τ que gera o máximo local de $R_{mm}(\tau)$. A Figura 3.5(b) apresenta o gráfico de $R_{mm}(\tau)$. Nota-se que o primeiro máximo local, depois da origem, ocorre em $\tau = 16$. Uma vez que a sequência de vídeo é capturada a 30 qps (quadros por segundo), o período calculado é de $T = 16/30 = 0,53$ s. Considerando que a velocidade do veículo seja $55 \text{ km/h} = 15,28 \text{ m/s}$, o comprimento estimado entre o início das marcas longitudinais adjacentes é $L = 8,15 \text{ m}$.

Dado T (e conseqüentemente o comprimento entre as marcas L), o próximo passo consiste em selecionar um quadro da sequência de vídeo em que seja possível detectar de maneira acurada o início de duas marcas longitudinais tracejadas adjacentes. Tendo selecionado um quadro e detectado as bordas da pista, desloca-se as duas regiões retangulares r_e e r_i ao longo das bordas da pista (e de forma paralela) e estima-se a razão entre os *pixels* relacionados as marcas da pista no retângulo externo, similarmente ao procedimento adotado para se

Figura 3.3 – Quadro típico e as regiões retangulares usadas para estimar o comprimento entre as marcas longitudinais adjacentes ao longo das bordas de pista detectadas (em verde).

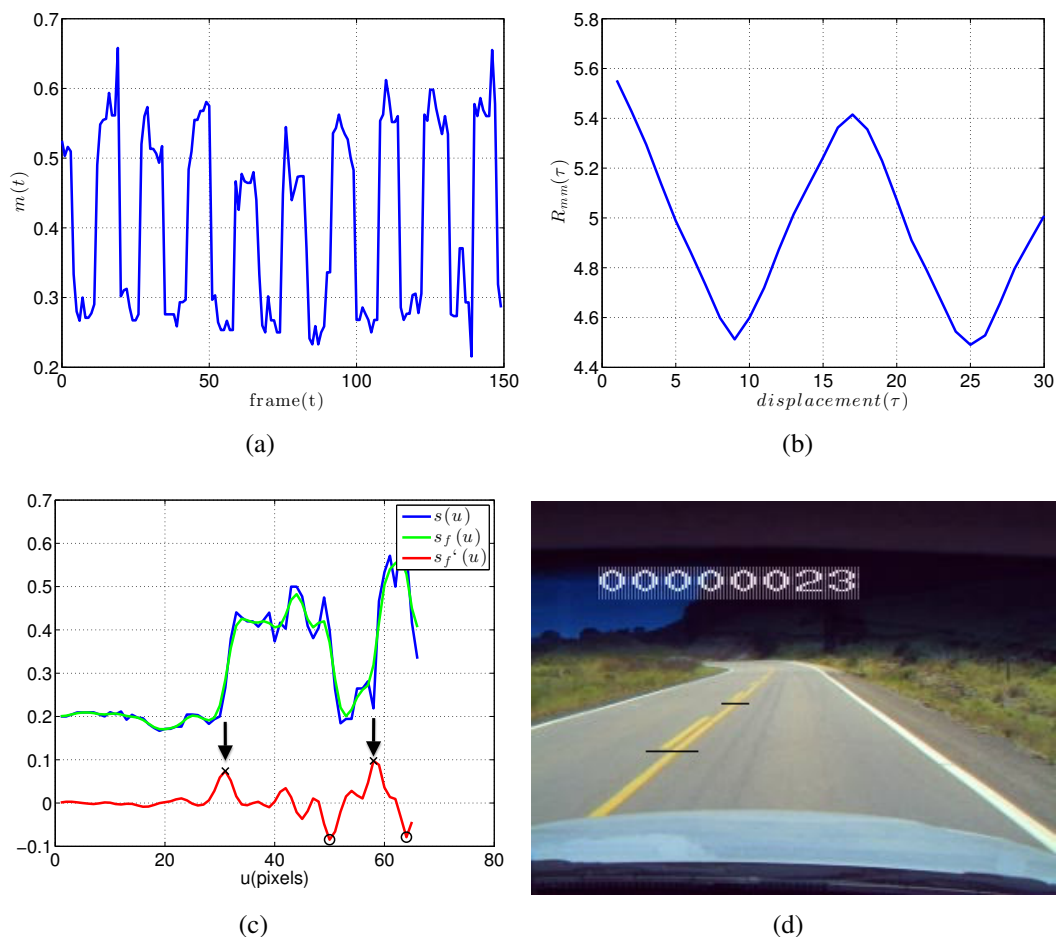


Fonte: do autor (2014).

obter $m(t)$. De forma mais detalhada, considera-se uma região retangular externa deslizante $r_e(u)$ que apresenta a sua borda inferior na coordenada vertical da imagem u com dimensão $0,025w_l(u) \times 0,15w_l(u)$, onde $w_l(u)$ é a distância horizontal (em *pixels*) entre as bordas da pista como uma função de u (o tamanho da janela deslizante é re-escalado para diferentes posições da imagem com o intuito de compensar questões relacionadas a perspectiva). Ainda, uma região interna retangular $r_i(u)$ com a mesma dimensão, porém deslocada $0,3w_l(u)$ unidades de *pixel* para o interior da pista, também é considerada. Por fim, calcula-se a razão dos *pixels* $s(u)$ com $r_e(u)$, de forma similar ao procedimento descrito pelas Equações (3.13) e (3.14).

Espera-se que no gráfico de $s(u)$ haja uma série de platôs: ao longo do pavimento, $s(u)$ deva ser aproximadamente uniforme e com valor pequeno, enquanto que ao longo das marcas, $s(u)$ deva ser aproximadamente uniforme, mas com valor alto. Transições significativas entre estes valores são esperadas no início e no fim das marcas tracejadas da pista. Portanto, os picos de $s'(u)$ podem ser usados para detectar os limites de cada marca de pista. Entretanto, observa-se na Figura 3.5(c), que o ruído em $s(u)$ afeta significativamente o valor de sua derivada. Para contornar este problema, utiliza-se um filtro de *kernel* Gaussiano, gerando uma versão filtrada, denominada $s_f(u)$. Para tanto, extrai-se todos os máximos locais de $s'_f(u)$, e considera-se apenas os picos que são maiores que 50% do maior pico obtido (para remover máximos locais pequenos). Similarmente, extrai-se todos os mínimos locais de $s'_f(u)$, e considera-se aqueles picos que são menores que 50% do menor pico (em valores absolutos). O máximo e mínimo locais de $s'_f(u)$ são mostrados na Figura 3.5(c) com os símbolos cruz e círculo, respectivamente. O primeiro máximo local é a primeira transição significativa em $s_f(u)$, e é atribuído ao início da

Figura 3.4 – (a) Gráfico da função $m(t)$. (b) Função de autocorrelação $R_{mm}(\tau)$. (c) Gráfico de $s(u)$; versão com filtro $s_f(u)$ e a correspondente derivada $s'_f(u)$ ao longo dos pontos de máximo e mínimo bem como a determinação do início das marcas longitudinais adjacentes. (d) Início das marcas adjacentes mapeadas no quadro da imagem.

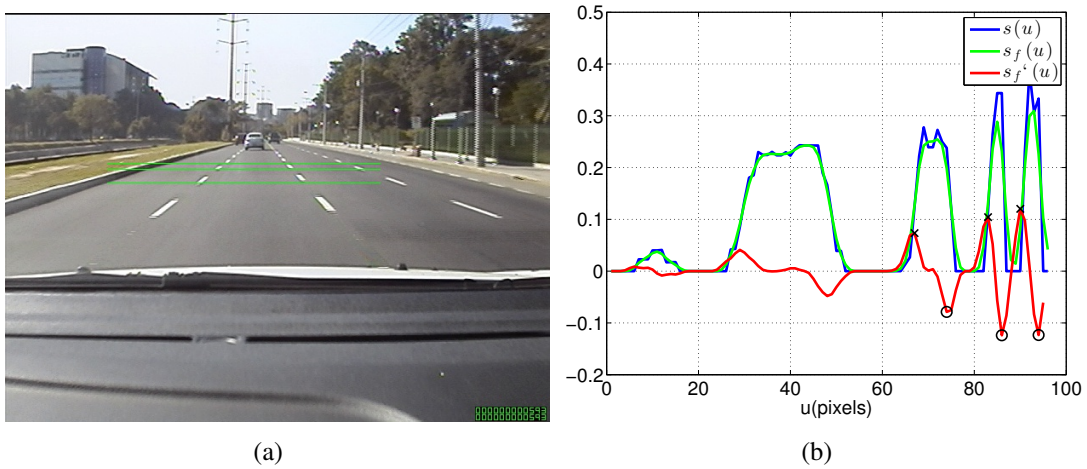


Fonte: do autor (2014).

primeira marca tracejada da pista que aparece ao longo da borda. De forma análoga, o primeiro mínimo local é atribuído ao final da referida marca. O primeiro máximo local subsequente² refere-se ao início da segunda marca tracejada da pista. Estas posições estão marcadas por setas na Figura 3.5(c), e os pontos correspondentes no quadro da imagem são apresentados na Figura 3.5(d). Na Figura 3.5 pode ser observado o início de três marcas longitudinais adjacentes.

²Se não há o segundo máximo local, o quadro selecionado é descartado no procedimento de calcular o valor de h .

Figura 3.5 – (a) Início das marcas adjacentes no quadro da imagem. (b) Respectiveos máximos locais em $s'_f(u)$ sinalizados por um \times .



Fonte: do autor (2014).

3.4 Obtendo a altura da câmera: segunda abordagem

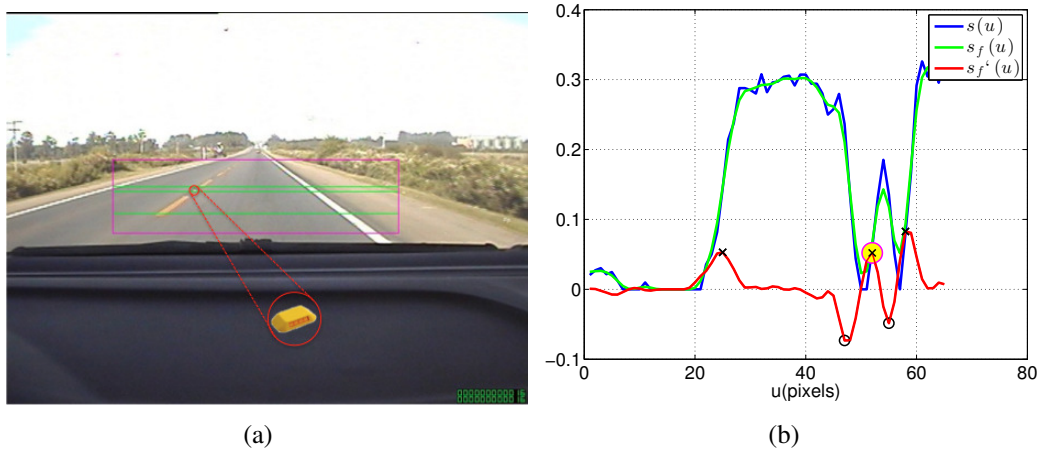
Um problema da abordagem descrita anteriormente em estimar o valor de h é ilustrado na Figura 3.7(a). A Figura mostra a sequência de vídeo 3 (maiores detalhes estão na Seção 3.7), onde há uma marca refletiva fixada entre as marcas tracejadas adjacentes. O segundo máximo local de $s'_f(u)$, apresentado na Figura 3.7(b) - círculo amarelo, não condiz com o início da segunda marca. Consequentemente, a detecção das marcas conduz a uma falha ao longo dos quadros usados na análise. A seguir, é apresentada uma abordagem alternativa a esta questão que é fundamentada no movimento de pontos planares, similar ao procedimento de calibração com base na estrutura a partir do movimento (HARTLEY; ZISSERMAN, 2004).

Considerando um ponto no espaço 3D $\mathbf{W}_3 = (0, y_3, z_3)^T$, localizado no plano do chão, e assumindo que a matriz de rotação \mathbf{R} seja conhecida, projetando-se \mathbf{W}_3 num ponto de imagem $\mathbf{I}_3 = (u_3, v_3)^T$ e resolvendo para y_3 e z_3 , tem-se:

$$\begin{pmatrix} u_3 \\ v_3 \end{pmatrix} = \begin{pmatrix} -\frac{f_u h v_3 \cos^2 \beta + f_u h v_3 \sin^2 \beta}{f_u f_v \sin \beta + f_v u_3 \cos \beta} \\ -\frac{h u_3 \sin \beta - f_u h \cos \beta}{u_3 \cos \beta + f_u \sin \beta} \end{pmatrix} \quad (3.16)$$

Assumindo que o veículo esteja se movimentando ao longo do eixo central da pista com uma velocidade v conhecida, a posição de \mathbf{W}_3 no quadro seguinte será dada por $\mathbf{W}_4 = (0, y_3, z_3 - d_z)^T$, onde $d_z = v/\text{qps}$ é o deslocamento ao longo do eixo z . Subtraindo d_z de z_3 na

Figura 3.6 – (a) Tachão refletivo no pavimento entre marcas longitudinais tracejadas. (b) Gráfico de $s'_f(u)$ que leva a estimativa errada de L .



Fonte: do autor (2014).

Equação (3.16) e calculando o mapeamento direto, obtém-se as correspondentes coordenadas de imagem $\mathbf{I}_4 = (u_4, v_4)^T$, expressa por:

$$\begin{pmatrix} u_4 \\ v_4 \end{pmatrix} = \begin{pmatrix} \frac{f_u(h \cos \beta - \Phi \sin \beta)}{h \sin \beta + \Phi \cos \beta} \\ \frac{f_v(f_u h v_3 \cos^2 \beta + f_u h v_3 \sin^2 \beta)}{(f_u f_v \sin \beta + f_v u_3 \cos \beta)(h \sin \beta + \Phi \cos \beta)} \end{pmatrix} \quad (3.17)$$

$$\text{onde } \Phi = d_z - \frac{h u_3 \sin \beta - f_u h \cos \beta}{u_3 \cos \beta + f_u \sin \beta}.$$

Por este motivo, se um par de pontos correspondentes $\mathbf{I}_3, \mathbf{I}_4$, em coordenadas de imagem, forem conhecidos, cada coordenada da Equação (3.17) pode ser usada para se estimar a altura da câmera. De fato, isolando h em ambas as coordenadas, chega-se a:

$$h_u = \frac{d_z f_u^2 \sin^2 \beta + d_z u_3 u_4 \cos^2 \beta + d_z f_u u_3 \cos \beta \sin \beta + d_z f_u u_4 \cos \beta \sin \beta}{f_u u_3 \cos^2 \beta - f_u u_4 \cos^2 \beta + f_u u_3 \sin^2 \beta - f_u u_4 \sin^2 \beta} \quad (3.18)$$

$$h_v = \frac{d_z u_3 v_4 \cos^2 \beta + d_z f_u v_4 \sin \beta \cos \beta}{f_u v_3 \cos^2 \beta - f_u v_4 \cos^2 \beta + f_u v_3 \sin^2 \beta - f_u v_4 \sin^2 \beta} \quad (3.19)$$

Se um conjunto de pontos correspondentes n , entre os quadros t e $t + 1$, estiver disponível, tem-se n equações redundantes para se determinar a mesma estimativa da altura h_u e h_v no quadro t . Se $h_u^{(j)}(t)$ e $h_v^{(j)}(t)$ denotam as estimativas de forma independente, para $j = 1, \dots, n$, propõe-se determinar um único valor para $h_u(t)$ e $h_v(t)$ no quadro t através de $h_u(t) = \mu_{\kappa_j}(h_u^{(j)}(t))$ e $h_v(t) = \mu_{\kappa_j}(h_v^{(j)}(t))$. A expressão $\mu_{\kappa_j}(\cdot)$ representa a média truncada κ (κ -trimmed mean) na sequência de valores da variável j (BEDNAR; WATT, 1984), que con-

siste na remoção de uma fração ($0 \leq \kappa < 1$) das observações que estejam relativamente longe do valor mediano (possivelmente relacionado a valores discrepantes) e, ainda, calcular a média dos valores restantes. Nos experimentos realizados verificou-se que um valor $\kappa = 0,2$ é o suficiente para remover os valores discrepantes e permitir valores suficientes para o cálculo da média.

Há várias técnicas que podem ser usadas para se encontrar pontos correspondentes em duas imagens, necessários para se estimar h_u e h_v . Uma vez que se tenha uma coerência espaço-temporal destes pontos, um rastreador de características (*feature tracker*) é selecionado. Mais especificamente, utilizou-se o algoritmo Lucas-Kanade (LK) piramidal (BOUGUET, 2000; LUCAS; KANADE, 1981) para realizar o rastreamento de pontos localizados no plano do chão. Com a intenção de selecionar boas características para realizar o rastreamento (*good features to track*), aplicou-se o clássico detector de Harris (HARRIS; STEPHENS, 1988) para encontrar pontos com alta curvatura em curvas planas, o que é esperado nos cantos das marcas longitudinais tracejadas e no pavimento, caso haja uma textura suficiente. Tais pontos são selecionados em uma região da imagem compreendida entre as bordas de pista detectadas. Portanto, estes pontos no plano do chão são susceptíveis de serem selecionados (mesmo que alguns pontos sejam selecionados fora do chão, eles tendem a ser removidos pela média truncada κ).

3.5 Melhorando a estimativa com informação temporal

As abordagens descritas nas Seções 3.3 e 3.4 podem ser utilizadas para estimar de forma independente α , β e h em cada quadro, sendo que uma janela temporal é necessária para estimar h . De fato, cada quadro t gera três estimativas para a altura: $h_s(t)$, usando as restrições espaciais descritas na Seção 3.3, e $h_u(t)$, $h_v(t)$, obtidas usando o movimento conforme descrito na Subseção 3.4. Entretanto, os parâmetros relacionados à imagem variam significativamente quando quadros diferentes são selecionados, devido à ruídos na imagem, sombra, condições da pintura da pista, etc.. Para se obter uma estimativa mais robusta aos desejados parâmetros extrínsecos, considera-se que $\{\alpha(t), \beta(t), h_s(t), h_u(t), h_v(t)\}$ seja um conjunto de parâmetros estimados de um dado quadro (t), para $t = 1, \dots, N_T$, onde N_T é o número de quadros utilizados na análise. Os parâmetros finais estimados são obtidos a partir de:

$$\hat{\alpha} = \mu_{\kappa}(\alpha(t)), \quad \hat{\beta} = \mu_{\kappa}(\beta(t)), \quad \hat{h}_p = \mu_{\kappa}(h_p(t)), \quad (3.20)$$

onde $\mu_{\kappa}(\cdot)$ é a média truncada κ da variável t , e $p \in \{s, u, v\}$ denota as diferentes possibilidades de se estimar a altura. Conforme as subseções anteriores, utilizou-se $\kappa = 0, 2$.

3.6 Análise de erros

Os parâmetros desejados α , β e h são calculados com base nos parâmetros estimados da imagem v_u , v_v , L , u_1 e u_2 , que são suscetíveis a erros. Considere que v_u^* , v_v^* , L^* , u_1^* e u_2^* são valores exatos, e que $v_u = v_u^* + \Delta v_u$, $v_v = v_v^* + \Delta v_v$, $L = L^* + \Delta L$, $u_1 = u_1^* + \Delta u_1$ e $u_2 = u_2^* + \Delta u_2$, são os valores estimados. Assumindo que os parâmetros estimados contenham pequenos erros, a aproximação de primeira ordem dos erros aplicada às Equações (3.5) e (3.6) usando diferenciais, leva às seguintes aproximações para os erros angulares $\Delta\beta$ e $\Delta\alpha$:

$$\Delta\beta \approx -\frac{f_u}{f_u^2 + v_u^2} \Delta v_u, \quad (3.21)$$

$$\Delta\alpha \approx -\frac{f_v (\Delta v_v \cos \beta - v_v \Delta\beta \sin \beta)}{f_v^2 + v_v^2 \cos^2 \beta}. \quad (3.22)$$

O erro no ângulo de arfagem (*pitch*) $\Delta\beta$ tende a ser pequeno, uma vez que o erro em Δv_u é geralmente pequeno (em torno de 5 *pixels* no máximo de acordo com os experimentos realizados), e tipicamente $f_u \gg \Delta v_u$. O erro no ângulo de guinada (*yaw*) $\Delta\alpha$ é afetado por $\Delta\beta$ e também por Δv_v . Uma vez que o termo $\Delta\beta \sin \beta$ tende a ser pequeno e $f_v \gg \Delta v_v$ em termos gerais, também é esperado que o erro $\Delta\alpha$ seja pequeno.

Para estimar a altura da câmera h , a primeira abordagem descrita na Seção 3.3 depende dos parâmetros u_1 , u_2 e L (veja a Equação (3.9)), que são extraídas da sequência de vídeo, e dos parâmetros estimados α e β (que também possuem erros). Para simplificar a análise de erros, foi assumido que o parâmetro α (*yaw*) é irrelevante, bem como o erro $\Delta\beta$ do parâmetro β (*pitch*). Também foi assumido que $\Delta u_1 = \Delta u_2 = \Delta u$, i.e., o erro obtido para obter o início das marcas adjacentes é o mesmo. Com essas premissas, obtém-se:

$$\begin{aligned} \Delta h \approx & \frac{f_u (u_1 + u_2) \sin \beta \cos \beta + f_u^2 \sin^2 \beta + u_1 u_2 \cos^2 \beta}{f_u (u_2 - u_1)} \Delta L \\ & + \frac{L \cos \beta (2f_u \sin \beta + (u_1 + u_2) \cos \beta)}{f_u (u_2 - u_1)} \Delta u. \end{aligned} \quad (3.23)$$

Uma análise teórica da Equação (3.23) não é uma tarefa fácil, mas em geral, algumas observações do erro esperado podem ser feitas.

No primeiro termo, $f_u^2 \sin^2 \beta$ pode ser um termo dominante, dependendo do ângulo β (*pitch*) da câmera. O segundo termo depende linearmente do espaçamento L entre as marcas adjacentes, tal que trechos da pista com pequenos espaçamentos tendem a produzir melhores resultados.

Para as estimativas de altura que usam a segunda abordagem, dada pelas Equações (3.18) e (3.19), o erro depende das características (*features*) do rastreador que estão relacionadas com os pontos I_3 e I_4 na imagem. Como consequência, uma análise teórica de erros é difícil de ser elaborada. Entretanto, o rastreador tende a trabalhar melhor quando pontos com alta curvatura são detectados, que ocorrem quando os vértices (cantos) das marcações são pontiagudas e/ou quando o pavimento é altamente texturado.

3.7 Resultados

Os sistema proposto foi implementado em C++ (compilador GCC 4.2.1), utilizando a biblioteca **Open** source **Computer Vision**, (OpenCV Versão 2.4.0³). O tempo médio de execução para cada quadro 240×320 e 480×640 , utilizando um *notebook* com processador Core 2 Duo de 2,53GHz, com 4GB de RAM, e uma placa gráfica NVIDIA GPU GeForce 9400M, foi de 22,64 ms e 64,80 ms, respectivamente. É importante observar que o conjunto de quadros deve ser analisado duas vezes (num primeiro momento, estima-se L ; na segunda vez, obtém-se as coordenadas de imagem das marcas longitudinais tracejadas e a estimação dos parâmetros extrínsecos) e ambos os processos são realizados totalmente na CPU, enquanto que o vídeo é executado na placa gráfica.

A abordagem proposta foi aplicada em sete (7) sequências de vídeo adquiridas com duas câmeras diferentes. Os parâmetros intrínsecos de cada câmera foram obtidos de forma *offline* com as funções de calibração do OpenCV (BRADSKI, 2000)⁴ antes da fixação ao para-brisa (uma vez que os parâmetros intrínsecos permanecem constantes e, conseqüentemente, este procedimento deve ser realizado uma única vez para cada câmera). Salienta-se que as coordenadas de imagem (u, v) , definidas na Seção 3.1 são relativas ao eixo ótico da câmera (u_o, v_o) , também obtidos no procedimento *offline* de calibração. A validação foi realizada de forma quantitativa pela comparação entre os parâmetros estimados com os dados de *ground truth*.

³<<http://opencv.willowgarage.com/wiki/>>

⁴Também, realizou-se a calibração usando o *Toolbox* de calibração de câmeras da Caltech para Matlab (BOUGUET, 2008), e os resultados obtidos foram similares.

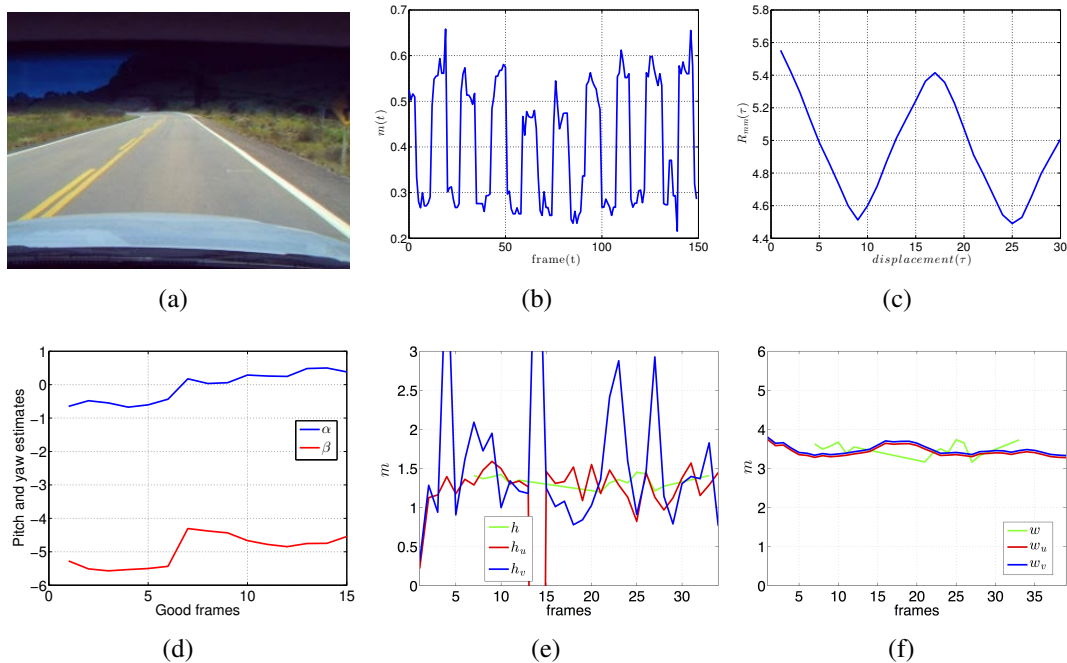
O primeiro vídeo (*Cambará*), foi adquirido usando uma câmera Jenoptik JD C 3.1 z3, com uma resolução de 240×320 pixels, a 30 qps, e os parâmetros intrínsecos estimados foram $f_u = 337,32$, $f_v = 345,88$, $u_o = 165,47$ e $v_o = 117,45$. O veículo estava a uma velocidade constante de 55 km/h e 50 quadros foram utilizados para calibrar a câmera (apenas 15 dos 50 quadros geraram parâmetros válidos para estimar a altura da câmera h). A Figura 3.7 ilustra alguns aspectos da abordagem proposta aplicada a esta sequência de vídeo. A Figura 3.8(a) apresenta um quadro da sequência de vídeo; a Figura 3.8(b) mostra o gráfico de $m(t)$ e a Figura 3.8(c) ilustra o gráfico da função de autocorrelação $R_{mm}(\tau)$. Como pode ser observado, embora $m(t)$ possua um certo ruído, o gráfico de $R_{mm}(\tau)$ é bastante suave, tal que T pode ser obtido de forma robusta. Os gráficos estimados do *pitch* e *yaw* ao longo dos quadros de calibração são mostrados na Figura 3.8(d), e as três estimativas para a altura h , ao longo do tempo, estão apresentadas na Figura 3.8(e). As estimativas do *pitch* e *yaw* são consistentes no tempo (em torno de 2°), enquanto a estimativa de h apresenta vários valores discrepantes, os quais são removidos de forma temporal com a média truncada κ . Por uma questão de ilustração, a estimativa da largura da pista W , relacionada a cada estimativa de h , é apresentada na Figura 3.8(f).

Os vídeos 2-6 (*Castelo Branco, BR116, BR290, Ipiranga, Assis Brasil*) foram capturados utilizando uma câmera analógica Watec Wat 240R, com uma lente fixa de 3,8 mm, e digitalizados com uma resolução de 480×640 pixels, a 30 qps, usando um dispositivo de captura Pinnacle PCTV USB2, tendo como parâmetros intrínsecos $f_u = 674,30$, $f_v = 674,30$, $u_o = 319,50$ e $v_o = 239,50$. As velocidades do veículo para as referidas sequências foram de 70, 40, 80, 50 e 50 km/h, respectivamente.

Finalmente, para testar a performance da abordagem proposta foi utilizada a sequência de vídeo número 7 (*Germany*) em que a velocidade do veículo foi de 120 km/h. Foram usados 100 quadros desta sequência para se estimar os parâmetros (apenas 28 quadros foram considerados para se estimar h). A câmera utilizada foi a mesma da sequência de vídeo 1.

A Tabela 3.1 mostra o resumo dos parâmetros estimados α , β e h para todas as sequências de vídeo. Para algumas sequências, tem-se o *ground truth* para h , medido diretamente (distância da lente ao chão), e β , medido utilizando-se um inclinômetro analógico. O *ground truth* para α deveria ser zero, mas podem haver erros devido a instalação da câmera. Também, desde que foi proposto três maneiras de se estimar a altura (\hat{h}_s , \hat{h}_u e \hat{h}_v), apresenta-se uma estimativa combinada da altura (\hat{h}) calculando simplesmente a média aritmética das três estimativas. Como pode ser observado, o valor estimado para β possui um erro em torno de

Figura 3.7 – (a) Quadro da sequência de vídeo *Cambará*. (b) Gráfico de $m(t)$. (c) Gráfico de $R_{mm}(\tau)$. (d) Gráfico da estimativa do *pitch* e *yaw* em alguns quadros. (e) Gráfico da estimativa da altura em alguns quadros. (f) Gráfico da estimativa da largura da pista em alguns quadros com base nas diferentes estimativas da altura.



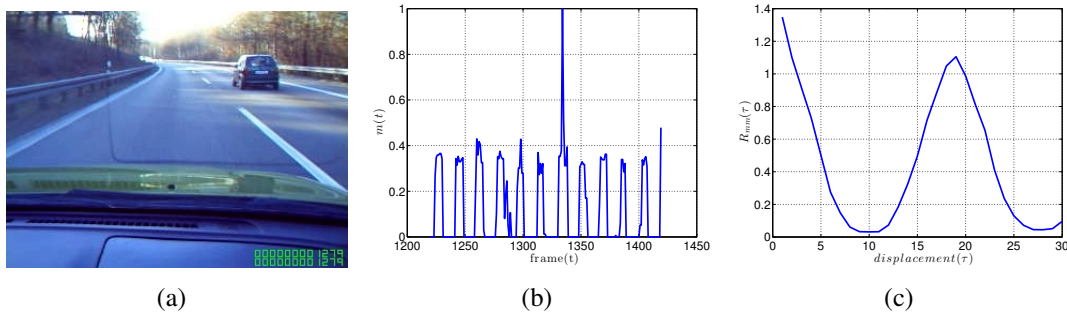
Fonte: do autor (2014).

$0,5^\circ$, enquanto que para α é de 2° . As estimativas para a altura h apresentam maiores variações (o erro relativo médio para \hat{h}_e é de 10,19%, o que corresponde a aproximadamente 10,56 centímetros em valores absolutos).

É importante ressaltar que a sequência de vídeo 7 apresenta fortes marcas de sombra na estrada, como pode ser observado na Figura 3.9(a). Embora a estimativa de $m(t)$ seja bastante ruidosa, como apresentada na Figura 3.9(b), a função de autocorrelação $R_{mm}(t)$ é suave o suficiente para permitir uma boa estimativa para a determinação de L . Entretanto, a abordagem de se detectar as marcas longitudinais adjacentes tende a falhar quando a estrada é coberta com uma sombra dispersa, uma vez que a região retangular usada para aprender as estatísticas do pavimento pode conter regiões de sombra e não sombra, o que gera estimativas erradas para a média e o desvio padrão do pavimento.

⁵Ground Truth

Figura 3.8 – (a) Quadro da sequência de vídeo *Germany*. (b) Gráfico da função $m(t)$. (c) Gráfico da função $R_{mm}(\tau)$.



Fonte: do autor (2014).

Tabela 3.1 – Resultados da calibração

Vídeo	Nome	α	β	β (GT ⁵)	\hat{h}_s	\hat{h}_u	\hat{h}_v	\hat{h}_e	h (GT)
1	Cambará	0,15474°	-4,85565°	–	1,40856m	1,27154m	1,11126m	1,26378m	–
2	Castelo Branco	1,63193°	-17,3463°	-17,0°	0,95158m	1,0709m	1,34371m	1,12206m	1,20m
3	BR116	-0,73957°	-6,02636°	-6,0°	1,56614m	1,43159m	1,27068m	1,4228m	1,20m
4	BR290	1,51358°	-7,89357°	-8,0°	1,39344m	1,07218m	1,47298m	1,31287m	1,20m
5	Assis Brasil	2,11023°	+1,56602°	+2,0°	1,17516m	1,41459m	1,20358m	1,26445m	1,20m
6	Ipiranga	-1,52376°	-7,60324°	-8,0°	1,38141m	1,54253m	1,29388m	1,40594m	1,20m
7	Germany	0,68985°	-11,4328°	–	1,08752m	0,90583m	0,91467m	0,82681m	–

Fonte: do autor (2014).

3.8 Conclusões

Neste Capítulo, foi apresentada uma nova abordagem dinâmica para se obter os parâmetros extrínsecos de uma câmera veicular embarcada de uma maneira automática. As suposições consideradas foram que os parâmetros intrínsecos sejam conhecidos, e que o veículo esteja se movendo em linha reta ao longo de um trecho planar da estrada com marcas longitudinais tracejadas em uma velocidade conhecida e constante durante o período de calibração. Salienta-se que nenhum parâmetro relacionado a estrada (como por exemplo, a largura da pista) é necessário na técnica proposta. Os valores de *pitch* e *yaw* da câmera dependem do ponto de fuga relacionado às bordas da pista, e os erros obtidos nos experimentos foram menores do que 0,5°. A altura da câmera foi estimada usando duas diferentes abordagens: a primeira extrai a distância (em coordenadas de mundo) de pontos do início das marcas longitudinais adjacentes, e enquanto que a segunda abordagem é fundamentada no rastreamento de pontos (*tracking*). A estimativa final é uma combinação destas duas abordagens, e o erro médio encontrado nos experimentos foi em torno de 10 cm (erro relativo aproximado de 10%). O procedimento para se obter a relação mundo-*pixels* automaticamente é fundamentado na estimativa das estatísticas do pavimento e na

detecção das discrepâncias ao longo da região onde as marcas da pista são esperadas. O método mencionado é robusto com respeito as alterações da iluminação global e sombras homogêneas, mas apresenta limitações quando sombras dispersas estão presentes (conforme discutido nas Seções 3.1 e 3.7). Também, salienta-se que o algoritmo usado neste trabalho (JUNG; KELBER, 2005) tende a falhar quando o limpador de para-brisa está ligado, por exemplo. Portanto, o procedimento de calibração provavelmente irá falhar em dias de chuva, baixa iluminação, noite ou forte neblina, por exemplo. Ainda, no Apêndice A são apresentadas possíveis aplicações baseadas em realidade aumentada para sistemas de assistência ao motorista. Em particular, o foco foi na sinalização horizontal e vertical, as quais poderiam ter sido utilizadas para advertir o condutor do veículo sobre a existência de uma curva, do limite máximo de velocidade em que o veículo pode trafegar na pista, dentre outras informações úteis. Nas sequências de vídeo com realidade aumentada, os objetos inseridos são visualmente coerentes com a geometria da estrada, indicando que os parâmetros extrínsecos são obtidos com a acurácia suficiente para aplicações de AR.

4 DETECÇÃO DO TIPO DE MARCAÇÃO HORIZONTAL

Para a detecção do tipo de sinalização horizontal foram desenvolvidas duas técnicas: (i) câmera não calibrada, (ii) câmera calibrada. Para ambas, a primeira etapa consiste em detectar e rastrear as bordas da pista de acordo com o trabalho proposto por Jung e Kelber (2005).

A primeira técnica foi desenvolvida em paralelo com a abordagem apresentada no Capítulo 3, visto que não requer uma câmera calibrada para realizar a identificação e classificação das marcas horizontais. Uma vez detectadas as bordas da pista, extraem-se algumas propriedades estatísticas do pavimento com base em uma região retangular localizada entre as bordas detectadas (como descrito em 3.3). Ainda, estes parâmetros servirão para detectar os *pixels* relacionados as marcações da pista em uma região de interesse (ROI - *Region Of Interest*) centrada na parte inferior da borda da pista. Enquanto o veículo se movimenta, a evolução temporal das marcações extraídas da pista fornece indicações sobre o tipo de marcação de pista, as quais são reconhecidas utilizando-se um classificador em cascata¹.

Para a segunda técnica, que explora uma câmera monocular calibrada, um esquema de suavização adaptativo é usado com o propósito de reduzir ruídos de segmentos unidimensionais extraídos do pavimento, preservando as bordas das marcações de pista. Máximos e mínimos locais, oriundos do gradiente da intensidade dos *pixels*, são utilizados como entrada para um classificador Bayesiano baseado em Misturas de Gaussianas.

4.1 Análise da sinalização horizontal usando câmera não calibrada

Assim como apresentado na Seção 3.3, para realizar a detecção do tipo de pintura torna-se necessário avaliar as marcações ao longo do tempo. Portanto, para cada *pixel* (u, v) da região r , verifica-se a consistência da distribuição dos *pixels* relacionados ao pavimento. Mais precisamente, uma região/imagem, também binária, $r_b^t(u, v)$ em cada quadro t é definida como:

$$r_b^t(u, v) = \begin{cases} 1, & \text{se } r(u, v) > \mu + k\sigma \\ 0, & \text{caso contrário} \end{cases}, \quad (4.1)$$

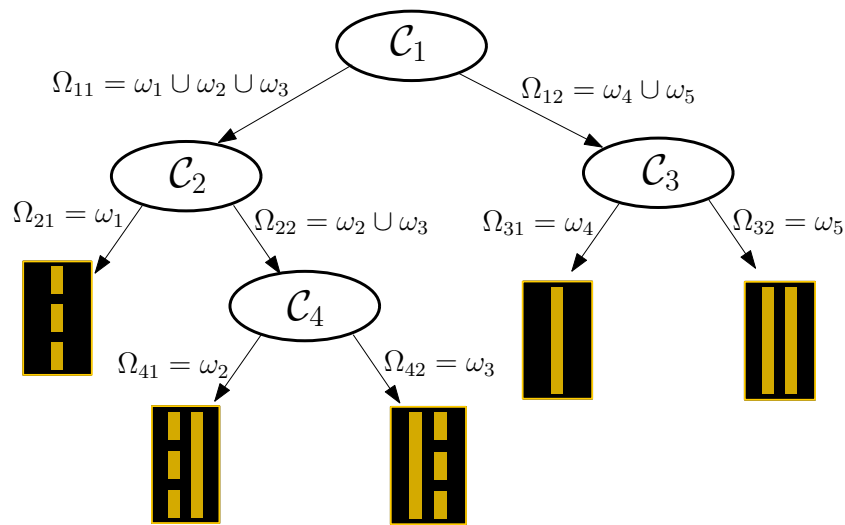
onde $k = 5$ é a tolerância usada para distinguir os *pixels* do pavimento e as marcações de pista.

As regiões $r_b^t(u, v)$ possuem dados binários relacionados a existência de marcações de

¹A cascata de classificadores, diferentemente do *Cascade Classifier com atributos HAAR*, é baseada na concatenação de classificadores que usam a saída de um classificador como entrada de outro. Maiores detalhes serão apresentadas nas Seções 4.1 e 4.2.

pista na ROI para cada quadro t . Um conjunto de atributos é extraído de $r_b^t(u, v)$, com o propósito de discriminar todas as cinco classes num modelo hierárquico. De fato, um classificador em cascata de três níveis foi desenvolvido, conforme ilustrado na Figura 4.1, e a seguir descreve-se as escolhas realizadas para os quatro classificadores binários \mathcal{C}_1 , \mathcal{C}_2 , \mathcal{C}_3 e \mathcal{C}_4 , bem como os vetores de atributos usados para cada classificador. Também, denota-se Ω_{i1} e Ω_{i2} as duas possíveis classes de saída do classificador \mathcal{C}_i , para $i = 1, 2, 3, 4$.

Figura 4.1 – Diagrama esquemático do classificador em cascata.

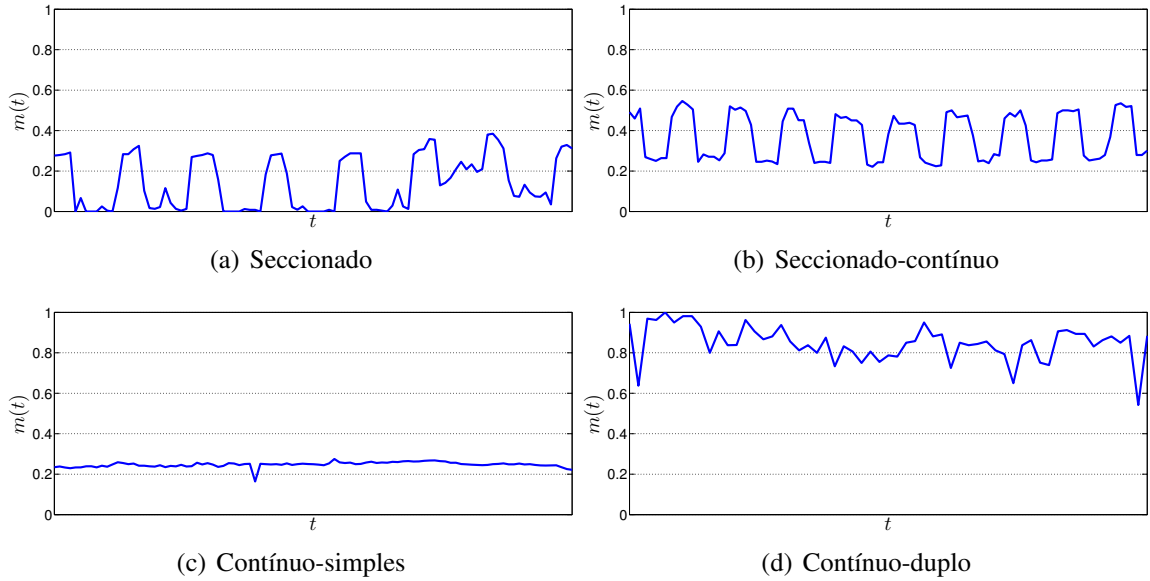


Fonte: do autor (2013).

4.1.1 Classificador \mathcal{C}_1

O primeiro classificador possui como objetivo separar as marcações de pista em componentes seccionados (Ω_{11}) e contínuos (Ω_{12}). No entanto, há poucos parâmetros que podem ser usados para discriminar estas duas classes: em partes com apenas componentes contínuos, $r_b^t(u, v)$ tende a apresentar o mesmo número de *pixels* relacionados as marcas de pista. Em contrapartida, o número destes *pixels* em partes da pista que possuem componentes seccionados tende a apresentar maiores variações (e de uma forma periódica). Neste trabalho, o classificador \mathcal{C}_1 recebe como entrada um vetor de características tridimensional $\mathbf{f}_1(t) = (f_{11}(t), f_{12}(t), f_{13}(t))^T$, contendo elementos que quantificam as variações supracitadas

Figura 4.2 – Gráficos da função $m(t)$ para diferentes tipos de marcações de pista.



Fonte: do autor (2013).

que serão definidas a seguir. Assim como apresentado na Equação (3.14), a fração de *pixels* com marcações relacionadas (com a região retangular) a cada quadro t é dada por

$$m(t) = \frac{\sum_{u,v} r_b^t(u, v)}{\#r_b^t}, \quad (4.2)$$

onde $\#r_b^t$ é o número total de *pixels* em r_b^t .

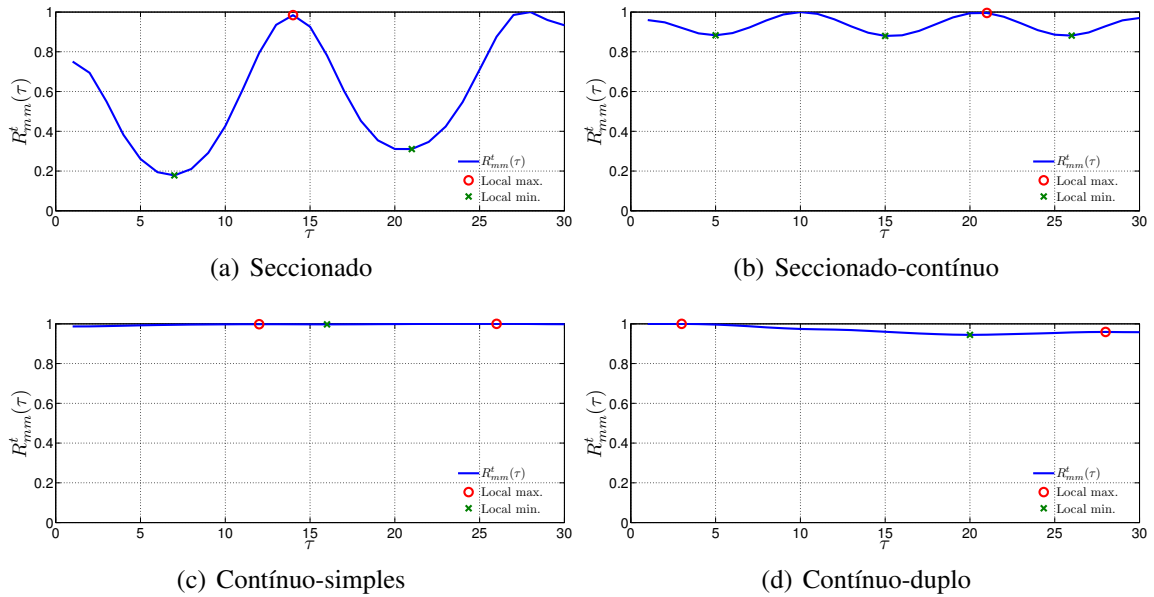
A função $m(t)$ deve oscilar quando os componentes seccionados estão presentes e manter-se constante na presença de componentes contínuos, como pode ser observado na Figura 4.2². Com base nesta premissa, para cada quadro t , avalia-se uma janela temporal $\mathcal{T}(t) = \{t - T + 1, t - T + 2, \dots, t - 1, t\}$ e computa-se a variação da amplitude de forma ponderada de $m(t)$ com a janela:

$$f_{11}(t) = \frac{1}{\mu_{\mathcal{T}(t)}\{m\}} \left(\max_{\mathcal{T}(t)}\{m\} - \min_{\mathcal{T}(t)}\{m\} \right), \quad (4.3)$$

onde $\mu_{\mathcal{T}(t)}$, $\max_{\mathcal{T}(t)}$ e $\min_{\mathcal{T}(t)}$ representam os valores da média, máximo e mínimo em $\mathcal{T}(t)$, respectivamente.

²Na Figura 4.2, o gráfico do tipo de marcação contínuo-seccionada é omitido, uma vez que é bastante similar ao caso seccionada-contínuo.

Figura 4.3 – Gráfico da autocorrelação da função $R_{mm}^t(\tau)$ para diferentes tipos de marcações (normalizada pelo maior valor da autocorrelação).



Fonte: do autor (2013).

Para explorar as características relacionadas a uma possível periodicidade de $m(t)$, calcula-se a autocorrelação de m com $\mathcal{T}(t)$:

$$R_{mm}^t(\tau) = \frac{1}{T'} \sum_{u=0}^{T'-1} m(t-u)m(t-u-\tau), \quad (4.4)$$

onde $T' < T$ é uma janela sobreposta entre $m(t-u)$ e $m(t-u-\tau)$, onde $\tau = 0, 1, \dots, T-T'-1$ é a variação de deslocamentos do sinal que podem ser computados com o sinal de tamanho fixo. Se $m(t)$ é periódica, $R_{mm}^t(\tau)$ deve produzir um pico quando τ se iguala ao valor do período. Por outro lado, funções constantes geram funções de autocorrelação constantes. As funções de autocorrelação correspondentes aos gráficos de $m(t)$, apresentados na Figura 4.2, estão indicadas na Figura 4.3.

Uma vez que as estimativas de $m(t)$ são ruidosas, valores maiores de T' , para calcular a função de autocorrelação, são necessários para melhor suprimir os ruídos. Entretanto, se T' (e consequentemente T) forem muito grandes, os valores de $m(t)$ usados na análise podem corresponder a partes da estrada com diferentes tipos de marcações de pista, ou seja, o sinal de $m(t)$ talvez não seja estacionário. Para os experimentos realizados, utilizou-se $T = 100$ e $T' = 70$, ambos definidos de forma experimental.

Dado um quadro t , calcula-se $R_{mm}^t(\tau)$, e recupera-se o primeiro máximo local $R_{mm}^t(\tau_{max})$ ³ e o primeiro mínimo local $R_{mm}^t(\tau_{min})$, calculando-se a diferença normalizada

$$f_{12}(t) = \frac{1}{R_{mm}^t(0)} \left(R_{mm}^t(\tau_{max}) - R_{mm}^t(\tau_{min}) \right) \quad (4.5)$$

como segundo atributo. Para marcações de pista com apenas componentes contínuos, espera-se que $f_{12}(t)$ possua valores pequenos, uma vez que a função de autocorrelação deva ser (idealmente) constante. No entanto, supõe-se que $f_{12}(t)$ tenha valores maiores quando componentes seccionados estejam presentes, devido a natureza periódica de $m(t)$.

Também, se a distância entre as marcações seccionadas de pista forem constantes num determinado trecho da estrada, a posição do máximo local deve ocorrer aproximadamente no mesmo lugar, o que corresponde ao período de $m(t)$. Outrossim, quando apenas marcas contínuas estão presentes, espera-se que os valores de $R_{mm}^t(\tau)$ sejam constantes e o primeiro máximo local ocorra em diferentes posições devido a pequenas variações em $R_{mm}^t(\tau)$. Por este motivo, o terceiro atributo usado no classificador \mathcal{C}_1 explora a variação da posição do máximo local com $\mathcal{T}(t)$. Embora o desvio padrão forneça esta métrica, decidiu-se usar:

$$f_{13}(t) = \text{MAD}_{u \in \mathcal{T}(t)} \{ \tau_{max}(u) \}, \quad (4.6)$$

onde $\tau_{max}(u)$ é a posição do primeiro máximo local de $R_{mm}^u(\tau)$, e o MAD (*Median Absolute Deviation*) é o desvio absoluto da mediana, a qual é menos afetada por valores discrepantes (*outliers*) do que o desvio padrão.

Quanto ao classificador \mathcal{C}_1 , optou-se por utilizar Máquinas de Vetores de Suporte (SVM - *Support Vector Machine*), e os resultados experimentais indicam que um *kernel* de Função de Base Radial (RBF - *Radial Basis Function*) apresentou os melhores resultados. Maiores detalhes sobre os dados de treinamento e de teste estão presentes na Seção 4.1.5.

4.1.2 Classificador \mathcal{C}_2

Como mostrado na Figura 4.1, o classificador \mathcal{C}_2 recebe amostras com componentes seccionados de acordo com \mathcal{C}_1 (classes ω_1 , ω_2 e ω_3), e classifica os mesmos como seccionado ($\Omega_{21} = \omega_1$) ou contínuo-seccionado/seccionado-contínuo ($\Omega_{22} = \omega_2 \cup \omega_3$).

Para este propósito, primeiramente estima-se a quantidade de marcações de pista ao

³Com o intuito de se evitar máximos locais espúrios, impôs-se que $R_{mm}^t(\tau_{max})$ deve ser ao menos a metade de $R_{mm}^t(0)$.

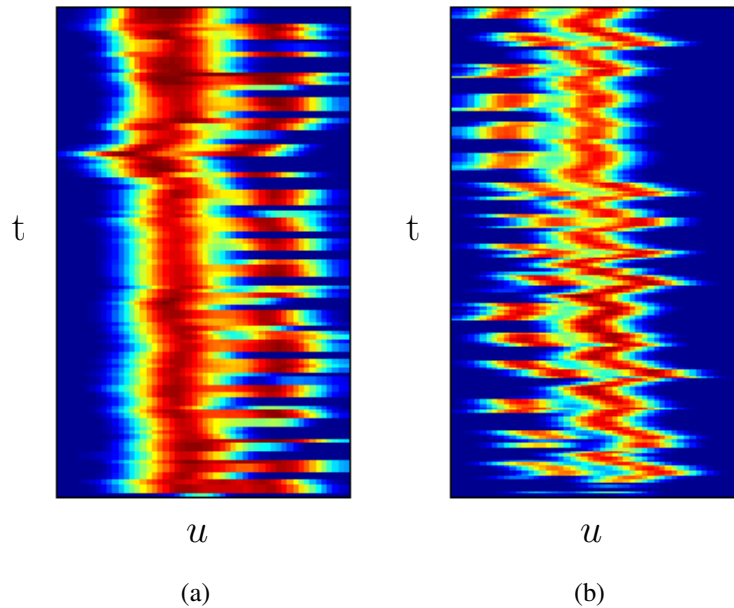
longo da direção horizontal. Mais especificamente, para cada quadro t , calcula-se

$$\rho_b^t(u) = \sum_u r_b^t(u, v), \quad (4.7)$$

a qual provê as evidências acumuladas das marcações na direção vertical u com a região retangular de interesse.

Para as marcas seccionadas, a forma de $\rho_b^t(u)$ irá alternar de um platô, quando nenhuma marca é localmente presente no quadro t , para um pico, quando marcas estão presentes. Para as marcações seccionada-contínua ou contínua-seccionada, a forma de $\rho_b^t(u)$ irá alternar de um gráfico de um único pico (apenas componentes contínuos) para um gráfico com dois picos (componentes seccionado e contínuo). A Figura 4.4 apresenta os gráficos correspondentes as superfícies $\rho_b^t(u)$ como uma função de t e u para partes da pista com marcações contínua-seccionada (Figura 4.5(a)) e seccionada-contínua (Figura 4.5(b)).

Figura 4.4 – Exemplo de evidências de marcações de pista $\rho_b^t(u)$ para (a) contínuo-seccionado e (b) seccionado-contínuo. Valores altos em vermelho e baixos em azul.



Fonte: do autor (2013).

Portanto, calcula-se o número de picos (máximos locais) $n_p(t)$ presentes em $\rho_b^t(u)$ ao longo do eixo u e define-se como atributo f_2 , para o classificador \mathcal{C}_2 , a média do número de picos na janela temporal T_2 :

$$f_2(t) = \frac{1}{T_2} \sum_{u=0}^{T_2-1} n_p(u), \quad (4.8)$$

onde $T_2 = 20$ quadros, definido de forma experimental. Para as amostras oriundas de marcas seccionadas, os valores de $f_2(t)$ devem estar próximos a 0,5, ao passo que as amostras de marcas do tipo seccionada-contínua ou contínua-seccionada devam apresentar valores para $f_2(t)$ próximos a 1,5. Por este motivo, uma simples regra foi adotada para o classificador \mathcal{C}_2 : se $f_2(t) < 1$, a amostra é atribuída a classe ω_1 ; caso contrário, a mesma é atribuída a classe Ω_{22} e o classificador \mathcal{C}_4 irá, posteriormente, discriminar entre ω_2 e ω_3 .

4.1.3 Classificador \mathcal{C}_3

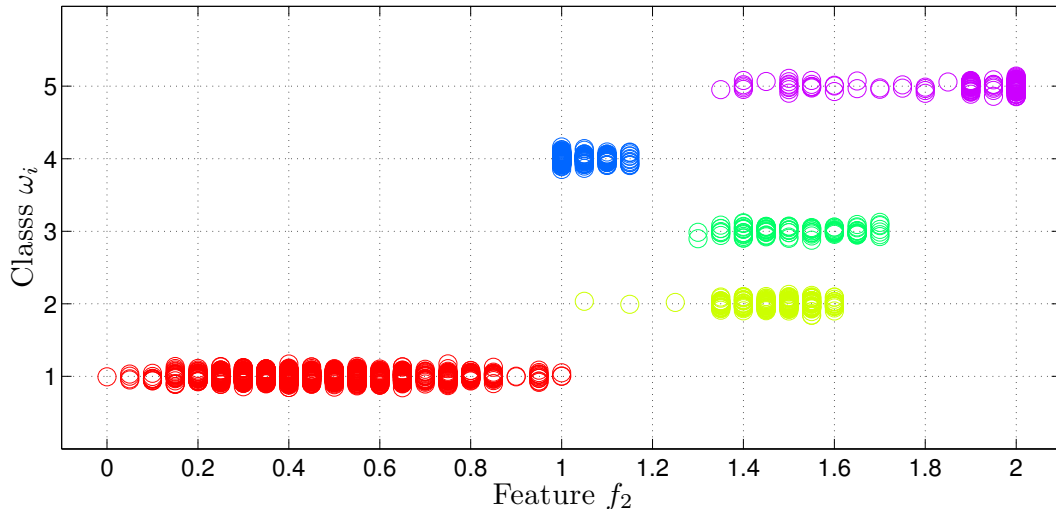
O objetivo do classificador \mathcal{C}_3 é estabelecer uma diferença entre as marcas do tipo contínuo-simples ($\Omega_{31} = \omega_4$) e contínuo-duplo ($\Omega_{32} = \omega_5$). Para as marcas do tipo contínuo-simples, o gráfico esperado de $\rho_b^t(u)$ em todos os quadros t é um único pico relacionado a marcação, enquanto dois picos distintos são esperados quando uma marcação contínuo-duplo está presente. Por este motivo, foi utilizado $f_3(t) = f_2(t)$ para o classificador \mathcal{C}_3 , conforme definido na Equação (4.8). A regra de decisão é, também, muito similar ao classificador \mathcal{C}_2 : se $f_3(t) < 1,5$, a amostra é atribuída a classe ω_4 ; caso contrário, a amostra é atribuída a classe ω_5 .

É importante salientar que o mesmo atributo $f_2(t)$ é usado em ambos os classificadores \mathcal{C}_2 e \mathcal{C}_3 , e os valores esperados para as classes Ω_{21} , Ω_{22} , Ω_{31} e Ω_{32} são, respectivamente, 0,5, 1,5, 1,0 e 2,0. Assim, pode-se pensar que, se um único classificador multiclases, baseado no atributo $f_2(t)$, poderia substituir os classificadores \mathcal{C}_1 , \mathcal{C}_2 e \mathcal{C}_3 . Embora este fato seja possível, os resultados apresentados não foram melhores quanto ao uso do classificador em cascata, visto que, as classes $\Omega_{21} - \Omega_{22}$ e $\Omega_{31} - \Omega_{32}$ apresentam uma maior separação (o que é fundamental para a escolha dos dois classificadores) do que todas as classes em conjunto num único classificador. Logo, a Figura 4.5 mostra a distribuição de f_2 para as diferentes classes, e nota-se que as classes Ω_{22} e ω_5 apresentam uma considerável sobreposição. Já as classes $\Omega_{21} - \Omega_{22}$ e $\Omega_{31} - \Omega_{32}$ apresentam ser disjuntas.

4.1.4 Classificador \mathcal{C}_4

O último classificador na cascata possui como objetivo distinguir marcas do tipo seccionada-contínua ($\Omega_{41} = \omega_2$) e contínua-seccionada ($\Omega_{42} = \omega_3$). Em ambos os casos, o número de picos $n_p(t)$ ao longo de u em $\rho_b^t(u)$ alterna entre um e dois como uma função de t . Todavia, no segundo caso (Ω_{42}), a posição do primeiro pico é consistente para todos os quadros

Figura 4.5 – Distribuição das amostras do atributo f_2 ao longo das cinco classes ω_i . Amostras de diferentes classes são apresentadas em diferentes linhas horizontais.



Fonte: do autor (2013).

t e o segundo pico emerge apenas em alguns quadros (ver Figura 4.5(a)). Para o primeiro caso, o comportamento oposto é esperado: o segundo pico é consistente e o primeiro surge em alguns quadros (ver Figura 4.5(b)).

Para detectar esse comportamento, mantem-se dois *buffers* p_1 e p_2 com as posições dos picos referentes aos quadros anteriores. Quando apenas um pico é identificado, num dado quadro t , a sua posição é comparada aos últimos valores armazenados em p_1 e p_2 , respectivamente, e atribuído ao *buffer* que apresenta a menor diferença absoluta (no outro *buffer*, o valor -1 é armazenado para indicar a ausência de pico naquele quadro).

A cada quadro t , conta-se a quantidade de posições com picos válidos (ou seja, valores diferentes de -1) em cada *buffer* nos últimos T quadros: o *buffer* p_i com o maior número de picos válidos refere-se as marcações contínuas. Enquanto que o *buffer* p_j refere-se as marcas seccionadas. Por fim, calcula-se a média das posições $P_{p_i}(t)$ e $P_{p_j}(t)$ de cada *buffer* (considerando apenas os picos válidos), e decide-se pela classe ω_3 (contínua-seccionada) se $P_{p_i}(t) < P_{p_j}(t)$, e para a classe ω_2 , caso contrário.

4.1.5 Resultados

Para os experimentos realizados foram utilizadas 10 sequências de vídeo adquiridas com duas diferentes câmeras, 2 sequências do *dataset* público *Environment Perception and Driver*

*Assistance*⁴ - set3 (Suburban Bridge and Trailer Follow) e 1 vídeo do *dataset Cambridge-driving Labeled Video Database*⁵ (CamVid) (Veja a Tabela 4.1).

Tabela 4.1 – Sequências de vídeo

Vídeo	Nome	Adquirido	Público	Resolução
1	Assis Brasil	x		480 × 640
2	BR-116 1: Pelotas - POA	x		480 × 640
3	BR-116 6: Pelotas - POA	x		480 × 720
4	BR-290	x		480 × 640
5	Cambará	x		240 × 320
6	Cambridge		x	600 × 800
7	Castelo Branco	x		480 × 640
8	Germany	x		240 × 320
9	Ipiranga	x		480 × 640
10	RS-040	x		480 × 640
11	RS-287: São Francisco - Canela	x		240 × 320
12	Suburban Bridge		x	640 × 800
13	Trailer Follow		x	640 × 800

Fonte: do autor (2013).

Para o primeiro classificador \mathcal{C}_1 , utilizou-se um *kernel* de Função de Base Radial (RBF), definido por:

$$K_\gamma(x, y) = e^{-\gamma\|x-y\|^2}, \quad (4.9)$$

onde $\gamma > 0$ é o parâmetro de largura. Ainda, foi usado um SVM com margem flexível, controlado pelo parâmetro $C > 0$. Foram avaliados de forma experimental diferentes valores para ambos os parâmetros C e γ com a validação cruzada (com o particionamento dos dados usando o método *holdout*) de forma aleatória, selecionando uma parte $g < 1$ para o modelo de treinamento⁶, e outra parte $1 - g$ para a validação. A acurácia da validação cruzada foi de 100% e os parâmetros estimados foram: $C = 1$ e $\gamma = 4$.

O conjunto de treinamento foi composto por 590 instâncias contendo um vetor de atributos tridimensional $\mathbf{f}_1(t)$. Ressalta-se que as sequências de vídeo usadas no processo de treinamento foram particionadas em múltiplas partes (clipes) contendo o mesmo tipo de marcações de pista. Para o conjunto de testes 1, foram utilizadas as 590 amostras restantes, além de amostras de outros clipes, cada um com um único tipo de marcação de pista. Para avaliar a transição

⁴<http://www.mi.auckland.ac.nz/>

⁵<http://http://www.mi.auckland.ac.nz/>

⁶Foi usado $g = 0,5$ em todos os experimentos.

entre diferentes tipos de marcações, outro conjunto de testes foi utilizado (chamado número 2), contendo sequências de vídeo maiores e intercalando diferentes tipos de marcações. As características dos trechos usados nos dois *datasets* estão sumarizadas na Tabela 4.2.

Tabela 4.2 – Trechos de vídeo

Trecho	Tipo	Quadros	Treinamento	Teste 1	Teste 2
Assis Brasil	seccionado	86	x	x	
BR-116 (1)	seccionado	86	x	x	
BR-116 (6)	contínuo-duplo	86	x	x	
BR-290	seccionado	86	x	x	
Cambará (1)	seccionado/contínuo	36	x	x	
Cambará (3)	seccionado/contínuo	66	x	x	
Cambará (5)	contínuo/seccionado	55	x	x	
Cambridge (1)	contínuo-simples	46	x	x	
Cambridge (2)	seccionado	46	x	x	
Castelo Branco	seccionado	86	x	x	
Germany	seccionado	86	x	x	
Ipiranga (1)	seccionado	86	x	x	
Ipiranga (2)	seccionado	36	x	x	
RS-040 (1)	mais do que um tipo	10797			x
RS-287 (1)	contínuo/seccionado	35	x	x	
RS-287 (2)	mais do que um tipo	706			x
Suburban (1)	seccionado	126		x	
Suburban (2)	contínuo-simples	46	x	x	
Suburban (4)	mais do que um tipo	717			x
Trailer Follow	seccionado	86	x	x	

Fonte: do autor (2013).

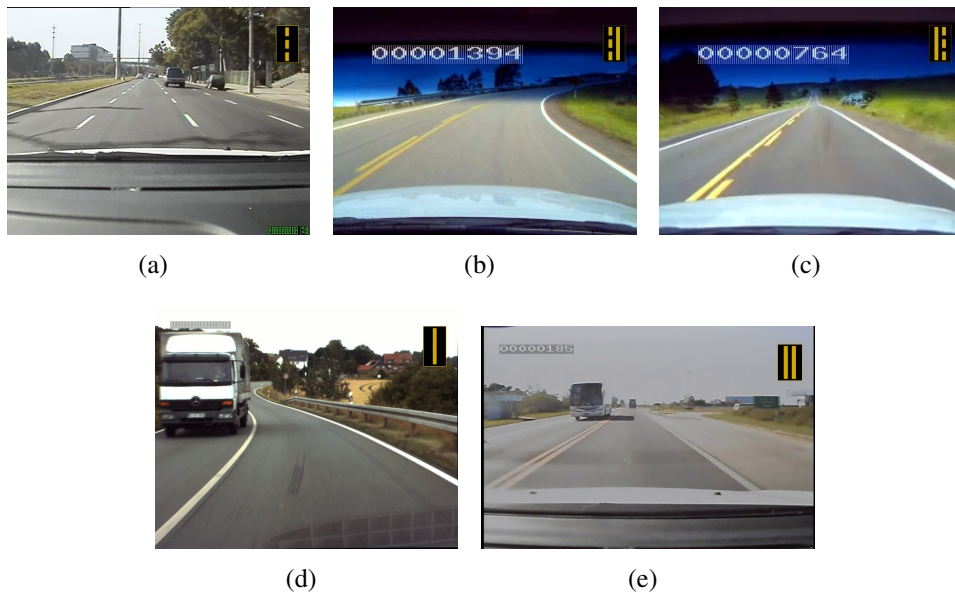
A matriz de confusão referente aos resultados de classificação para o conjunto de testes 1 é apresentada na Tabela 4.3. Salienta-se que o número de amostras utilizadas considera as 590 de treinamento, totalizando 1180 amostras. A acurácia global para o conjunto de testes 1 foi de 99,32%, com apenas algumas amostras da classe ω_1 sendo classificados como ω_2 ou ω_4 .

Tabela 4.3 – Matriz de confusão para o conjunto de testes 1

		Classe predita					
		ω_1	ω_2	ω_3	ω_4	ω_5	
Classe real	ω_1	802	7	0	1	0	99,01%
	ω_2	0	102	0	0	0	100%
	ω_3	0	0	90	0	0	100%
	ω_4	0	0	0	92	0	100%
	ω_5	0	0	0	0	86	100%
						99,32%	

Fonte: do autor (2013).

Figura 4.6 – Exemplos de diferentes tipos de marcações contidas no *dataset 1*.

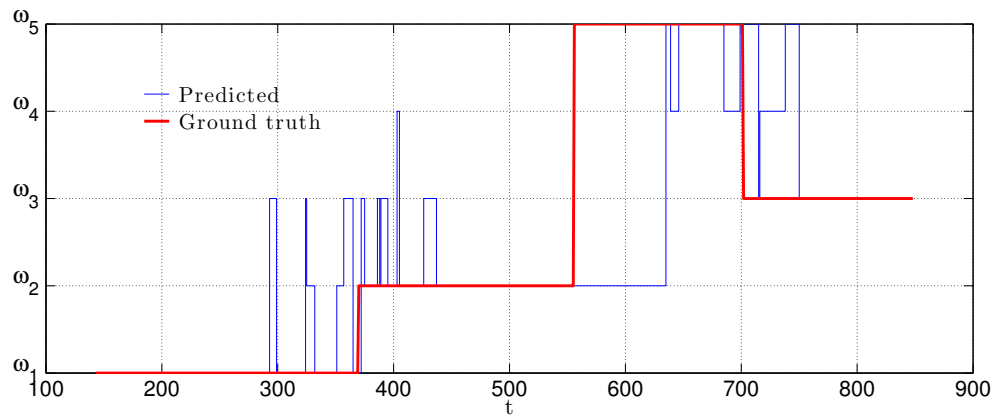


Fonte: do autor (2013).

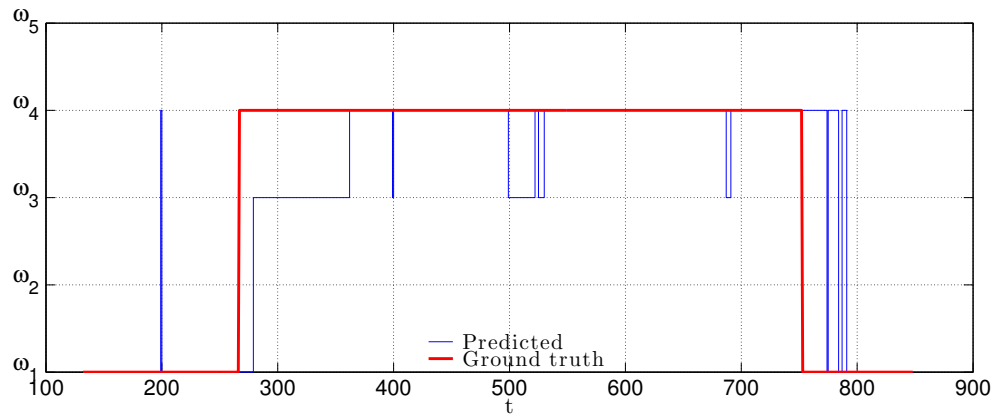
Como mencionado anteriormente, o conjunto de testes 2 contém seqüências de vídeos mais longas e com diferentes tipos de marcas de pista. Uma vez que todas as características usadas no esquema de classificação proposto explorem uma janela temporal, transições ao longo de diferentes tipos de marcação são um problema (uma janela temporal pode conter diferentes tipos de marcações). Os gráficos apresentados na Figura 4.7 ilustram os valores de *ground truth* (ao longo do tempo) para as três seqüências de vídeo usadas no conjunto de testes número 2, bem como os resultados da classificação produzidos pela abordagem proposta.

O primeiro gráfico mostra que há algumas flutuações em torno das classes, mas os resultados são, em sua maioria, coerentes. Pode-se ainda observar que existe um atraso (*lag*) quando acontecem alterações nos tipos de marcações. De fato, este comportamento é esperado, uma vez que se utiliza uma janela temporal com $T = 100$ quadros para calcular os atributos (o atraso é de aproximadamente 50 quadros, ou seja, a metade do tamanho da janela temporal). A matriz de confusão é mostrada na Tabela 4.4, e onde houve erros na classificação, foram de amostras da classe ω_5 como ω_2 , relacionados a troca do tipo de marcação em torno do quadro 550. Apesar disso, a taxa de classificação global foi de 71,53%.

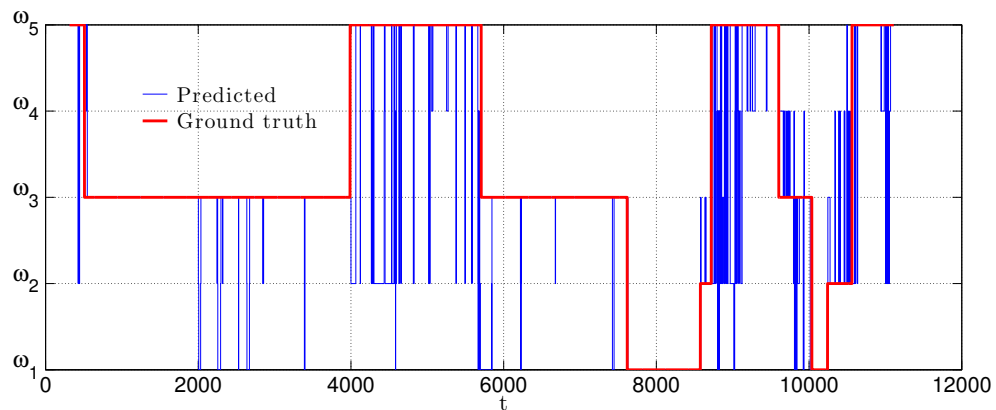
Figura 4.7 – *Ground truth* (em vermelho) e os resultados obtidos (em azul) para as três sequências de vídeo usadas no conjunto de testes 2.



(a)



(b)



(c)

Fonte: do autor (2013).

No segundo vídeo, *Suburban (4)*, nota-se o atraso durante as alterações do tipo de marcação (em torno dos quadros 280 e 750). A Tabela 4.5 mostra a matriz de confusão desse experimento, e a respectiva taxa classificação foi de 77,27%.

Tabela 4.4 – Matriz de confusão para o vídeo *RS-287 (2)*

		Classe predita					
		ω_1	ω_2	ω_3	ω_4	ω_5	
Classe real	ω_1	199	13	15	0	0	87,67%
	ω_2	2	160	22	2	0	86,02%
	ω_3	0	0	100	22	25	68,03%
	ω_4	0	0	0	0	0	—
	ω_5	0	79	0	21	46	31,51%
						71,53%	

Fonte: do autor (2013).

O último gráfico, apresentado na Figura 4.8(c), refere-se a maior sequência de vídeo (*RS-040 (1)*), como mais de 10.000 quadros). Como visto nos dois vídeos anteriores, este também apresenta algumas flutuações espúrias em torno das classes, particularmente, em torno dos quadros 4.000-5.000 e 8.700-10.000. Parte dessas flutuações podem ser explicadas pela presença da sinalização horizontal (como mostrado na Figura 4.9), uma vez que a região usada para extrair as marcações de pista não está localizada em regiões sem marcação. A taxa global de classificação para esta longa sequência de vídeo foi de 85,42% e a matriz de confusão é mostrada na Tabela 4.6. Cabe ressaltar que nenhum dos quadros dessa sequência de vídeo foram usadas para o treinamento do classificador SVM \mathcal{C}_1 . Ainda, salienta-se que o número de objetos classificados é altamente dependente da qualidade das marcações de pista.

Tabela 4.5 – Matriz de confusão para o vídeo *Suburban (4)*

		Classe predita					
		ω_1	ω_2	ω_3	ω_4	ω_5	
Classe real	ω_1	196	0	0	35	0	84,85%
	ω_2	0	0	0	0	0	—
	ω_3	0	0	0	0	0	—
	ω_4	12	0	116	358	0	73,66%
	ω_5	0	0	0	0	0	—
						77,27%	

Fonte: do autor (2013).

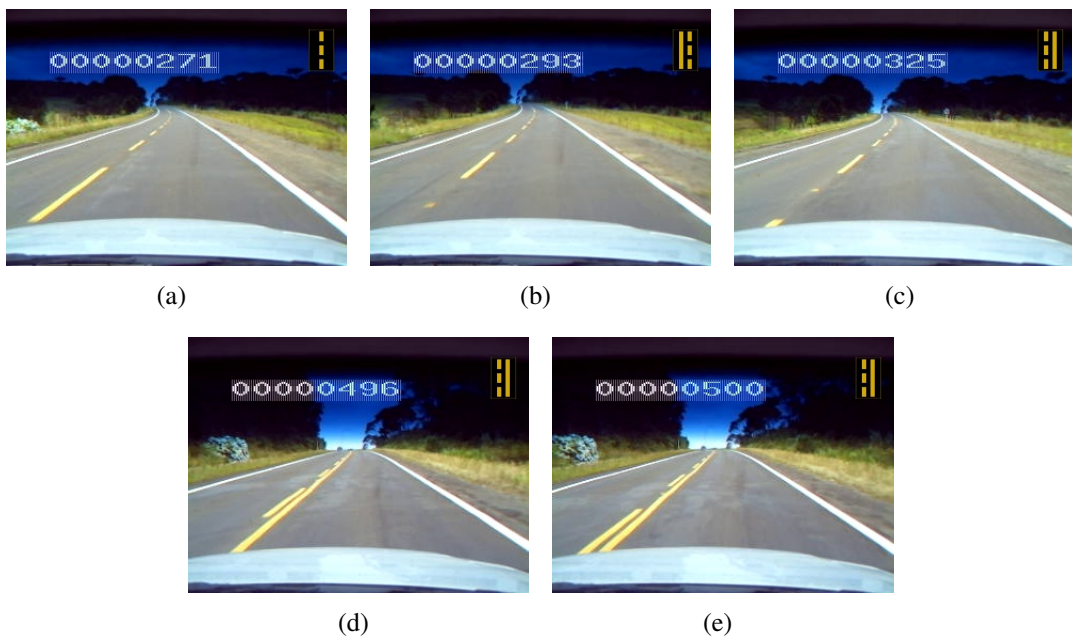
Os experimentos realizados nesta seção também foram implementados em C++ (compilador GCC 4.2.1), usando o OpenCV Versão 2.4.4 para detectar e rastrear a pista, bem como extrair os atributos necessários. Os testes foram realizados em um *notebook* com processador Core 2 Duo de 2,53GHz com 8GB de RAM. Ainda, foi usada a biblioteca libSVM (CHANG; LIN, 2011) para o classificador SVM. O tempo médio de execução para cada quadro, para diferentes resoluções, está descrito na Tabela 4.7. Como pode ser visto, o sistema proposto

Tabela 4.6 – Matriz de confusão para o vídeo *RS-040 (1)*

		Classe predita					
		ω_1	ω_2	ω_3	ω_4	ω_5	
Classe real	ω_1	1156	0	8	0	0	99,31%
	ω_2	6	374	44	37	2	80,78%
	ω_3	319	76	5296	95	38	90,93%
	ω_4	0	0	0	0	0	—
	ω_5	47	623	26	253	2397	71,64%
						85,42%	

Fonte: do autor (2013).

Figura 4.8 – Quadros da sequência de vídeo *RS-287 (2)*.



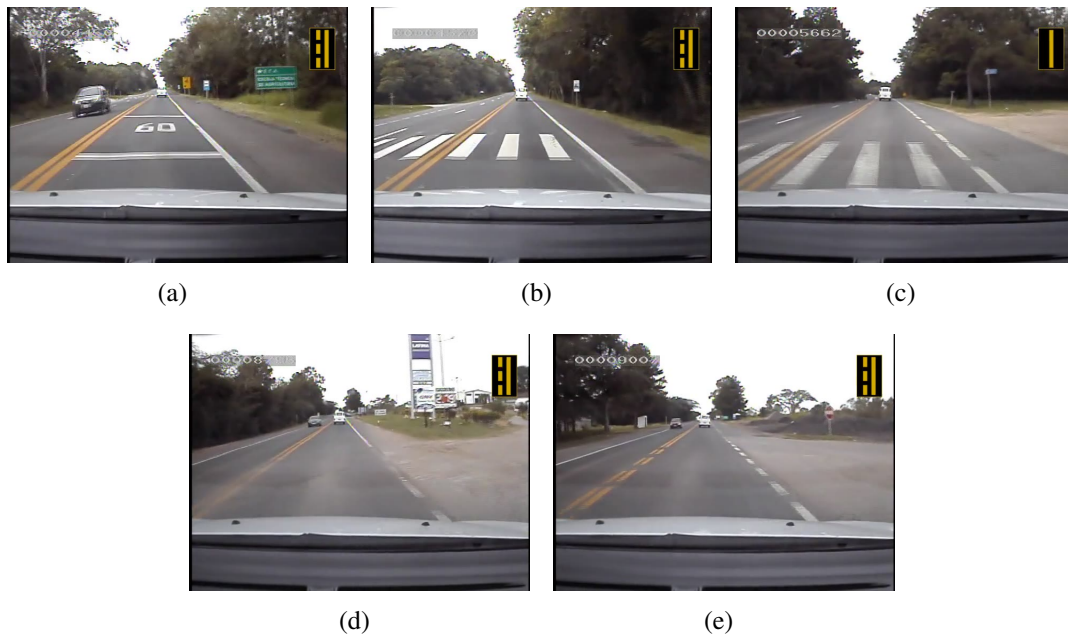
Fonte: do autor (2013).

poderia ser executado a 30 qps para resoluções até 480×640 (VGA), e o código ainda pode ser otimizado para reduzir o tempo de execução.

Tabela 4.7 – Tempo médio de execução.

Resolução	Tempo (ms)
240×320	7,82 ms
480×640	25,10 ms
480×720	94,53 ms
600×800	165,65 ms

Fonte: do autor (2013).

Figura 4.9 – Quadros da sequência de vídeo *RS-040 (1)*.

Fonte: do autor (2013).

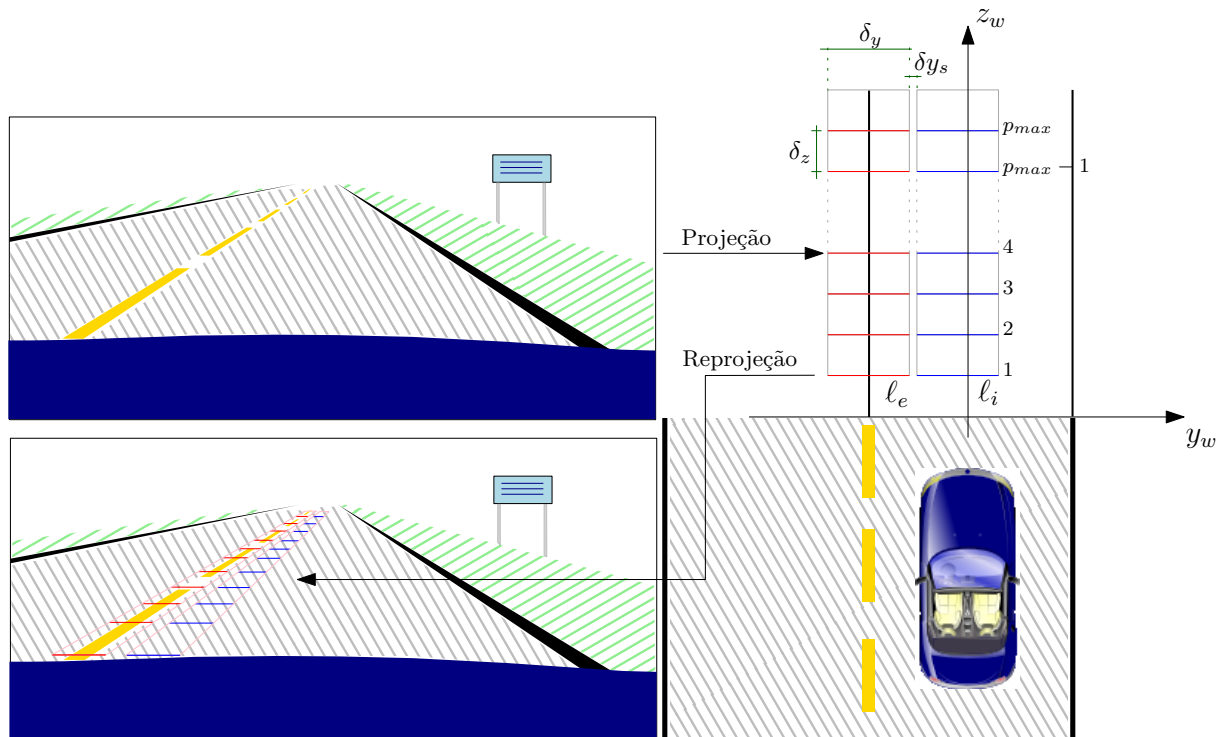
4.2 Análise da sinalização horizontal usando câmera calibrada

Em decorrência do atraso gerado durante as transições de marcações de pista, utilizou-se a informação de câmera para realizar uma análise da marcação num único quadro, sem a necessidade do uso de uma janela temporal para estimar o tipo de marcação. Diferentemente da técnica apresentada na Seção 4.1, a partir das bordas de pista detectadas, utiliza-se o método de calibração de câmera, descrito no Capítulo 3, para extrair os parâmetros extrínsecos. Após a câmera ser calibrada, a próxima etapa consiste em extrair dados referentes a geometria local das marcações de pista. Como citado em Satzoda e Trivedi (2013), o cálculo da variação da intensidade ao longo da direção horizontal com linhas de varredura pode ser útil para extrair as marcações de pista. Entretanto, a metodologia proposta dos autores explora a varredura em toda imagem e um limiar global é usado para recuperar bordas relevantes, o que torna o método dependente das condições de iluminação, da cor e qualidade da pintura. A principal ideia do presente método é fixar segmentos unidimensionais ao longo das bordas de pista detectadas e avaliar o comportamento da primeira derivada.

Dada uma imagem de entrada com as respectivas bordas de pista detectadas, calcula-se a vista aérea (ou de topo - *bird's eye view*) usando informação da câmera calibrada, e definem-se p_{max} segmentos de linhas horizontais l_e^p com comprimento de δ_y metros alinhados ao centro da

borda da pista, para $p = 1, 2, \dots, p_{max}$, e igualmente espaçados por δ_z metros. Ainda, utiliza-se um conjunto de segmentos de linhas horizontais internos ℓ_i^p , com as mesmas dimensões e espaçamento de ℓ_e , mas deslocados δy_s unidades para o interior da pista. As linhas horizontais de varredura ℓ_e e ℓ_i são projetadas de volta para a imagem original, como ilustrado na Figura 4.10.

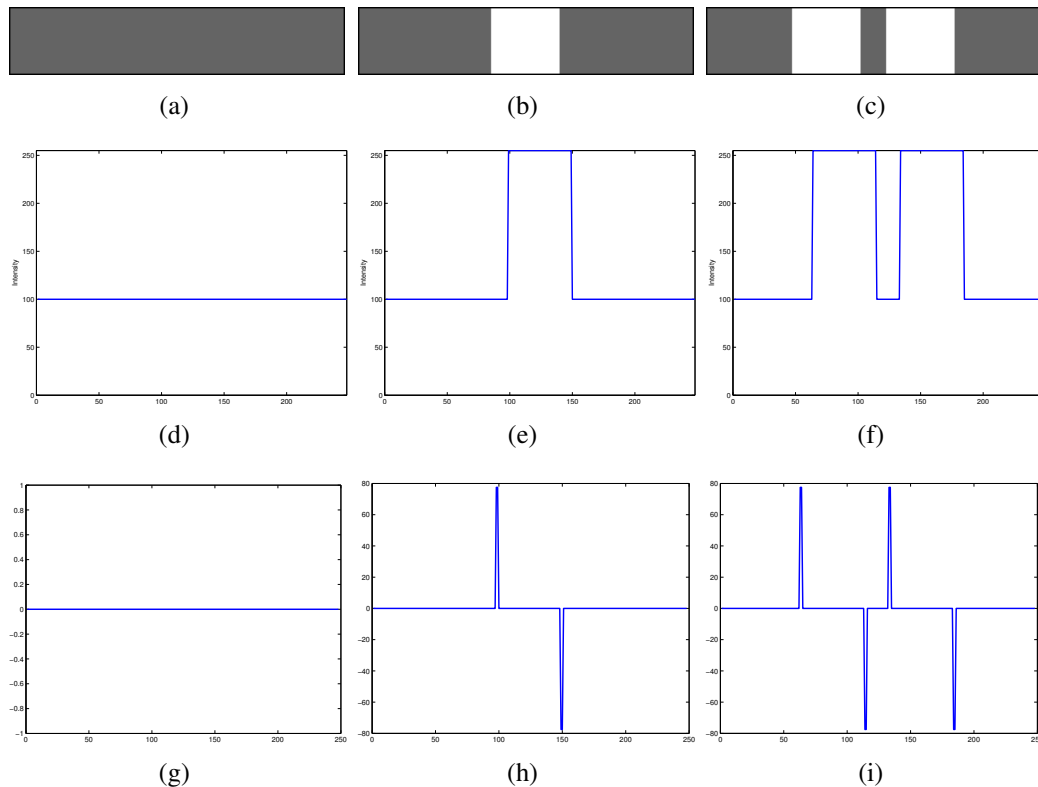
Figura 4.10 – Projeção das bordas de pista do sistema de coordenadas de imagem (ICS) para o sistema de coordenadas de mundo (WCS). Definição dos segmentos de linha em coordenadas de mundo. Reprojecção dos segmentos de linha para coordenadas de imagem.



Fonte: do autor (2015).

A partir de ℓ_e^p , a imagem $I_e^p(v)$ ao longo da coordenada horizontal v deve estar relacionada com a geometria da marcação de pista: se nenhuma marcação está presente, $I_e^p(v)$ tende a ser aproximadamente constante, tal que sua derivada $I_e^{p'}(v)$ deva ser baixa e plana. Se uma marcação horizontal simples está presente (como os tipos de marcação seccionada ou contínua), $I_e^{p'}(v)$ deve apresentar um máximo local forte e outro mínimo local forte (isto é, um par de picos), e se duas marcações lateralmente adjacentes estão presentes (como em marcações do tipo contínua dupla, contínua-seccionada ou seccionada-contínua), $I_e^{p'}(v)$ deve apresentar dois pares de picos. Estas três situações são ilustradas na Figura 4.11. As Figuras 4.12(a) a 4.12(c) mostram três imagens sintéticas com diferentes tipos de marcas horizontais locais. As Figuras 4.12(d) a 4.12(f) apresentam a intensidade dos níveis de cinza dos *pixels* de $I_e^p(v)$ ao longo de

Figura 4.11 – (a) Imagem sem marcação. (b)-(c) Imagem com uma e duas marcações, respectivamente. (d)-(f) Linha horizontal unidimensional (1D) na escala de cinza ao longo do centro da imagem. (g)-(i) Respectivas derivadas de cada linha horizontal da imagem.

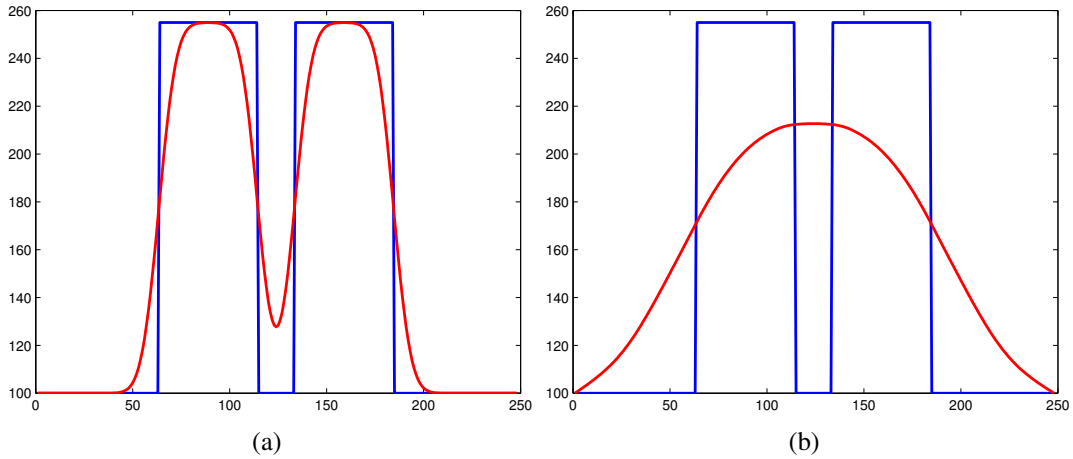


Fonte: do autor (2015).

uma varredura horizontal, e suas respectivas derivadas $I_e^p(v)$ são mostradas nas Figuras 4.12(g) a 4.12(i).

Entretanto, a detecção de picos (ou a detecção de bordas num caso geral) não é uma tarefa trivial em cenários práticos e reais. As condições de iluminação, cor das marcações e a pintura fraca podem gerar pequenos picos a partir de $I_e^p(v)$. De forma recíproca, ruídos, textura do pavimento e outros artefatos podem produzir picos espúrios. A suavização de imagens e sinais é uma solução clássica, mas a escolha do tamanho do *kernel* é um problema: *kernels* grandes filtram mais ruídos, mas tendem a borrar bordas que estão próximas, como previsto pela teoria espaço-escala (WITKIN, 1983). Assim, o tamanho do *kernel* deve ser grande o suficiente para remover os ruídos, mas ao mesmo tempo, suficientemente pequeno ao ponto de preservar as bordas que são próximas, as quais são um cenário comum em se tratando de detectar as marcações de pista (veja a Figura 4.12(f)). A Figura 4.12 ilustra o possível problema de interferência nas bordas quando um *kernel* com tamanho grande é usado.

Figura 4.12 – Efeitos do uso de um filtro Gaussiano: (a) Segmento horizontal 1D com valor adequado de σ : original (azul) e filtrado (vermelho). (b) Segmento horizontal 1D com valor alto de σ , fundindo as bordas vizinhas em uma única borda.



Fonte: do autor (2015).

No mapeamento perspectivo inverso (IPM) da imagem, o tamanho do *kernel* deve ser baseado na distância lateral entre as marcações (nos casos das marcações contínua dupla, contínua-seccionada ou seccionada-contínua), que são constantes ao longo de toda a imagem, de modo que um *kernel* de tamanho máximo poderia ser definido. Na abordagem proposta, entretanto, todo o processamento é realizado diretamente no domínio da imagem original. Na imagem projetada, o espaçamento entre as marcas varia ao longo dos segmentos devido à perspectiva da câmera. Felizmente, o uso da câmera calibrada permite mapear o maior valor para o tamanho do *kernel* em WCS para o espaço de imagem. Mais precisamente, para cada segmento de linha p , filtra-se a imagem $I_e^p(v)$ usando um filtro Gaussiano $g_p(v)$, dado por

$$g_p(v) = \frac{1}{\sqrt{2\pi\sigma_p^2}} e^{-\frac{v^2}{2\sigma_p^2}}, \quad (4.10)$$

onde $\sigma_p = \text{proj}(\sigma_w, u_p)$ é o comprimento da projeção do segmento de linha horizontal com comprimento original σ_w (ajustado para $0.020m$ de acordo com os experimentos realizados) em WCS para ICS usando a coordenada de imagem vertical u_p , com base nas Equações (3.1)-(3.2).

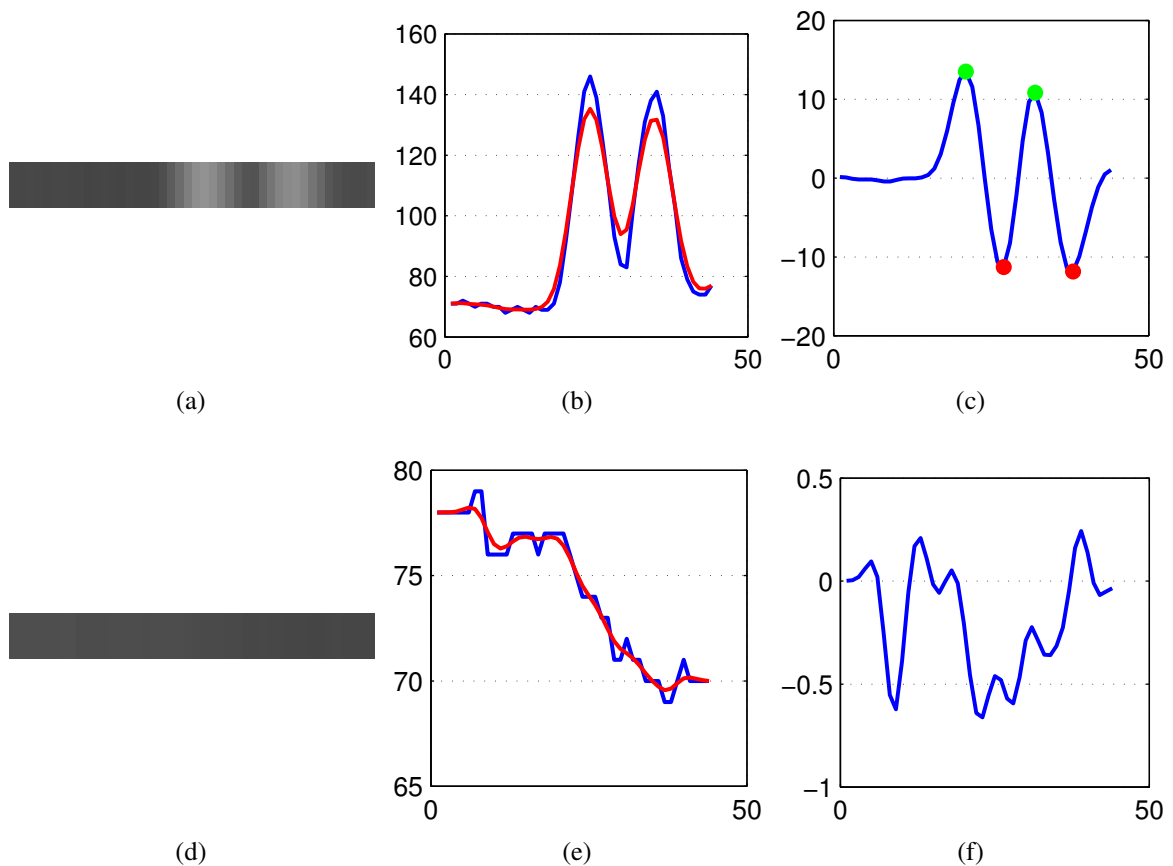
Uma vez que cada marcação produz um máximo local (quando a intensidade aumenta) e um mínimo local (quando a intensidade diminui), a próxima etapa consiste em encontrar pares de picos adjacentes de máximos-mínimos para cada segmento de linha, que devem possuir alturas semelhantes (uma vez que o pavimento apresenta tipicamente uma cor e a marcação

apresenta outra). Para esse propósito, todos os extremos locais v_j são computados, e o máximo global $M_{max}^p = \max_j |I_e^{p'}(v_j)|$ é extraído. Após, recupera-se apenas o extremo local cuja altura (em valores absolutos) é maior que εM_{max}^p , onde $0 < \varepsilon < 1$ é o limiar usado para remover pequenos picos (o valor de $\varepsilon = 0.35$ foi definido experimentalmente).

É importante ressaltar que o limiar ε é relativo à máxima altura do pico com relação a ℓ_e . Mesmo assim, os segmentos de linha sem marcação continuarão a produzir pequenos picos relacionados a irregularidades/ruídos que não foram removidos após a aplicação do limiar. Como uma última etapa de validação, recupera-se cada par remanescente de picos adjacentes de máximo-mínimo, e calcula-se a diferença de altura entre picos do j -ésimo par $\delta h_j^p = I_e^{p'}(v_{max}^j) - I_e^{p'}(v_{min}^j)$, onde v_{max}^j e v_{min}^j são as posições correspondentes dos extremos adjacentes. Em seguida, cada segmento de linha interno ℓ_i (que se presume conter *pixels* relacionados com pavimento) é suavizado com o *kernel* Gaussiano adaptativo apresentado na Equação (4.10). Ainda, calculam-se as derivadas $I_i^{p'}(v)$, extraindo o seu desvio padrão σ_i^p , que representa a irregularidade/ruído com relação aos *pixels* do pavimento. Por fim, o j -ésimo par de picos é validado se $\delta h_j^p > k\sigma_i^p$, onde k é o número de desvios padrões usado na regra de rejeição (os experimentos indicaram que $k = 10$ apresenta um bom compromisso entre as detecções errôneas e falsas). A Figura 4.13 ilustra o processo de extração dos picos.

É importante mencionar que o processo de extração de marcações de pista é realizado em uma imagem monocromática. Apesar do uso da informação de cor de forma explícita possa ser útil, a maioria das marcações de pista são brancas ou amarelas, enquanto o pavimento é tipicamente em escala de cinza (asfalto), ou seja, não possui informação de crominância. Uma vez que utiliza-se a luminância $Y = 0.299R + 0.587G + 0.114B$ é usada como imagem de entrada, e a cor amarela é principalmente distinguível nos canais vermelho e verde (que carregam maiores pesos no cálculo da luminância), tem-se suficiente contraste para se detectar as marcações de pista (Figura 4.14).

Figura 4.13 – Exemplo da extração de picos ao longo das bordas da pista. (a) imagem contendo a marcação contínua-dupla, (b) intensidades original (azul) e filtrada (vermelho), (c) derivada da intensidade da imagem filtrada e pares de picos detectados. (d)-(f) apresenta as imagens análogas, mas para uma imagem sem marcação.



Fonte: do autor (2015).

4.2.1 Classificação das marcações de pista

Para cada segmento de linha p , o número de pares de picos extraídos n_p caracteriza a geometria local (horizontal) da marcação de pista: nenhum (zero) par indica que não há marcação; um par (1) refere-se a marcação do tipo simples; e dois (2) pares significa que existe duas marcações lateralmente adjacentes. A caracterização das marcações de pista ao longo de um quadro da imagem t em análise é dado pela composição de p_{max} valores $n_p(t)$: em marcações seccionadas (classe ω_1), $n_p(t)$ deve variar entre 0 e 1; para marcações seccionada-contínua (ω_2) ou contínua-seccionada (ω_3), entre 1 e 2; para contínua-simples (ω_4), deve ser geralmente 1; e para contínuo-duplo (ω_5), geralmente 2.

Uma vez que a análise de $n_p(t)$ não pode ser usada para distinguir entre as marcações seccionada-contínua e contínua-seccionada, um classificador em dois níveis foi desenvolvido,

Figura 4.14 – Imagem com marcação de pista: (a) imagem RGB, (b) canal R, (c) canal G, (d) canal B.

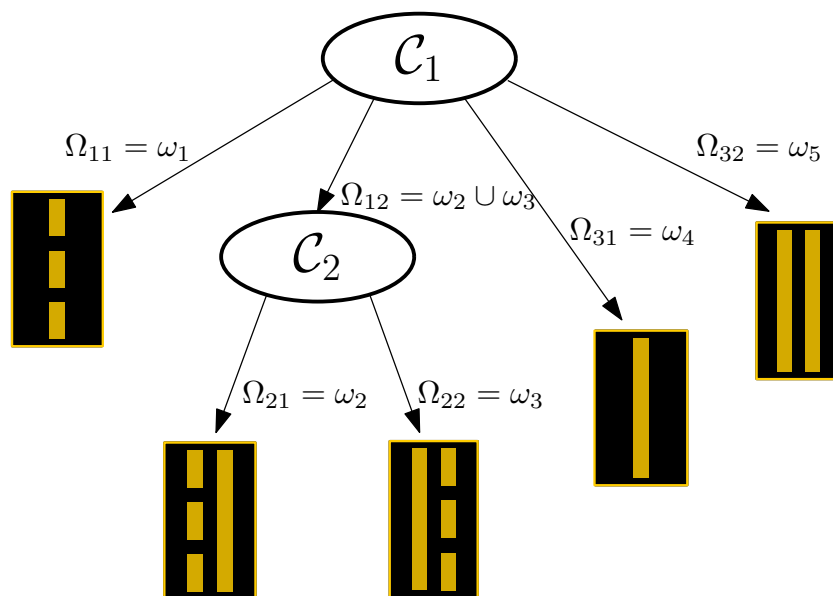


Fonte: do autor (2015).

como apresentado na Figura 4.15.

Analogamente à técnica descrita na Seção 4.1, um classificador em cascata é explorado, consistindo de dois classificadores \mathcal{C}_1 e \mathcal{C}_2 . Seja Ω_{ij} a classe de saída do classificador \mathcal{C}_i , como também apresentado na Figura 4.15.

Figura 4.15 – Diagrama esquemático do classificador de dois estágios.



Fonte: do autor (2015).

4.2.2 Classificador \mathcal{C}_1

O primeiro classificador consiste em identificar as marcações de pista em quatro categorias: seccionada (Ω_{11}), contínua-simples (Ω_{13}), contínuo-duplo (Ω_{14}) e mista (Ω_{12}), conhecida como seccionada-contínua ou contínua-seccionada.

Para cada quadro t , definiu-se um conjunto p_{max} de valores $n_p(t)$ relacionados ao p -ésimo segmento de linha, com os possíveis valores 0, 1 ou 2. Como o pavimento deve conter problemas devido a inúmeros fatores, aplica-se o filtro da mediana ao longo de p em $n_p(t)$ com o propósito de remover os segmentos de linha com respostas de picos isolados, provavelmente devido a *outliers* (valores atípicos). Após o filtro da mediana, propõe-se o seguinte vetor de características para o classificador \mathcal{C}_1 :

$$\mathbf{f}_1(t) = \begin{pmatrix} f_{11}(t) \\ f_{12}(t) \end{pmatrix} = \begin{pmatrix} \#\{n_p|n_p = 0\}/p_{max} \\ \#\{n_p|n_p = 1\}/p_{max} \end{pmatrix}, \quad (4.11)$$

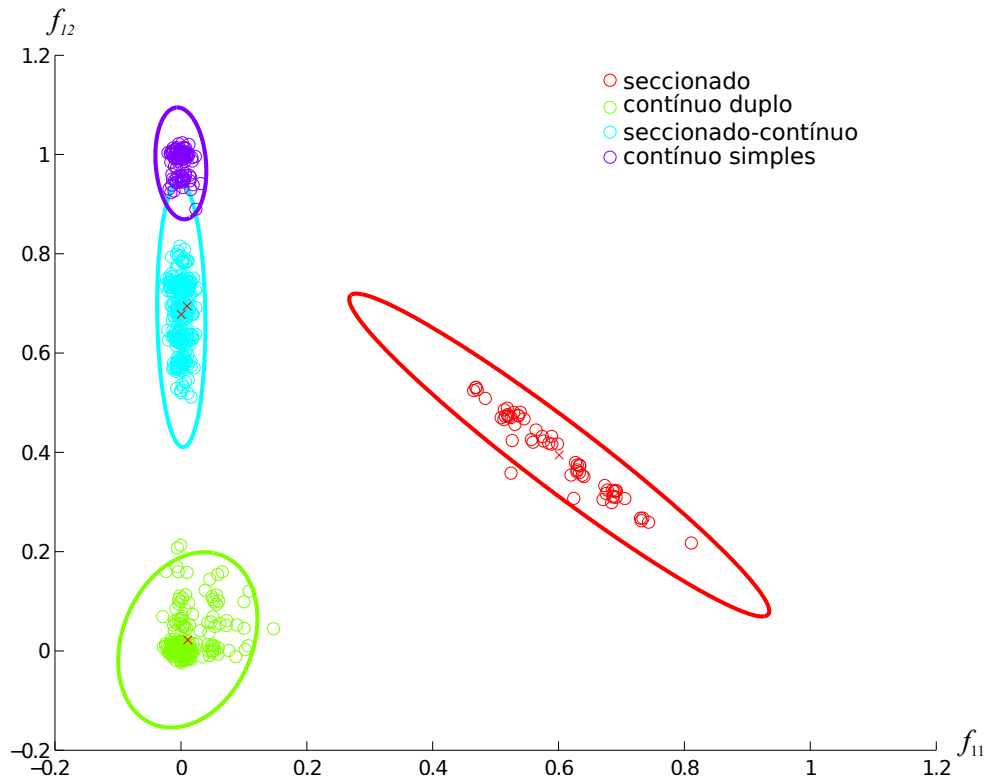
onde $\#$ denota o número de elementos no conjunto. Logo, $f_{11}(t)$ contém a fração dos segmentos de linha que não apresentam picos, ou seja zero (0) picos, e $f_{12}(t)$ a fração com um par de picos (a fração dos segmentos de linha com dois picos estão contidos em $f_{11}(t)$ e $f_{12}(t)$, uma vez que a soma das frações é igual a um).

A Figura 4.16 apresenta um gráfico de dispersão do vetor de características $\mathbf{f}_1(t)$ referente às quatro classes do classificador \mathcal{C}_1 e observa-se que elas formam agrupamentos bem separados num espaço de características 2D. A mesma figura ainda apresenta os contornos equiprováveis (elipses) de uma distribuição Gaussiana 2D ajustada em cada classe, indicando que a Mistura de Gaussianas (MoG) é uma boa escolha para modelar a distribuição global.

Por fim, foi escolhido um classificador Bayesiano para \mathcal{C}_1 baseado em uma distribuição Gaussiana com probabilidade a priori igual para todas as classes de modo que a classe escolhida é Ω_{1j} , com

$$j = \arg \max_i P(\Omega_{1i}|\mathbf{f}_1) = \arg \max_i \frac{p(\mathbf{f}_1|\Omega_{1i})}{\sum_{i=1}^4 p(\mathbf{f}_1|\Omega_{1i})}, \quad (4.12)$$

com $p(\mathbf{f}_1|\Omega_{1i}) \sim \mathcal{N}(\boldsymbol{\mu}_i, \mathbf{C}_i)$ sendo a distribuição Gaussiana 2D da classe Ω_{1i} com média $\boldsymbol{\mu}_i$ e matriz de covariância \mathbf{C}_i . A Seção 4.2.5 apresenta em detalhes o conjunto de treinamento e testes usados nos experimentos.

Figura 4.16 – Distribuição das classes de f_1 para diferentes classes.

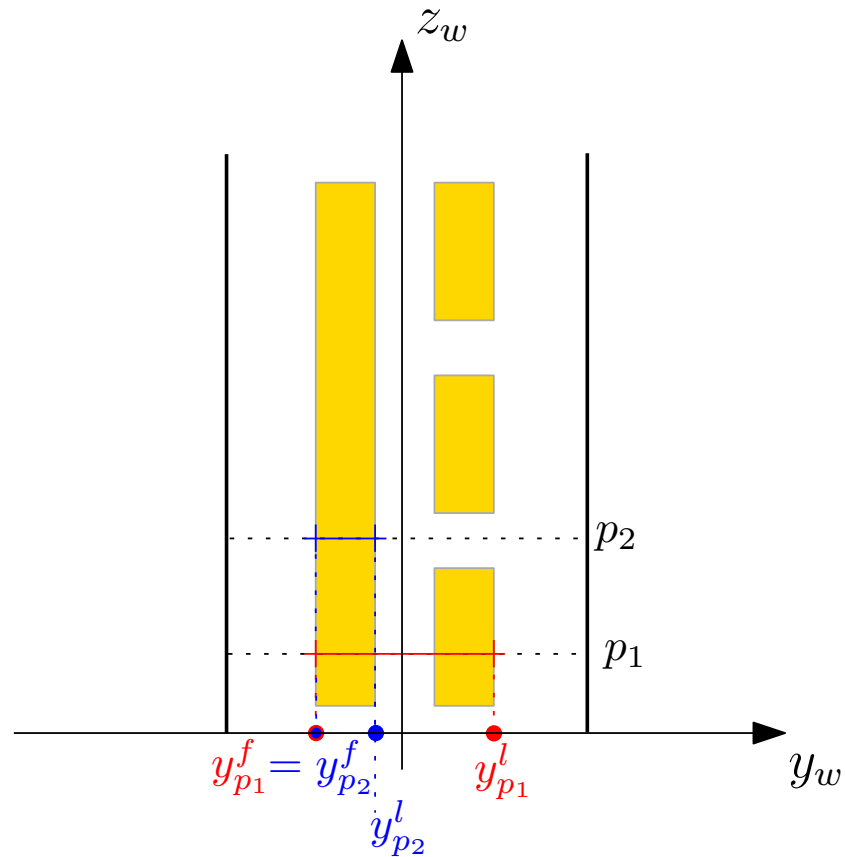
Fonte: do autor (2015).

4.2.3 Classificador \mathcal{C}_2

Como mostrado na Figura 4.15, o classificador \mathcal{C}_2 é suprido com amostras que possuem componentes seccionados e contínuos. Assim, o seu objetivo é distinguir marcações seccionada-contínua ($\Omega_{21} = \omega_2$) das marcações contínua-seccionada ($\Omega_{22} = \omega_3$). Para este propósito, é avaliada a posição do máximo e mínimo local ao longo de diferentes segmentos de linha num dado quadro e no sistema de coordenadas de mundo (WCS). Para o tipo de marcação contínuo-seccionada, o primeiro máximo local tende a ocorrer aproximadamente na mesma coordenada y_w , enquanto que a localização do último mínimo local varia ao longo do eixo, como pode ser visualizado na Figura 4.17. Para as marcações do tipo seccionada-contínua, a variação deve ocorrer no primeiro máximo.

Para detectar tal comportamento, procura-se identificar a localização do primeiro máximo y_p^f e do último mínimo y_p^l para cada segmento de linha no WCS, e calculam-se as variâncias (ao longo de p) $\sigma^2(y^f)$ e $\sigma^2(y^l)$. Então, uma simples regra é adotada para o classificador

Figura 4.17 – Estimando as ordenadas do máximo e mínimo local.



Fonte: do autor (2015).

\mathcal{C}_2 : se $\sigma^2(y^f) < \sigma^2(y^l)$, o quadro é atribuído à classe ω_3 (contínuo-seccionado); caso contrário, é atribuído à classe ω_2 (seccionada-contínua).

4.2.4 Melhorando os resultados da classificação

4.2.4.1 Detecção de outliers

Quadros que contém dois tipos de marcações de pista (na transição entre quadros) tendem a gerar vetores de atributos $f_1(t)$ que não se encaixam em nenhuma das quatro classes do classificador \mathcal{C}_1 , mostradas na Figura 4.16. Ruídos, oclusões, asfalto degradado ou a má condição de pintura da sinalização horizontal também podem gerar vetores de atributos espúrios. Entretanto, mesmo nestes casos, grandes valores a posteriori podem ser produzidos pela Equação (4.12). Uma vez que a função de distribuição de probabilidade (FDP) das classes são

Gaussianas bivariadas, a distância de Mahalanobis ao quadrado

$$d(\mathbf{f}_1; \boldsymbol{\mu}_i, \mathbf{C}_i)^2 = (\mathbf{f}_1 - \boldsymbol{\mu}_i)^T \mathbf{C}_i^{-1} (\mathbf{f}_1 - \boldsymbol{\mu}_i) \quad (4.13)$$

pode ser usada para avaliar quão bem um dado vetor exemplo \mathbf{f}_1 se encaixa em uma distribuição Gaussiana $\mathcal{N}(\boldsymbol{\mu}_i; \mathbf{C}_i)$. De fato, os valores $d(\mathbf{f}_1; \boldsymbol{\mu}_i; \mathbf{C}_i)^2$ seguem uma distribuição chi-quadrado com 2 graus de liberdade (χ_2^2), tal que os *outliers* podem ser detectados se eles ocorrem na cauda da distribuição. Mais precisamente, seleciona-se um limiar $\epsilon(\rho)$ definido com base no quantil ρ da distribuição chi-quadrado (valor crítico da cauda superior) com nível de probabilidade ρ como regra de detecção de *outliers*. Se $c(t) \in \{\Omega_{11}, \Omega_{12}, \Omega_{13}, \Omega_{14}\}$ denota o rótulo da classe para o vetor de características $\mathbf{f}_1(t)$, baseado na Equação (4.12), a classe atualizada $\hat{c}(t)$ é dada por

$$\hat{c}(t) = \begin{cases} \hat{c}(t-1), & \text{se } d(\mathbf{f}_1(t); \boldsymbol{\mu}_{c(t)}, \mathbf{C}_{i(t)})^2 > \epsilon(\rho) \\ c(t), & \text{caso contrário} \end{cases}, \quad (4.14)$$

significando que a última classe confiável estimada é mantida quando um *outlier* é detectado. Para todos os experimentos realizados, foi selecionado $\rho = 0.975$, levando à $\epsilon(\rho) = 7,37$.

4.2.4.2 Informação temporal

Outro aspecto importante do problema em questão é a continuidade temporal: as marcações de pista da estrada tendem à pertencer a uma classe para um certo conjunto de quadros adjacentes, tal que $\Omega_{ij}(t)$ está propensa a ser uma função constante definida por partes, onde cada platô corresponde a um conjunto de quadros adjacentes com o mesmo tipo de marcação de pista. Embora haja várias alternativas (como HMM - Modelo Oculto de Markov) para impor a coerência temporal, neste trabalho foi adotado uma maneira simples mas efetiva de evitar “*outliers* temporais”. Com uma janela temporal $\mathcal{T}(t) = \{t-T+1, t-T+2, \dots, t-1, t\}$ contendo os últimos T quadros, é calculado e armazenado o tipo de marcação de pista $\hat{c}(t)$ a cada quadro usando a Equação (4.14). O rótulo de classe temporalmente coerente $c'(t)$ no quadro corrente t é dado pelo mais frequente rótulo encontrado na janela temporal, ou seja

$$c'(t) = \underset{\tau \in \mathcal{T}(t)}{\text{moda}} \hat{c}(\tau), \quad (4.15)$$

que corresponde a um esquema de classificação em que o “vencedor-leva-tudo” (*winner-takes-all*) em $\mathcal{T}(t)$.

A escolha da janela com duração T é um compromisso entre a suavidade e o atraso na detecção: altos valores de T reduzem a ocorrência de falsas detecções, mas com o custo de introduzir atrasos na detecção quando o tipo de marcação de pista muda. Foram testados diferentes valores para T e foi verificado que $T = 20$, o que corresponde a 667ms de vídeo capturado a 30 qps, apresenta ser suficiente para controlar as respostas isoladas (*outliers*) ao mesmo tempo em que não introduz atrasos significantes.

4.2.5 Resultados

O algoritmo proposto na Seção 4.2 foi testado com diferentes sequências de vídeo contendo uma variedade de marcações de pista. Mais precisamente, foram testados 12 sequências de vídeo adquiridas com diferentes câmeras (Watec Wat 240R, iPad Mini, iPhone 3G, iPhone4, iPhone5S), resoluções e taxas de amostragem, em diferentes estradas no Brasil. A única câmera analógica usada foi a Watec Wat 240R, equipada com uma lente de 3.8mm, e todos os vídeos da câmera Watec foram digitalizados usando um dispositivo de captura AverTV iPro HD (modelo H869M) a 29,97 qps e com resolução de 480×640 . O conjunto de dados usado neste trabalho com os correspondentes dados de *ground truth*, bem como as sequências de vídeo apresentadas nos resultados do método, estão publicamente disponíveis em <http://inf.ufrgs.br/~mbpaula/roadLaneMarkings/>

Os dados de *ground truth* contêm um rótulo para cada borda de pista (esquerda e direita) a cada quadro. Em quadros em que existam mais de um tipo de marcação visível (isto é, durante transições), dois rótulos foram usados para a correspondente borda da pista, e ambos são considerados como corretos no estágio de classificação.

Para todas as sequências de vídeo, os parâmetros intrínsecos foram obtidos de forma *offline* (apenas uma única vez para cada câmera) usando o *Caltech Camera Calibration Toolbox* para Matlab (BOUGUET, 2008), e os parâmetros extrínsecos foram obtidos em tempo de execução *on-the-fly* de acordo com a abordagem proposta no Capítulo 3, em trechos da sequência de vídeo que: (i) o veículo estava se movendo ao longo de uma região plana e em linha reta da pista, (ii) as marcações eram seccionadas, (iii) a velocidade do veículo neste trecho foi constante e conhecida.

Salienta-se que, com poucos quadros a 30 qps por exemplo, se consegue ter a câmera calibrada em apenas 3~5 segundos. Os parâmetros intrínsecos e extrínsecos de câmera, a velocidade do veículo e o número de quadros usados para obter os respectivos valores são apresentados na Tabela 4.8.

Os parâmetros extrínsecos das sequências de vídeo 5 e 10-12 (em que a velocidade é desconhecida), foram calculados usando os vídeos clip_iPhone3G_1395431301_70kmh e clip_i5s_0085, respectivamente, que apresentam a mesma configuração câmera e velocidade conhecida.

Tabela 4.8 – Clipes: parâmetros de calibração de câmera

Clipe	Nome	u_0	v_0	f_u / f_v^7	α°	β°	h (m)	Velocidade (km/h)	Quadros ⁸
1	clip_watec_DMA291113_HMS120139	319,50	239,50	674,30	-7,70	-0,81	1,16	90	200
2	clip_watec_DMA311013_HMS143419_100kmh	319,50	239,50	674,30	-4,01	2,27	1,44	100	200
3	clip_watec_DMA300114_HMS154757_80kmh	319,50	239,50	674,30	1,17	-0,89	1,16	80	200
4	clip_watec_DMA311013_HMS145905	319,50	239,50	674,30	-4,01	2,27	1,44	100	200
5	clip_iPhone3G_1395434215_01	191,50	143,50	441,50	4,36	-0,75	1,25	desconhecida	-
6	clip_iPhone3G_1395431301_70kmh	191,50	143,50	441,50	4,36	-0,75	1,25	70	150
7	clip_iPad_DMA190414_1	959,50	539,50	2,03e+03	-5,90	0,60	1,08	80	100
8	clip_iPhone4_1807_1560	639,50	395,50	1,64e+03	3,89	-3,10	1,05	60	200
9	clip_i5s_0085	959,50	539,50	1,81e+03	5,99	-0,85	1,22	70	200
10	clip_i5s_0094	959,50	539,50	1,81e+03	5,99	-0,85	1,22	desconhecida	-
11	clip_i5s_0096	959,50	539,50	1,81e+03	5,99	-0,85	1,22	desconhecida	-
12	clip_i5s_0104	959,50	539,50	1,81e+03	5,99	-0,85	1,22	desconhecida	-

Fonte: do autor (2015).

Para a etapa de classificação propriamente dita, os valores de todos os parâmetros que foram usados estão descritos na Seção 4.2, e estão sumarizados na Tabela 4.9. Com base nestes parâmetros e, de acordo com Manual Brasileiro de Sinalização Vertical (CNT, 2007), o número de linhas de varredura foi definido levando-se em consideração: (i) o comprimento entre duas marcas adjacentes ($l_{adj} = 6\text{m}$); (ii) o comprimento de uma marca seccionada ($l_{mk} = 2\text{m}$). Portanto, o número de linhas de varredura usados para analisar as marcações no pavimento é expresso por $p_{max} = \lfloor 1 + (1,5l_{mk} + l_{adj})\delta_z^{-1} \rfloor = 19$, onde $\lfloor \cdot \rfloor$ denota o operador de arredondamento para o inteiro mais próximo.

Tabela 4.9 – Sumário das variáveis e valores padrão

Símbolo	Definição	Valor padrão
p_{max}	número de segmentos de linhas de varredura	19
δ_z	espaçamento entre os segmentos as linhas de varredura	0,5m
δ_x	largura das linhas de varredura	1,0m
δx_s	distância entre as linhas de varredura l_e e l_i	0,12m
σ_w	desvio padrão do filtro Gaussiano	0,020m
ε	limiar de rejeição para picos espúrios	0,35
$\epsilon(\rho)$	limiar de rejeição para a distância de Mahalanobis	7,37
T	número de quadros usado na janela temporal	20

Fonte: do autor (2015).

⁷A distância focal em ambos os eixos é a mesma.

⁸Número de quadros usados na calibração da câmera.

De maneira a avaliar a robustez e generalização das capacidades do modelo, foi utilizado outro conjunto de sequências de vídeo (clipes 3-12) adquiridos com diferentes câmeras variando em resolução, como mostrado na Tabela 4.10. Para este conjunto, a taxa global de classificação foi 96,36%, com o clipe de menor acurácia de 82,39%. A matriz de confusão apresentada na Tabela 4.11 indica que os erros de classificação ocorrem nas classes ω_3 e ω_4 , sendo classificadas como ω_1 , e ω_4 como ω_3 , as quais estão relacionadas ao classificador \mathcal{C}_1 .

Tabela 4.10 – Resolução, taxa de amostragem e acurácia global dos clipes usados

Clipe	Nome	Resolução	qps	Tipo	Quadros	Treino	Teste	Acurácia
1	clip_watec_DMA291113_HMS120139	480 × 640	29,97	+ de um tipo	990	x	x	99,80%
2	clip_watec_DMA311013_HMS143419_100kmh	480 × 640	29,97	seccionada	201	x	x	100,00%
3	clip_watec_DMA300114_HMS154757_80kmh	480 × 640	29,97	seccionada	261		x	100,00%
4	clip_watec_DMA311013_HMS145905	480 × 640	29,97	+ de um tipo	498		x	100,00%
5	clip_iPhone3G_1395434215_01	288 × 384	14,64	+ de um tipo	73		x	98,63%
6	clip_iPhone3G_1395431301_70kmh	288 × 384	14,64	seccionada	151		x	100,00%
7	clip_iPad_DMA190414_1	1080 × 1920	29,97	+ de um tipo	1501		x	99,57%
8	clip_iPhone4_1807_1560	720 × 1280	23,976	+ de um tipo	1190		x	99,41%
9	clip_i5s_0085	1080 × 1920	29,97	+ de um tipo	200		x	100,00%
10	clip_i5s_0094	1080 × 1920	29,97	+ de um tipo	1000		x	82,39%
11	clip_i5s_0096	1080 × 1920	29,97	+ de um tipo	1100		x	98,75%
12	clip_i5s_0104	1080 × 1920	29,97	+ de um tipo	1100		x	97,96%

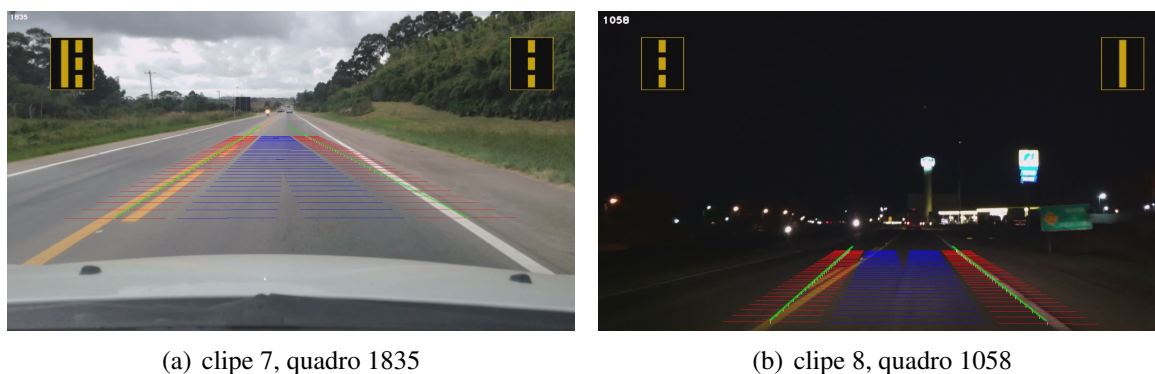
Fonte: do autor (2015).

De fato, a maioria dos erros são devidos a problemas no rastreador das bordas de pista, como ilustrado na Figura 4.18. Na Figura 4.19(a), o rastreador das bordas de pista apresentou variações/desvios da borda direita no campo de visão distante e na Figura 4.19(b), a borda esquerda da pista não está corretamente rastreada. Como consequência, os correspondentes segmentos de linha de varredura ℓ_p^e são mal posicionados, levando a falsas detecções de picos e conseqüentemente à falsa classificação. Em geral, certas situações em que o rastreador de bordas da pista apresenta limitações, tal como em dias de chuva (uma vez que o limpador de para-brisa afeta o rastreador), iluminação adversa (sol frontal que pode saturar a imagem de câmera), neblina, poeira, também tendem a corromper os resultados do esquema de classificação.

A Figura 4.20(a) mostra outros exemplos em que a abordagem proposta falha. Na Figura 4.20(a), há um veículo muito próximo a câmera, interferindo nas segmentos de linha de varredura, que comprometem o processo de extração de picos. Na Figura 4.20(b), a falta de pintura na borda direita da pista também interfere no processo de extração de picos, uma vez que parte da pintura é contínua, a distribuição global de picos é similar a uma linha seccionada.

Apesar dos poucos erros de classificação, é importante salientar que o método proposto teve foi eficaz em sequências de vídeo com baixas taxas de amostragem e resolução, tal como os clipes clip_iPhone3G_1395434215_01 e clip_iPhone3G_1395431301_70kmh, e

Figura 4.18 – Exemplos de quadros com erros de classificação devido a problemas com o rastreador de bordas de pista.



(a) clipe 7, quadro 1835

(b) clipe 8, quadro 1058

Fonte: do autor (2015).

Figura 4.19 – Exemplos das limitações enfrentadas pelo método. (a) pequena distância do veículo a frente. (b) falta da pintura da borda direita da pista.



(a) clipe 11, quadro 779

(b) clipe 11, quadro 617

Fonte: do autor (2015).

também para vídeos com alta resolução, como o clip_iPad_DMA190414_1. Também, algumas destas sequências apresentam várias transições entre os tipos de marcações de pista, o que poderiam causar atrasos na detecção. Entretanto, o método proposto teve um desempenho eficiente mesmo durante as transições, como pode ser visto na Figura 4.20 para o vídeo clip_watec_DMA311013_HMS145905.

Como mostrado na Tabela 4.10, o método também apresentou uma boa taxa de acertos para a sequência de vídeo clip_iPad_DMA190414_1, apesar da presença de rachaduras e ondulações na pavimentação durante alguns quadros, como pode ser visto na Figura 4.22(d). Também, um dos clipes (clip_iPhone4_1807_1560) foi adquirido a noite, e a acurácia obtida foi superior a 99%.

Figura 4.20 – Quadros da sequência clip_watec_DMA311013_HMS145905 sem atrasos na detecção durante as transições.



(a) Quadro 116



(b) Quadro 117



(c) Quadro 118



(d) Quadro 119

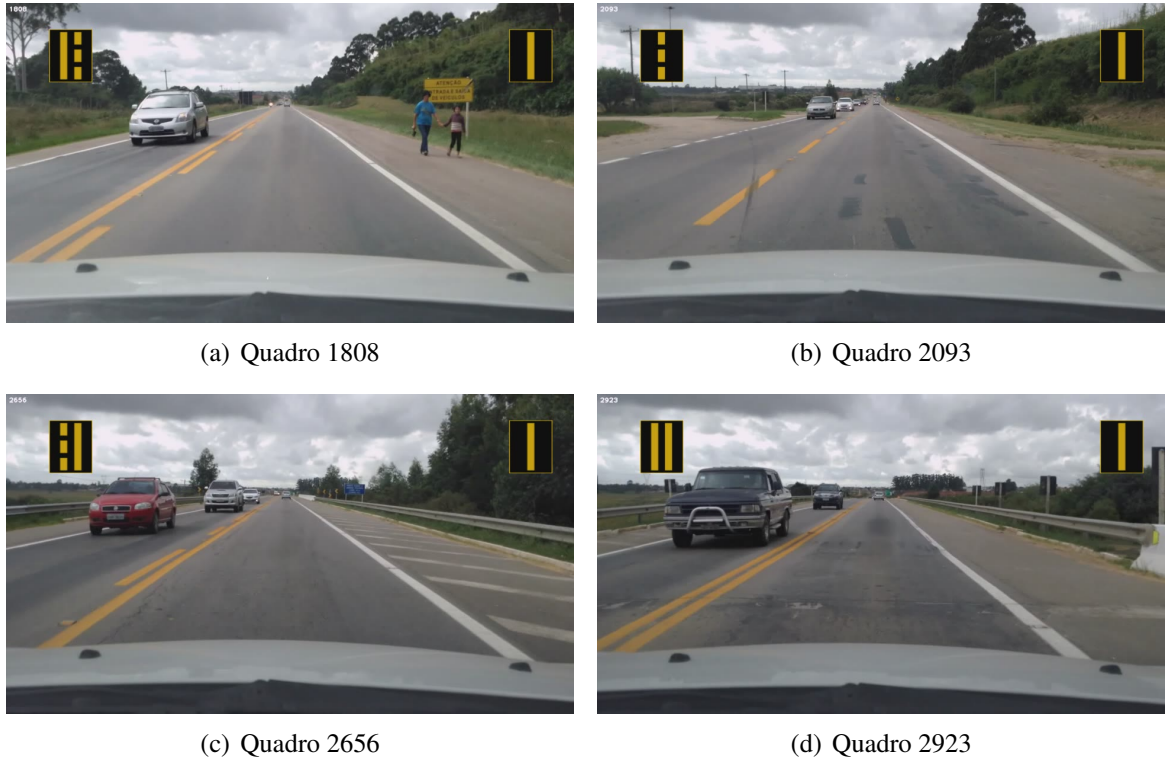
Fonte: do autor (2015).

Quanto à comparação com outros métodos, é importante salientar que até o presente momento não foram encontradas na literatura outras técnicas que classifiquem as mesmas marcações em cinco (5) tipos. Portanto, foi realizada uma comparação entre as duas técnicas propostas neste trabalho.

A Figura 4.22 mostra os resultados comparativos da análise da sinalização horizontal usando a câmera calibrada e não calibrada para o vídeo clip_iPad_DMA190414_1. Como pode ser observado, o método que realiza a análise com a câmera calibrada não apresenta atrasos de detecção face ao que não usa a câmera calibrada, o qual apresenta respostas isoladas.

Ainda, também foi implementado a técnica descrita em (SUCHITRA; SATZODA; SRIKANTHAN, 2013), que classifica as marcações de pista em contínuo e seccionado, e uma avaliação de desempenho foi realizada em um subconjunto do conjunto apresentado na Tabela 4.10

Figura 4.21 – Quadros da sequência clip_iPad_DMA190414_1.



Fonte: do autor (2015).

contendo apenas trechos com marcações contínuo e seccionado (um total de 410 quadros). A acurácia alcançada por (SUCHITRA; SATZODA; SRIKANTHAN, 2013) foi 96,43%, enquanto a técnica aqui apresenta (submetida a mesmo subconjunto) foi de 100%.

Tabela 4.11 – Matriz de confusão para os cliques 3-12

		Classe predita					Acurácia
		ω_1	ω_2	ω_3	ω_4	ω_5	
Classe real	ω_1	2312	0	1	2	0	99,87%
	ω_2	2	1225	43	9	0	95,78%
	ω_3	193	2	1747	60	4	87,09%
	ω_4	130	0	73	6555	0	97,00%
	ω_5	0	3	0	0	1995	99,85%
						99,36%	

Fonte: do autor (2015).

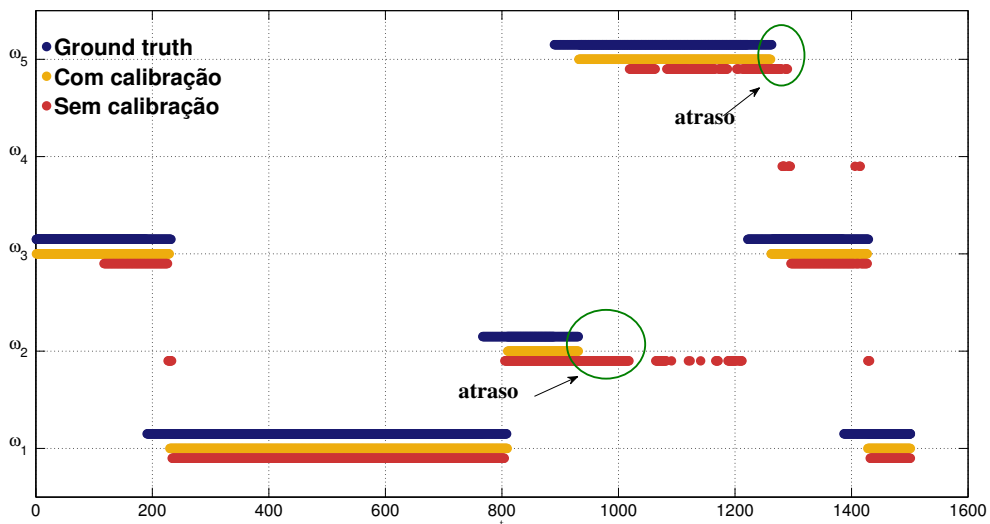
O método proposto também foi implementado em C++ (compilador GCC 4.8.2), usando o OpenCV Versão 2.4.9 e a API **Open Multi-Processing** (OpenMP⁹). Todos os experimentos foram conduzidos em um computador *desktop* Intel Core i7-2600 CPU com 8GB de RAM, e a Tabela 4.12 apresenta o tempo médio de execução para cada quadro, considerando toda a etapa (da aquisição do quadro à classificação). Do total, 24,70% é dedicado a recuperação/decodificação do vídeo, 22,89% à detecção e rastreamento das bordas da pista, 5,60% em *overhead* e 46,81% ao processo de classificação das marcações de pista, tal que a implementação em dispositivos embarcados (como por exemplo em *smartphones*) é viável.

Tabela 4.12 – Tempo médio de execução

Resolução	Tempo médio (ms)
288 × 384	2,69
480 × 640	3,71
720 × 1280	6,53
1080 × 1920	16,59

Fonte: do autor (2015).

Figura 4.22 – Comparação das duas técnicas de detecção de marcação horizontal: sem e com calibração.



Fonte: do autor (2015).

⁹<<http://openmp.org/>>

4.3 Conclusões

Neste Capítulo, foram propostas duas abordagens para detecção e reconhecimento dos principais tipos de sinalização horizontal. Ambas as técnicas realizam a identificação automática e em tempo real do tipo de pintura da pista usando uma câmera veicular embarcada. A primeira (sem calibração) utiliza um simples modelo estatístico para representar os *pixels* relacionados ao pavimento, os quais são usados para extrair as marcações da pista. A seguir, um conjunto de atributos é gerado com base na evolução temporal das marcas detectadas, e um classificador em cascata é utilizado para reconhecer os cinco tipos de marcações (conforme apresentado na Figura 1.2). Os experimentos reportam que os resultados da classificação possuem uma acurácia alta quando são usados trechos do vídeo com o mesmo tipo de marcação em todos os quadros, mas o desempenho diminui quando longos trechos de vídeo com diversas transições estão presentes (em particular, há um atraso em torno das transições dos tipos de marcação). Mesmo assim, acredita-se que os resultados globais, em torno de 78,07%, para um problema de cinco classes é bem promissor.

O segundo método, que usa a informação de câmera calibrada, diminui o número de erros de classificação por não apresentar atrasos na detecção durante as transições de marcações horizontais, conforme comparação apresentada na Figura 4.22. Isso é possível porque já é possível determinar o tipo de marcação em um único quadro através das linhas horizontais de varredura, ao contrário da primeira abordagem, que requer informação de vários quadros. Ressalta-se que a taxa global de classificação foi de 96,36% mesmo utilizando vídeos com baixas taxas de amostragem (14,64 qps), resolução (288×384) e com diversas transições entre os tipos de marcação de pista.

É importante mencionar que os resultados da classificação são afetados por erros ao longo de todo o *pipeline*: sombras e/ou mudanças na iluminação podem afetar o rastreamento da pista (bem como a qualidade dos atributos extraídos), particularmente quando a região de treinamento e de teste apresentam diferente iluminação (veja as Figuras 4.9(b) e 4.9(c)). Cabe comentar que as marcações de pista desgastadas (em decorrência de algum fator) representam um desafio adicional.

5 DETECÇÃO DO TIPO DE MARCAÇÃO VERTICAL

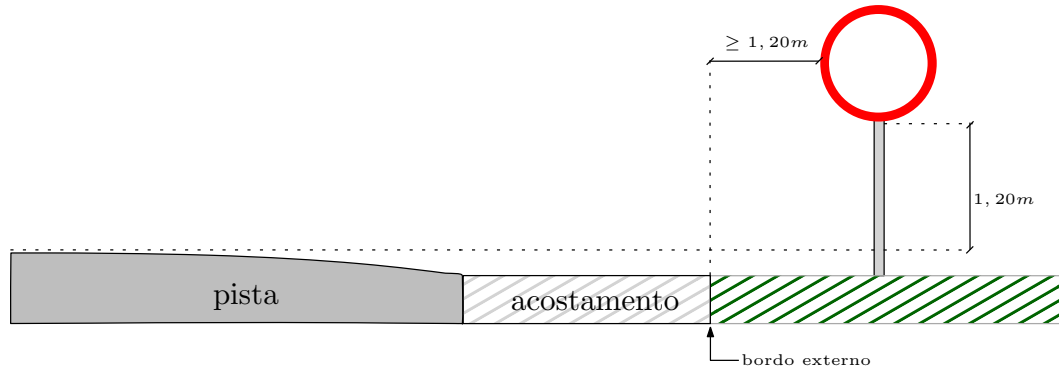
O reconhecimento de sinais verticais é um importante componente para os sistemas de assistência ao motorista. Embora várias técnicas e abordagens tenham sido estudadas, o presente trabalho visa explorar a calibração de câmera, descrita no Capítulo 3, para detectar e reconhecer a sinalização vertical de trânsito. É importante enfatizar que este trabalho não tem como objetivo principal desenvolver um novo método de detectar e reconhecer as placas de trânsito, mas sim contribuir de forma secundária usando a informação de câmera para restringir a região de busca para melhorar e acelerar o processo dos detectores existentes. Em específico, e de maneira a complementar a análise da marcação da sinalização horizontal, será realizada a detecção e identificação do sinal de proibido ultrapassar, embora a técnica aqui proposta possa ser facilmente estendida para outros sinais de trânsito verticais.

De acordo com o Manual Brasileiro de Sinalização de Trânsito (CONTRAN, 2007a), o posicionamento das placas de sinalização deve obedecer algumas regras. As placas devem ser colocadas na posição vertical, formando um ângulo de 93° a 95° em relação ao sentido do fluxo de tráfego e voltadas para o lado externo da via, ou seja, próximos da orientação ortogonal ao eixo central da via. A altura e o afastamento lateral de colocação das placas depende do tipo de via, urbana ou rural. Para este trabalho foram consideradas apenas as placas em vias rurais, que devem ser implantadas com 1,20m de altura, a contar da borda inferior da placa com relação a superfície da pista de rolamento (CONTRAN, 2007a). Ainda, o afastamento mínimo para a implantação da placa é de 1,20m do bordo externo do acostamento, ou pista.

As dimensões recomendadas para os sinais variam de acordo com a função do tipo de via. Os sinais de forma circular possuem diâmetro de 1 metro para vias rurais. Existem outros casos particulares no referido manual, como por exemplo as placas suspensas, que não serão abordadas neste trabalho. A Figura 5.1 apresenta o posicionamento da placa de acordo com as definições supracitadas.

De forma análoga à Seção 4.2, a partir da imagem de entrada com as bordas da pista detectadas e da vista aérea, calculada usando a informação da câmera calibrada, definem-se q_{max} regiões retangulares de busca r_q , para $q = 1, 2, \dots, q_{max}$, igualmente espaçadas por δ_z metros. Cada região retangular, ortogonal ao plano de chão e às bordas da pista, terá largura δ_w , altura δ_h e centro de simetria definido por:

Figura 5.1 – Posicionamento na via rural da sinalização vertical.



Fonte: CONTRAN (2007a).

$$\mathbf{W}_q = \left(p_h + \frac{p_d}{2}, IPM_y(u_{max}, v_{max}) + \delta y_{ac} + \delta y_{al} \frac{p_d}{2}, IPM_z(u_{max}, v_{max}) + (q - 1)\delta \right), \quad (5.1)$$

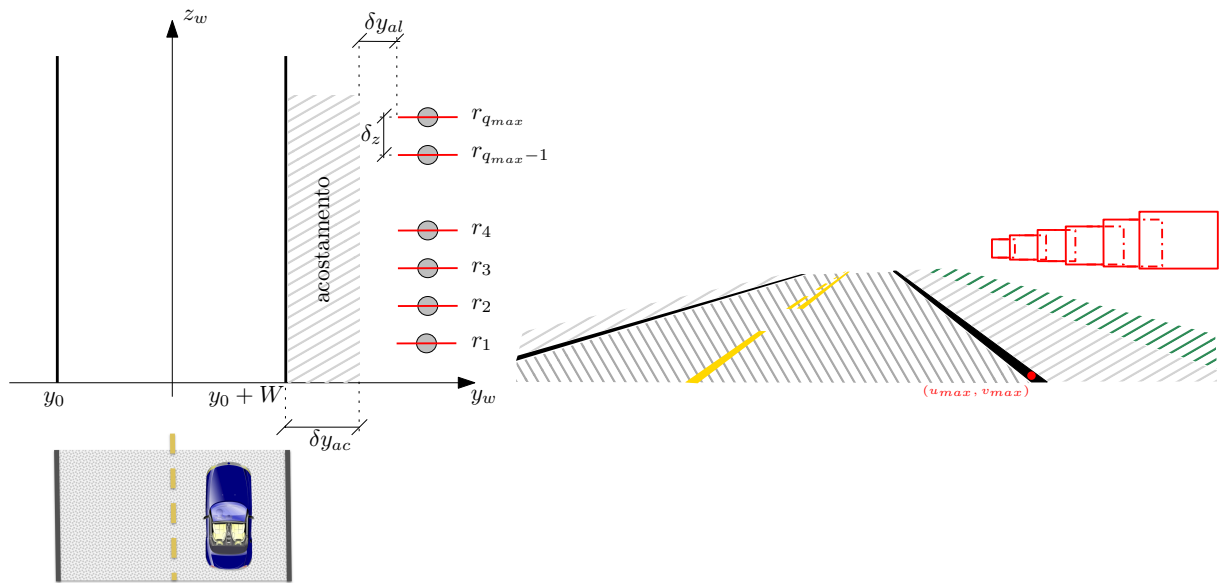
onde (u_{max}, v_{max}) é o ponto em coordenadas de imagem da borda direita da pista mais próxima da câmera, e $IPM_y(u_{max}, v_{max})$, $IPM_z(u_{max}, v_{max})$ representam as coordenadas y (lateral) e z (profundidade) da projeção inversa do ponto (u_{max}, v_{max}) de acordo com a câmera calibrada. Ainda,

- δy_{ac} é a largura esperada do acostamento.
- δy_{al} é o afastamento lateral esperado do bordo externo do acostamento a placa.
- p_d é o diâmetro esperado do sinal vertical de forma circular.
- p_h é a altura esperada da placa a partir da borda inferior.

As regiões retangulares são projetadas novamente na imagem original, como pode ser observado na Figura 5.2.

Para determinar a existência de uma placa de proibido ultrapassar para cada região r_q , foi utilizado um classificador SVM, treinado com descritores HOG. Assim como no trabalho de Dalal e Triggs (2005), a arquitetura proposta se divide em duas etapas: a primeira, chamada de aprendizado, consiste em constituir um classificador binário que verifique a existência, ou não, de uma placa de proibido ultrapassar. Já a segunda, denominada de teste, utiliza o classificador treinado para realizar uma busca em cada região da imagem de entrada, reportando a existência de placa ou não placa.

Figura 5.2 – Definição das regiões de busca retangulares em coordenadas de mundo (esquerda). Re-projeção das regiões na imagem (direita).



Fonte: do autor (2015).

A fase de aprendizado apoia-se inicialmente na construção de um conjunto de treinamento com várias amostras de imagens das placas de proibido ultrapassar, denominadas como exemplos positivos. Todas as imagens são amostradas com um tamanho fixo e contendo apenas a placa em questão. De forma semelhante às amostras negativas, ou seja, que não são placas de proibido ultrapassar, também são amostradas de forma aleatória e com o mesmo tamanho das amostras positivas. A Figura 5.3 apresenta algumas amostras que compõem o conjunto de dados de treinamento das amostras positivas e negativas. A partir desse conjunto de dados, o processo de extração de características da imagem é realizado, mapeando as amostras num espaço de atributos com tamanho fixo que de forma robusta codifica a forma visual do padrão a ser detectado.

A fase de testes consiste em uma varredura da região de busca, tipicamente usando janelas deslizantes (para detectar placas em diferentes posições) de diferentes escalas (para capturar placas com tamanhos diferentes). Com o uso da câmera calibrada, entretanto, pode-se estimar o tamanho da placa (em coordenadas de imagem) para cada região de busca r_q , de modo que apenas uma escala possa ser usada para cada região. Mais precisamente, para cada região r_q , situada a uma distância de z metros da câmera, calcula-se o diâmetro d_i esperado da placa em *pixels*. Sendo o tamanho do *template* do classificador $h_t \times w_t$ (altura \times largura),

Figura 5.3 – Amostras do conjunto de dados de treinamento do classificador SVM. (a)-(d) Amostras positivas. (e)-(h) Amostras negativas.



Fonte: do autor (2015).

define-se o redimensionamento para a região r_q de forma que d_i seja mapeado para a altura h_t , permitindo a busca usando apenas uma escala. O fator de escala da região r_q é dado por:

$$f_s = \alpha \cdot \frac{d_i}{h_t}, \quad (5.2)$$

com $\alpha < 1$, uma vez que podem haver erros nas medições.

Se o detector encontrar a placa, ela é automaticamente validada, visto que possui dimensões coerentes com a esperada. Entretanto, os detectores baseados em janelas deslizantes tendem a produzir múltiplas detecções em torno de cada objeto na imagem. Também, como há uma sobreposição das regiões r_q , poderá ocorrer mais de uma detecção por quadro de imagem. Consequentemente algumas regiões precisam ser combinadas ou excluídas utilizando algum critério.

Para resolver a possível detecção múltipla da mesma placa, foi utilizada a conhecida técnica de supressão de não-máximos (*Non-Maximum Supression*) com base no escore da detecção retornada pela SVM. Mais precisamente, uma estratégia em duas etapas foi proposta, conforme explicado a seguir. Tal estratégia é aplicada a cada quadro t da sequência de vídeo de forma independente, e por clareza de notação o índice temporal será omitido na notação que segue.

Considere $\mathcal{D}_q = \{\mathbf{D}_{q1}, \mathbf{D}_{q2}, \dots, \mathbf{D}_{qm_q}\}$ o conjunto com as m_q detecções na região de busca r_q , onde cada detecção $\mathbf{D}_{qi} = (b_{qi}, s_{qi})$ é caracterizada pela *bounding box* b_{qi} e o escore

s_{qi} , dado pela distância do vetor de atributos à fronteira de decisão na SVM. Com a hipótese de que não pode haver mais de uma placa na mesma região de busca r_q , escolhe-se como detecção vencedora $D_q = (b_q, s_q)$ na primeira etapa aquela que apresenta o maior escore, ou seja, toma-se

$$D_q = D_{qj}, \text{ onde } j = \arg \max_i s_{qi}. \quad (5.3)$$

Como as regiões r_q possuem sobreposição na imagem, conforme pode ser observado na Figura 5.4, uma mesma placa pode ser detectada em mais de uma região de busca. Entretanto, ao contrário da primeira etapa, não é razoável assumir que há apenas uma placa em cada quadro da sequência de vídeo. De fato, duas placas próximas podem ser detectadas em regiões r_q diferentes, e nesse caso é importante identificar que as detecções correspondem a placas distintas.

Figura 5.4 – Regiões de busca r_q (em amarelo).



Fonte: do autor (2015).

Para identificar se múltiplas detecções correspondem à mesma placa ou não, calcula-se a sobreposição entre as detecções. Mais precisamente, seja $\mathcal{D} = \{D_1, D_2, \dots, D_{q_{max}}\}$ o conjunto formado pelas detecções em cada região r_q de um mesmo quadro. Dada a detecção D_j com maior escore associado s_j , calcula-se o coeficiente de Jaccard para verificar se as demais detecções D_i (com $i \neq j$) apresentam superposição significativa com D_j . O coeficiente de Jaccard entre as *bounding boxes* b_i e b_j é dado por

$$J_c(i, j) = \frac{\#(b_i \cap b_j)}{\#(b_i \cup b_j)}, \quad (5.4)$$

onde $\#(b_i \cap b_j)$ e $\#(b_i \cup b_j)$ denotam, respectivamente, o número de pixels da intersecção e da união entre b_i e b_j .

De maneira semelhante a vários trabalhos que exploram janelas deslizantes (como Viola e Jones (2004)), o conjunto de detecções \mathcal{D} é particionado em dois conjuntos disjuntos. O primeiro conjunto contém as detecções que são consideradas diferentes de D_j , e contém os elementos D_i com pouca sobreposição com D_j , caracterizados por $J_c(i, j) < T_J$, onde T_J é um limiar que controla a fração de sobreposição (foi usado o valor $T_J = 0,5$). O segundo conjunto contém os elementos considerados detecções múltiplas do mesmo objeto (alta sobreposição), caracterizados por $J_c(i, j) \geq 0,5$. Nesse caso, todas as detecções do conjunto são consolidadas em uma única detecção para o objeto comum, dada pela média das coordenadas dos vértices das *bounding boxes* dos seus elementos.

5.1 Resultados

A análise da detecção e reconhecimento da sinalização vertical (placa de proibido ultrapassar) usando a informação de câmera calibrada foi aplicada em 3 sequências de vídeo, sendo uma já utilizada na Seção 4.2.5 (clip_i5s_0094) e as outras duas adquiridas com a câmera de um iPhone 5S.

É importante salientar que existem várias bases de dados para validação de placas de trânsito, tais como o LISA (*Laboratory for Intelligent and Safe Automobiles*) Dataset (MOGELMOSE; TRIVEDI; MOESLUND, 2012), que contém sinais de trânsito usados nos Estados Unidos, o GTSDB (HOUBEN et al., 2013) e o GTSRB (STALLKAMP et al., 2012) (*German Traffic Sign Detection and Recognition Benchmark*), que apresentam um conjunto de imagens com sinais verticais europeus. Como a técnica proposta é voltada para sequências de vídeo com câmera calibrada, foi necessário produzir uma nova base de dados.

Os três vídeos usados possuem resolução de 1080×1920 pixels e foram amostrados a 29.97 qps. Assim como mencionado na Seção 4.2.5, os parâmetros intrínsecos foram obtidos de forma *offline* e os parâmetros extrínsecos em tempo de execução *on-the-fly* para cada sequência de vídeo, utilizando a técnica descrita neste trabalho. A Tabela 5.1 apresenta os parâmetros de câmera para cada clipe.

O conjunto de treinamento foi formado por 159 amostras de imagens de proibido ultrapassar (amostras positivas) e 378 imagens aleatórias de não placa (amostras negativas). Todas

¹A distância focal em ambos os eixos é a mesma.

Tabela 5.1 – Clipes: parâmetros de calibração de câmera

Clipe	Nome	u_0	v_0	f_u / f_v^1	α°	β°	h (m)
1	clip_i5s_0094	959,50	539,50	1,81e+03	+5,99	-0,85	1,22
2	clip_i5s_0099	959,50	539,50	1,81e+03	+5,99	-0,85	1,22
3	clip_i5s_0789	959,50	539,50	1,81e+03	-5,54	+1,48	1,25

Fonte: do autor (2015).

as imagens são provenientes de diversas sequências de vídeo citadas nas Tabelas 3.1, 4.1, 4.8. As amostras positivas foram adquiridas de forma manual enquanto que as negativas de forma automática.

Os descritores HOG foram extraídos de cada imagem utilizando os seguintes parâmetros: célula com 8×8 *pixels*; bloco de tamanho 16×16 *pixels*; janela de 48×48 *pixels*; passo de deslocamento da janela deslizante com 8×8 *pixels*; número de intervalos para o histograma = 9.

Para realizar o processo de extração das características de cada sub-imagem através dos descritores HOG, todas as imagens foram redimensionadas para 48×48 *pixels*. Salienta-se que as imagens originais possuem resolução maior do que 48×48 *pixels*.

Como a intenção é classificar objetos em placa e não placa, foi utilizado um classificador SVM binário com *kernel* linear e margem flexível (*soft margin*) com o parâmetro $C = 0.01$, assim como Dalal e Triggs (2005). Na fase de classificação, os valores dos parâmetros que descrevem a posição e dimensão das regiões retangulares de busca, definidos no Capítulo 5, estão sumarizados na Tabela 5.2.

Tabela 5.2 – Sumário das variáveis e valores padrão

Símbolo	Definição	Valor padrão
q_{max}	número de regiões de busca r_q	10
δ_z	espaçamento entre as regiões de busca	3,0m
δ_w	largura da região de busca	2,0m
δ_h	altura da região de busca	1,5m
δy_{ac}	largura esperada do acostamento	2,0m
δy_{al}	afastamento lateral esperado do bordo externo do acostamento à placa	1,2m
p_d	diâmetro esperado do sinal vertical de forma circular	1,0m
p_h	altura esperada da placa a partir da borda inferior	1,2m

Fonte: do autor (2015).

A avaliação do sistema proposto foi realizada nas três sequências de vídeo, e os resultados podem ser observados na Tabela 5.3. A acurácia média de cada vídeo está indicada na última coluna.

Tabela 5.3 – Resolução, taxa de amostragem e acurácia global dos cliques usados

Clipe	Nome	Resolução	qps	Quadros	Acurácia
1	clip_i5s_0094	1080 × 1920	29,97	562	99,11%
2	clip_i5s_0099	1080 × 1920	29,97	100	100,00%
3	clip_i5s_0789	1080 × 1920	29,97	300	99,67%

Fonte: do autor (2015).

A taxa global de classificação para os 962 quadros analisados, oriundos das três sequências de vídeo, foi de 99,38%. Dos seis (6) erros de classificação indicados na matriz de confusão (Tabela 5.4), alguns são falsos negativos produzidos pela técnica em situações onde a placa está parcialmente fora na região de busca (veja a Figura 5.6(a)).

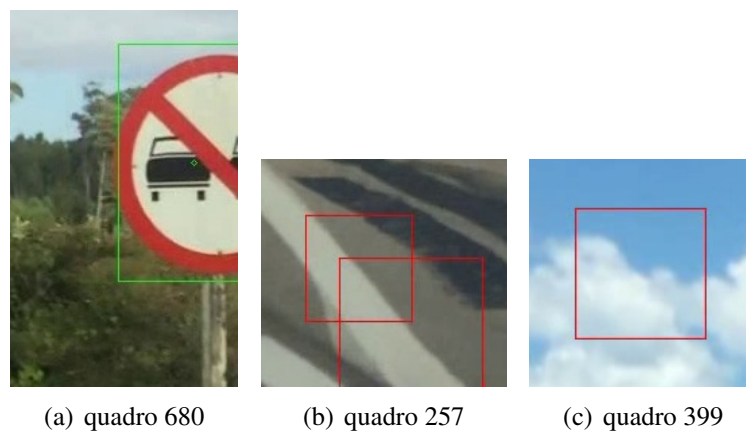
Tabela 5.4 – Matriz de confusão para os cliques 1-3

		Classe predita		
		Placa	Não placa	Acurácia
Classe real	Placa	408	6	98,55%
	Não placa	0	548	100%
				99,38%

Fonte: do autor (2015).

Para a construção da matriz de confusão, uma detecção é considerada correta se a sobreposição da região retangular da detecção e sua respectiva anotação² (*ground truth*) é maior ou igual a 50%, conforme o coeficiente de Jaccard dado pela Equação (5.4).

Figura 5.5 – (a) Falso negativo no quadro 680 do vídeo clip_i5s_0789 (nosso método). Região retangular de *ground truth* na cor verde. (b)-(c) Falso positivo no vídeo clip_i5s_0094 (HOG + SVM com varredura em toda imagem com múltiplas escalas e com fator de mudança de escala igual a 1.05).



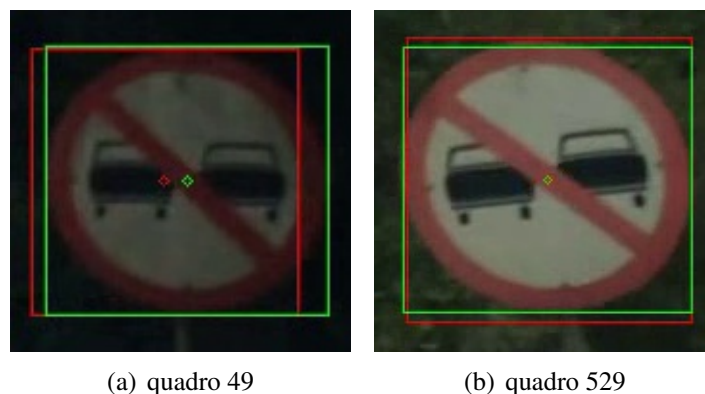
Fonte: do autor (2015).

²Os dados de *ground truth* foram anotados manualmente para cada quadro.

Os experimentos descritos nesta seção foram implementados em C++ (compilador LLVM³ versão 6.0), usando a biblioteca OpenCV versão 2.4.9. Para o classificador SVM, foi utilizada a biblioteca SVMLight (JOACHIMS, 2002). Os testes foram conduzidos em um *notebook* com processador Core 2 Duo de 2,53GHz com 8GB de RAM. O tempo médio de execução para cada quadro, considerando apenas a etapa de varredura da janela deslizante em todas as $q_{max} = 10$ regiões r_q , foi de aproximadamente 25,49 ms, que representa uma taxa de ≈ 39 qps.

Salienta-se que o tempo de execução pode ser reduzido, visto que a técnica de detecção das placas pode ser paralelizada, distribuindo a busca em cada janela r_q por *threads*. Na Figura 5.6 podem ser observadas duas detecções em quadros diferentes para o vídeo clip_i5s_0094.

Figura 5.6 – Exemplos de placas de proibido ultrapassar detectadas no vídeo clip_i5s_0094. *Ground truth* na cor verde e a detecção na cor vermelha.



Fonte: do autor (2015).

Por fim, as mesmas sequências de vídeo foram utilizadas para verificar o tempo de execução bem como a acurácia média aplicando o detector em toda a imagem, em múltiplas escalas e com quatro (4) fatores de mudança de escala.

Inicialmente, para um fator de mudança de escala igual a 1,01, a acurácia média foi de 86,90%. O tempo de execução para varrer todo quadro da imagem foi de aproximadamente 6,04s. Para um fator de mudança de escala igual a 1,05, a acurácia média foi de 90,44%, com 7 falsos negativos e 85 falsos positivos e com um tempo de execução 2,02s. Aumentando o fator de mudança de escala para 1,10 verificou-se que a acurácia média diminuiu para 87,32% em decorrência da diminuição do número de níveis da pirâmide e do aumento do número de falsos

³<<http://llvm.org>>

negativos. Conseqüentemente o tempo médio foi reduzido para 1,06s. As Figuras 5.6(b) e 5.6(c) apresentam dois quadros de imagens do vídeo clip_i5s_0094 com falsos positivos. Um último teste foi realizado alterando novamente o fator de mudança de escala para 1,20. A acurácia média diminuiu para 79,73% e o tempo médio para 0,58s. A Tabela 5.5 apresenta a acurácia média e os tempos de execução para cada quadro da imagem para o método proposto e o HOG + SVM utilizando a varredura em toda imagem, com múltiplas escalas e diferentes fatores de mudança de escala. Embora o tempo de execução no método proposto seja bem menor com relação aos demais, cabe ressaltar que o tempo médio consumido por quadro à detecção e rastreamento das bordas da pista, bem como para calcular o IPM, foi de aproximadamente 0,05s e 0,002s, respectivamente. Conseqüentemente, todo o processo consome em torno de 0,082s, que continua sendo inferior ao apresentado pelos demais métodos com um fator de mudança de escala de 1,01 a 1,10.

Tabela 5.5 – Acurácia média e tempo de execução.

Método	Acurácia	Tempo(s)
Proposto	99,38%	0,03
HOG+SVM (1,01)	86,90%	6,04
HOG+SVM (1,05)	90,44%	2,02
HOG+SVM (1,10)	87,32%	1,06
HOG+SVM (1,20)	79,73%	0,58

Fonte: do autor (2015).

Neste trabalho a técnica proposta de redução do espaço de busca para sinalização vertical foi combinada com o classificador HOG+SVM (DALAL; TRIGGS, 2005). Entretanto, salienta-se que qualquer outro método de detecção de objetos que permita a seleção da escala poderia ser usado em conjunção com a técnica proposta, visto que a principal contribuição desta tese é o uso da câmera calibrada na redução do espaço de busca e determinação da escala de busca. Embora a altura da câmera h possa apresentar erros em torno de 10 centímetros com a técnica proposta de calibração, conforme mencionado na Seção 3.7, todas as regiões de busca r_q foram estimadas em áreas da imagem que contêm placa. Para reforçar a hipótese que os erros na estimativa de h não impactam significativamente os resultados do classificador, foram realizados testes variando h (ver Tabela 5.1) com $\Delta_h = \pm 0,20\text{m}$ e os resultados apresentados na Tabela 5.3 foram mantidos.

5.2 Conclusões

A contribuição proposta no presente Capítulo foi explorar a informação de câmera calibrada para reduzir a região de busca e determinar a escala adequada para a aplicação de um detector/reconhecedor de sinais de trânsito baseado em janelas deslizantes. Como prova de conceito, foi utilizado um classificador SVM com atributos HOG. Os resultados apresentados indicam uma taxa média de classificação de 99% para os três vídeos, com apenas 962 quadros, utilizados durante a fase de testes, totalizando 962 quadros. Cabe ressaltar que o uso da câmera calibrada contribuiu significativamente para reduzir o tempo de processamento, uma vez que o espaço de varredura da janela deslizante do detector se resumiu as regiões onde provavelmente hajam placas segundo as especificações do CONTRAN (2007a). Ainda, com as informações de câmera, é possível estimar o tamanho da placa em coordenadas de imagem. Consequentemente, uma única escala pode ser usada para cada região. Os resultados comparativos mostram que, além do tempo de execução por quadro ser menor (ver Tabela 5.5) com o uso da câmera calibrada, a acurácia média é maior, visto que o espaço de busca é reduzido e consequentemente menos erros tendem a ocorrer. Embora a técnica tenha sido avaliada em um conjunto de dados relativamente pequeno, os resultados foram animadores.

6 CONCLUSÕES E CONTINUIDADE DO TRABALHO

Esta tese abordou o desenvolvimento de algoritmos de visão computacional para veículos inteligentes com potencial aplicação em sistemas de apoio ao motorista. Em específico, foram propostas três novas abordagens: (i) calibração automática de uma câmera embarcada; (ii) e (iii) detecção e classificação de marcas longitudinais horizontais e verticais, respectivamente. As técnicas apresentadas na literatura para a calibração de câmera normalmente empregam padrões de calibração de forma manual e/ou exploram a geometria da pista, necessitando de alguma informação prévia de um ou mais parâmetros como a largura da pista, por exemplo. Neste trabalho, foi apresentado um novo método de calibração automática de câmera que necessita apenas que o veículo se movimente numa velocidade constante num trecho planar com marcas longitudinais seccionadas. Duas novas abordagens para a detecção e classificação de marcas longitudinais horizontais foram propostas. A primeira (sem calibração) utiliza um simples modelo estatístico para representar os *pixels* relacionados ao pavimento, os quais são usados para extrair as marcações da pista. A segunda abordagem usa a informação de câmera calibrada contribuindo para diminuir o número de erros de classificação por não apresentar atrasos na detecção durante as transições de marcações horizontais. Por fim, uma nova técnica foi apresentada para detectar a sinalização vertical. Embora vários trabalhos tenham sido apresentados acerca do tema, a técnica proposta nesta tese busca usar a informação de câmera calibrada para explorar a relação imagem x mundo para reduzir o espaço de busca e evitar falsos-positivos que não possuem uma escala coerente. Durante a elaboração dessa tese foram geradas três publicações que podem ser observadas no Apêndice B.

Como extensões futuras deste trabalho, pretende-se realizar o uso de esquemas de estabilização de imagem com a finalidade de lidar de forma eficiente com movimentos irregulares (*jittering*) da câmera, desenvolvendo uma melhor maneira de se estimar a altura h com base em \hat{h}_s , \hat{h}_u e \hat{h}_v . Ainda, continuar investigando a necessidade de desenvolvimento de aplicações de AR em condições práticas de dirigibilidade. As duas técnicas para classificar a sinalização horizontal propostas no Capítulo 4 (câmera calibrada e não calibrada) exploram técnicas simples de consistência temporal para evitar mudanças bruscas no rótulo do classificador em quadros adjacentes. Embora os resultados atingidos já apresentem uma acurácia de classificação alta, acredita-se que o uso de técnicas mais elaboradas (como HMMs) para modelar a transição esperada dos rótulos ao longo do tempo possam melhorar ainda mais os resultados, deixando também a técnica mais robusta com relação ao ruído. Finalmente, o método que realiza a detecção das placas de trânsito usando a câmera calibrada pode ser estendido para rastrear a placa

detectada ao longo do tempo, procurando verificar o deslocamento retilíneo da mesma. Ainda, se a velocidade do veículo for conhecida, pode-se explorar o espaçamento entre detecções em quadros adjacentes para validar se as detecções são coerentes com a velocidade do veículo. O uso de outros sensores em conjunto com a câmera monocular pode contribuir para melhorar os resultados. De acordo com Levinson (2011), o uso de um laser (LIDAR) pode ajudar a mapear o ambiente, de forma a identificar as marcações de pista, desníveis dentre elementos fundamentais. O presente trabalho ainda pode ser estendido a algumas aplicações, tais como: (i) uso em empresas de locação de veículos e seguradoras para verificar a conduta do motorista no que se refere ao cumprimento da legislação de trânsito; (ii) inspeção de rodovias buscando verificar a coerência entre a sinalização horizontal e vertical, tendo como base a extração de características da via, visando verificar a adequação da sinalização horizontal de trânsito (em função das condições físicas tais como a geometria e topologia) e sua conformidade e complementação com a sinalização vertical. Cita-se, por exemplo, a existência de uma marcação contínua-seccionada que regulamenta o trecho em que a ultrapassagem, a transposição e deslocamento lateral são proibidos num sentido e permitido em outro. Consequentemente, um sinal vertical de proibido ultrapassar deve estar disposto na via, com o intuito de reforçar a sinalização horizontal de proibição de ultrapassagem. Esta aplicação servirá de subsídio para a tomada de decisão e gestão dos órgãos competentes quanto a manutenção e conservação da malha rodoviária.

REFERÊNCIAS

- ANTT. **Transporte de Passageiros**. 2010. Disponível em: <<http://www.antt.gov.br/passageiro/apresentacaopas.asp>>.
- AZUMA, R. A survey of augmented reality. **Presence Teleoperators and Virtual Environments**, v. 6, n. 4, p. 355–385, Aug 1997.
- BEDNAR, J.; WATT, T. Alpha-trimmed means and their relationship to median filters. **Acoustics, Speech and Signal Processing, IEEE Transactions on**, v. 32, n. 1, p. 145–153, 1984.
- BELLINO, M.; MERENDA, T.; KOLSKI, S. Calibration of an embedded camera for driver-assistant systems. In: **Proceedings of IEEE International Conference on Intelligent Transportation Systems**. [S.l.: s.n.], 2005. p. 354–359.
- BERTOZZI, M. et al. Artificial vision in road vehicles. **Proceedings of the IEEE**, v. 90, n. 7, p. 1258–1271, July 2002.
- BORKAR, A.; HAYES, M.; SMITH, M. T. Robust lane detection and tracking with Ransac and Kalman filter. In: **Proceedings - International Conference on Image Processing, ICIP**. [S.l.: s.n.], 2009. p. 3261–3264. ISBN 9781424456543. ISSN 15224880.
- BOUGUET, J.-y. Pyramidal implementation of the Lucas Kanade feature tracker. In: **Intel Corporation, Microprocessor Research Labs**. [S.l.: s.n.], 2000.
- BOUGUET, J. Y. **Camera calibration toolbox for Matlab**. 2008. Disponível em: <http://www.vision.caltech.edu/bouguetj/calib_doc/>.
- BRADSKI, G. The OpenCV Library. **Dr. Dobb's Journal of Software Tools**, 2000.
- BUCH, N.; VELASTIN, S.; ORWELL, J. A review of computer vision techniques for the analysis of urban traffic. **IEEE Transactions on Intelligent Transportation Systems**, v. 12, n. 3, p. 920–939, 2011. ISSN 1524-9050.
- CANNY, J. A computational approach to edge detection. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, PAMI-8, n. 6, p. 679–698, 1986. ISSN 0162-8828.
- CHANG, C.-C.; LIN, C.-J. Libsvm: A library for support vector machines. **ACM Trans. Intell. Syst. Technol.**, ACM, New York, NY, USA, v. 2, n. 3, p. 27:1–27:27, maio 2011. ISSN 2157-6904.
- CHIANG, H.-H. et al. Road speed sign recognition using edge-voting principle and learning vector quantization network. In: **Computer Symposium (ICS), 2010 International**. [S.l.: s.n.], 2010. p. 246–251.
- CHIRA, I.; CHIBULCUTEAN, A.; DANESCU, R. Real-time detection of road markings for driving assistance applications. In: **Computer Engineering and Systems (ICCES), 2010 International Conference on**. [S.l.: s.n.], 2010. p. 158–163.
- CNM. **Mapeamento das Mortes por Acidentes de Trânsito Mapeamento das mortes por acidentes de trânsito no Brasil**. [S.l.], 2009. Disponível em: <<http://portal.cnm.org.br/sites/9000/9070/Estudos/Transito/EstudoTransito-versaoconcurso.pdf>>.

- CNT, C. N. de T. **Manual Brasileiro de Sinalização de Trânsito**. 1^a. ed. [S.l.]: DENATRAN, Ministério das Cidades, 2007.
- COLLADO, J. et al. Detection and classification of road lanes with a frequency analysis. In: **Intelligent Vehicles Symposium, 2005. Proceedings. IEEE**. [S.l.: s.n.], 2005. p. 78–83.
- COLLADO, J. M. et al. Adaptive road lanes detection and classification. In: **Proceedings of the 8th international conference on Advanced Concepts For Intelligent Vision Systems**. Berlin, Heidelberg: Springer-Verlag, 2006. (ACIVS'06), p. 1151–1162. ISBN 3-540-44630-3, 978-3-540-44630-9.
- CONTRAN. **Manual Brasileiro de Sinalização de Trânsito**. 2^a. ed. Brasília, 2007a. I.
- CONTRAN. **Manual Brasileiro de Sinalização de Trânsito**. 1^a. ed. [S.l.], 2007b. IV.
- CREUSEN, I. et al. Color exploitation in hog-based traffic sign detection. In: **Image Processing (ICIP), 2010 17th IEEE International Conference on**. [S.l.: s.n.], 2010. p. 2669–2672. ISSN 1522-4880.
- 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 vol. 1. ISSN 1063-6919.
- DANESCU, R.; NEDEVSCI, S. Detection and classification of painted road objects for intersection assistance applications. In: **Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on**. [S.l.: s.n.], 2010. p. 433–438. ISSN 2153-0009.
- DEGUCHI, D. et al. Intelligent traffic sign detector: Adaptive learning based on online gathering of training samples. In: **Intelligent Vehicles Symposium (IV), 2011 IEEE**. [S.l.: s.n.], 2011. p. 72–77. ISSN 1931-0587.
- DNIT. **Anuário Estatístico de Acidentes de Trânsito - 2010**. 2010. Disponível em: <<http://www.dnit.gov.br/rodovias/operacoes-rodoviaras/estatisticas-de-acidentes>>.
- DOLLAR, P. et al. Fast feature pyramids for object detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 36, n. 8, p. 1532–1545, 2014. ISSN 01628828.
- DOLLAR, P. et al. Integral channel features. In: **Proceedings of the British Machine Vision Conference**. [S.l.]: BMVA Press, 2009. p. 91.1–91.11. ISBN 1-901725-39-1. Doi:10.5244/C.23.91.
- DOUXCHAMPS, D.; CHIHARA, K. High-accuracy and robust localization of large control markers for geometric camera calibration. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 31, p. 376–383, 2009.
- EICHNER, M.; BRECKON, T. Integrated speed limit detection and recognition from real-time video. In: **Intelligent Vehicles Symposium, 2008 IEEE**. [S.l.: s.n.], 2008. p. 626–631. ISSN 1931-0587.
- 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.

FISTREK, T.; LONCARIC, S. Traffic sign detection and recognition using neural networks and histogram based selection of segmentation method. In: **ELMAR, 2011 Proceedings**. [S.l.: s.n.], 2011. p. 51–54. ISSN 1334-2630.

GAO, X. W. et al. Recognition of traffic signs based on their colour and shape features extracted using human vision models. **J. Visual Communication and Image Representation**, v. 17, n. 4, p. 675–685, 2006.

GEHRIG, S. et al. Lane recognition on poorly structured roads-the bots dot problem in california. In: **Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on**. [S.l.: s.n.], 2002. p. 67–71.

GERÓNIMO, D. et al. Survey of pedestrian detection for advanced driver assistance systems. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 32, n. 7, p. 1239–1258, 2010.

GOMEZ-MORENO, H. et al. Goal evaluation of segmentation algorithms for traffic sign recognition. **Intelligent Transportation Systems, IEEE Transactions on**, v. 11, n. 4, p. 917–930, 2010. ISSN 1524-9050.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing (3rd Edition)**. 3. ed. [S.l.]: Prentice Hall, 2007. Hardcover. ISBN 013168728X.

GUIDUCCI, A. Parametric model of the perspective projection of a road with applications to lane keeping and 3d road reconstruction. **Computer Vision and Image Understanding**, v. 73, n. 3, p. 414 – 427, 1999. ISSN 1077-3142.

GUIDUCCI, A. Camera calibration for road applications. **Computer Vision and Image Understanding**, v. 79, n. 2, p. 250–266, August 2000.

HARRIS, C.; STEPHENS, M. A combined corner and edge detector. In: **Proc. Fourth Alvey Vision Conference**. [S.l.: s.n.], 1988. p. 147–151.

HARTLEY, R. I.; ZISSERMAN, A. **Multiple View Geometry in Computer Vision**. Second. [S.l.]: Cambridge University Press, ISBN: 0521540518, 2004.

HILLEL, A. B. et al. Recent progress in road and lane detection: a survey. **Machine vision and applications**, Springer, v. 25, n. 3, p. 727–745, 2014.

HOLD, S. et al. Efficient and robust extrinsic camera calibration procedure for lane departure warning. In: **Intelligent Vehicles Symposium, 2009 IEEE**. [S.l.: s.n.], 2009. p. 382 –387. ISSN 1931-0587.

HOUBEN, S. A single target voting scheme for traffic sign detection. In: **Intelligent Vehicles Symposium (IV), 2011 IEEE**. [S.l.: s.n.], 2011. p. 124–129. ISSN 1931-0587.

HOUBEN, S. et al. Detection of traffic signs in real-world images: The German traffic sign detection benchmark. **Proceedings of the International Joint Conference on Neural Networks**, 2013.

IRTAD. **Road Safety Annual Report 2011**. [S.l.], 2011. Disponível em: <<http://www.internationaltransportforum.org/irtadpublic/pdf/11IrtadReport.pdf>>.

JOACHIMS, T. **SVM light**, <http://svmlight.joachims.org>. [S.l.]: An implementation of Support Vector Machines (SVMs) in C. <http://svmlight.joachims.org>, 2002.

JUNG, C. R.; KELBER, C. R. An improved linear-parabolic model for lane following and curve detection. In: **Proceedings of the XVIII Brazilian Symposium on Computer Graphics and Image Processing**. Washington, DC, USA: IEEE Computer Society, 2005. (SIBGRAPI '05), p. 131–. ISBN 0-7695-2389-7.

JUNG, C. R.; SALDANHA, J. S.; JR., L. G. da S. Augmented reality with automatic camera calibration for driver assistance systems. In: **In: Proceedings of the XI Symposium on Virtual and Augmented Reality**. [S.l.: s.n.], 2009. p. 29–36.

KANHERE, N. K.; BIRCHFIELD, S. T. A taxonomy and analysis of camera calibration methods for traffic monitoring applications. **Trans. Intell. Transport. Sys.**, v. 11, n. 2, p. 441–452, jun. 2010. ISSN 1524-9050.

KOSCHER, K. et al. Experimental security analysis of a modern automobile. In: **Proceedings of IEEE Symposium on Security and Privacy**. [S.l.: s.n.], 2010. p. 447–462.

KUO, W.-J.; LIN, C.-C. Two-stage road sign detection and recognition. In: **Multimedia and Expo, 2007 IEEE International Conference on**. [S.l.: s.n.], 2007. p. 1427–1430.

LEVINSON, J. S. **Automatic laser calibration, mapping, and localization for autonomous vehicles**. Tese (Doutorado) — Stanford, August 2011.

LI, Z. et al. Road markings extraction based on threshold segmentation. In: **Fuzzy Systems and Knowledge Discovery (FSKD), 2012 9th International Conference on**. [S.l.: s.n.], 2012. p. 1924–1928.

LINDNER, P. et al. Applying multi level processing for robust geometric lane feature extraction. In: **Multisensor Fusion and Integration for Intelligent Systems (MFI), 2010 IEEE Conference on**. [S.l.: s.n.], 2010. p. 248–254.

LOPEZ, A. et al. Detection of lane markings based on ridgeness and ransac. In: **Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE**. [S.l.: s.n.], 2005. p. 254–259.

LUCAS, B. D.; KANADE, T. An iterative image registration technique with an application to stereo vision. In: . [S.l.: s.n.], 1981. p. 674–679.

MCCALL, J.; TRIVEDI, M. Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation. **Intelligent Transportation Systems, IEEE Transactions on**, v. 7, n. 1, p. 20 – 37, march 2006. ISSN 1524-9050.

MCCALL, J. et al. Lane change intent analysis using robust operators and sparse bayesian learning. In: **Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on**. [S.l.: s.n.], 2005. p. 59–59. ISSN 1063-6919.

MOGELMOSE, A.; LIU, D.; TRIVEDI, M. Detection of u.s. traffic signs. **Intelligent Transportation Systems, IEEE Transactions on**, PP, n. 99, p. 1–10, 2015. ISSN 1524-9050.

MOGELMOSE, A.; TRIVEDI, M.; MOESLUND, T. Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey. **Intelligent Transportation Systems, IEEE Transactions on**, v. 13, n. 4, p. 1484–1497, 2012. ISSN 1524-9050.

MUAD, A. et al. Implementation of inverse perspective mapping algorithm for the development of an automatic lane tracking system. **2004 IEEE Region 10 Conference TENCON 2004.**, A, 2004.

NEDEVSKI, S. et al. Accurate ego-vehicle global localization at intersections through alignment of visual data with digital map. **Intelligent Transportation Systems, IEEE Transactions on**, v. 14, n. 2, p. 673–687, June 2013. ISSN 1524-9050.

NGUWI, Y.-Y.; KOUZANI, A. Z. Detection and classification of road signs in natural environments. **Neural Computing and Applications**, v. 17, n. 3, p. 265–289, 2008.

PAULA, M. B. D.; JUNG, C. R.; SILVEIRA JR., L. G. D. Automatic on-the-fly extrinsic camera calibration of onboard vehicular cameras. **Expert Syst. Appl.**, Pergamon Press, Inc., Tarrytown, NY, USA, v. 41, n. 4, p. 1997–2007, mar. 2014. ISSN 0957-4174.

PAULA, M. Braga de; JUNG, C. R. Real-time detection and classification of road lane markings. In: **Graphics, Patterns and Images (SIBGRAPI), 2013 26th SIBGRAPI - Conference on**. [S.l.: s.n.], 2013. p. 83–90. ISSN 1530-1834.

PAULA, M. de; JUNG, C. Automatic detection and classification of road lane markings using onboard vehicular cameras. **Intelligent Transportation Systems, IEEE Transactions on**, PP, n. 99, p. 1–10, 2015. ISSN 1524-9050.

PRISACARIU, V. A. et al. Integrating object detection with 3d tracking towards a better driver assistance system. In: **ICPR**. [S.l.: s.n.], 2010. p. 3344–3347.

QINGSONG, X.; JUAN, S.; TIAN, L. A detection and recognition method for prohibition traffic signs. In: **International Conference on Image Analysis and Signal Processing**. [S.l.: s.n.], 2010.

REN, F. et al. General traffic sign recognition by feature matching. In: **Image and Vision Computing New Zealand**. [S.l.: s.n.], 2009. p. 409–414.

RIBEIRO, A. A. G. A.; DIHL, R.; JUNG, C. R. Automatic camera calibration for driver assistance systems. In: **in: Proceedings of the 13th International in: Proceedings of the 13th International Conference on Systems, Signals and Image Processing (IWSSIP)**. [S.l.: s.n.], 2006. p. 173 –176.

RUTA, A.; LI, Y.; LIU, X. Real-time traffic sign recognition from video by class-specific discriminative features. **Pattern Recognition**, Elsevier Science Inc., New York, NY, USA, v. 43, n. 1, p. 416 – 430, jan. 2010. ISSN 0031-3203.

SATZODA, R.; TRIVEDI, M. Vision-based lane analysis: Exploration of issues and approaches for embedded realization. In: **Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on**. [S.l.: s.n.], 2013. p. 604–609.

- SCHUBERT, R.; SCHULZE, K.; WANIELIK, G. Situation assessment for automatic lane-change maneuvers. **Intelligent Transportation Systems, IEEE Transactions on**, v. 11, n. 3, p. 607–616, Sept 2010. ISSN 1524-9050.
- STALLKAMP, J. et al. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. **Neural Networks**, n. 0, p. –, 2012. ISSN 0893-6080.
- SUCHITRA, S.; SATZODA, R.; SRIKANTHAN, T. Identifying lane types: A modular approach. In: **Intelligent Transportation Systems - (ITSC), 2013 16th International IEEE Conference on**. [S.l.: s.n.], 2013. p. 1929–1934.
- SUS, M. da Saúde / DATASUS Departamento de Informática do. **DATASUS**. 2015. Disponível em: <<http://www2.datasus.gov.br>>.
- TAN, J. et al. An interactive method for extrinsic parameter calibration of onboard camera. In: **Intelligent Vehicles Symposium**. [S.l.: s.n.], 2011. p. 236–241.
- TIMOFTE, R.; ZIMMERMANN, K.; GOOL, L. V. Multi-view traffic sign detection, recognition, and 3d localisation. In: **Applications of Computer Vision (WACV), 2009 Workshop on**. [S.l.: s.n.], 2009. p. 1–8. ISSN 1550-5790.
- TONNIS, M. et al. Experimental evaluation of an augmented reality visualization for directing a car driver's attention. In: **Proceedings of the International Symposium on Mixed and Augmented Reality**. Washington, DC, USA: IEEE Computer Society, 2005. p. 56–59.
- TSAI, R. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. **IEEE Journal of Robotics and Automation**, v. 3, n. 4, p. 323 –344, august 1987. ISSN 0882-4967.
- VASCONCELLOS, E. A.; SIVAK, M. **Road Safety in Brazil: Challenges and Opportunities**. [S.l.], 2009.
- VIOLA, P.; JONES, M. J. Robust Real-Time Face Detection. **International Journal of Computer Vision**, v. 57, n. 2, p. 137–154, 2004. ISSN 09205691.
- VIOLA, P.; JONES, M. J.; SNOW, D. Detecting pedestrians using patterns of motion and appearance. **International Journal of Computer Vision**, v. 63, n. 2, p. 153–161, 2005. ISSN 09205691.
- VISVIKIS, C. et al. **Study on lane departure warning and lane change assistant systems**. [S.l.], 2008.
- WANG, L. L.; TSAI, W.-H. Camera calibration by vanishing lines for 3-d computer vision. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 13, n. 4, p. 370 –376, apr 1991.
- WHO. **Global status report on road safety 2013: supporting a decade of action**. [S.l.], 2013. Disponível em: <http://www.who.int/violence_injury_prevention/road_safety_status/2013/en/>.
- WITKIN, A. Scale space filtering. In: **Proc. Eight International Joint Conference on Artificial Intelligence**. [S.l.: s.n.], 1983. p. 1019–1021.

XIE, Y. et al. Unifying visual saliency with hog feature learning for traffic sign detection. In: **Intelligent Vehicles Symposium, 2009 IEEE**. [S.l.: s.n.], 2009. p. 24–29. ISSN 1931-0587.

YENIKAYA, S.; YENIKAYA, G.; DÜVEN, E. Keeping the vehicle on the road: A survey on on-road lane detection systems. **ACM Computing Surveys (CSUR)**, ACM, v. 46, n. 1, p. 2, 2013.

ZHANG, K.; SHENG, Y.; LI, J. Automatic detection of road traffic signs from natural scene images based on pixel vector and central projected shape feature. **Intelligent Transport Systems, IET**, v. 6, n. 3, p. 282–291, 2012. ISSN 1751-956X.

ZHANG, Z. et al. Practical camera calibration from moving objects for traffic scene surveillance. **IEEE Transactions on Circuits and Systems for Video Technology**, PP, n. 99, p. 1, 2012. Available online.

APÊNDICE A — APLICAÇÃO: SINALIZAÇÃO VIRTUAL HORIZONTAL E VERTICAL

A sinalização horizontal é pintada no pavimento da estrada, e seu principal objetivo é informar o motorista acerca de específicas condições existentes na pista e/ou próxima a via, tais como: curvas a seguir, velocidade corrente permitida, presença de “zona escolar” ou faixas de segurança. Em uma aplicação comercial, a combinação de um sistema de GPS, contendo informações acerca de uma estrada georreferenciada, pode ser usada para definir a localização da sinalização horizontal que deve ser colocada de maneira virtual. Neste trabalho, indicou-se de forma manual a localização dos sinais, e avaliou-se que os mesmos, de fato, estivessem inseridos coerentemente (análise visual) com a sequência de vídeo.

Todo processo é ilustrado na Figura A.1. O vídeo capturado pela câmera real é mostrado de modo ortogonal na janela (vídeo de fundo), enquanto que os objetos virtuais são transformados de acordo com os parâmetros de câmera extraídos e sobrepostos sobre o vídeo.

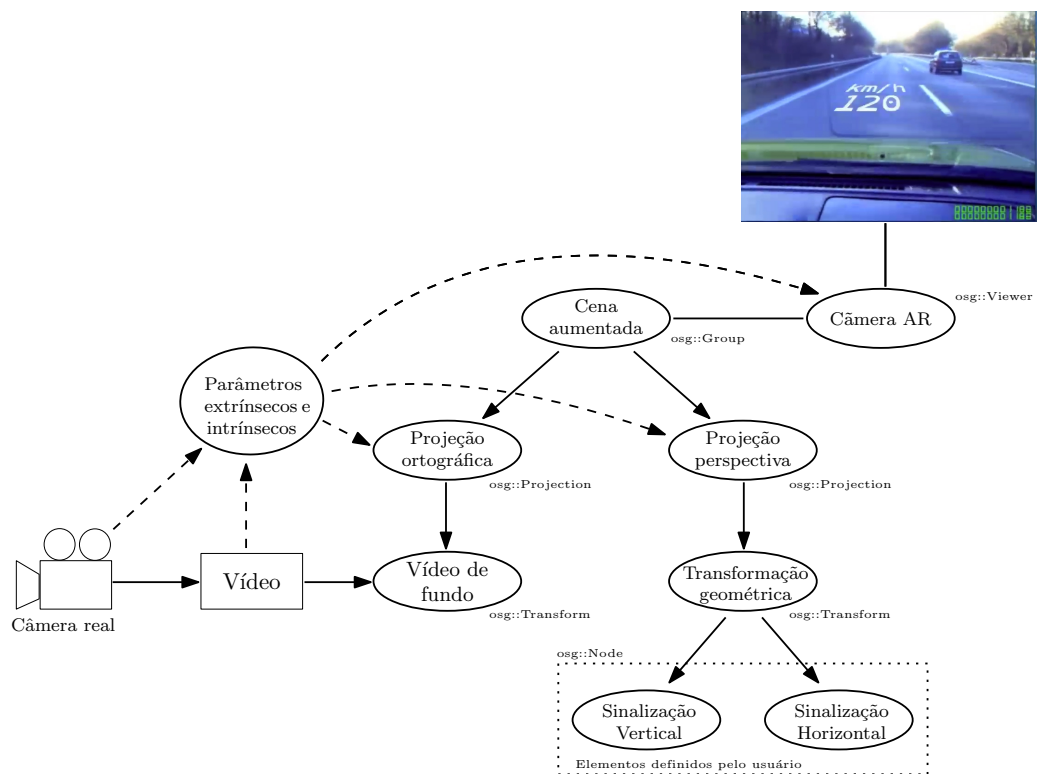


Figura A.1 – Visão geral de uma aplicação de AR. A cena real capturada por uma câmera embarcada é aumentada com os conteúdos/objetos virtuais. Os parâmetros da câmera são extraídos e usados para configurar a câmera virtual.

Um exemplo de sinalização horizontal é a indicação de proximidade escolar. A Figura A.2 apresenta alguns quadros que indicam a proximidade de uma região escolar, sobreposto a sequência de vídeo *Ipiranga*. O objeto virtual é colocado no plano da estrada (isto é, $x = 0$ em coordenadas de mundo) no meio da pista, e se aproxima do veículo a uma velocidade v (que é exatamente a mesma velocidade do veículo). Como pode ser observado, o objeto virtual inserido é coerente com o cenário real (como esperado, desde que os parâmetros de câmera estimados sejam próximos dos valores de *ground truth*, como mostrado na Tabela 3.1). A Figura A.3 apresenta outro exemplo de sinalização horizontal sobreposto no pavimento. Mais exatamente, a figura mostra alguns quadros da sequência de vídeo *Germany* e um sinal “velocidade máxima permitida”. É importante notar que, para este caso, não há valores de *ground truth* disponíveis, porém, a sinalização virtual é coerente.

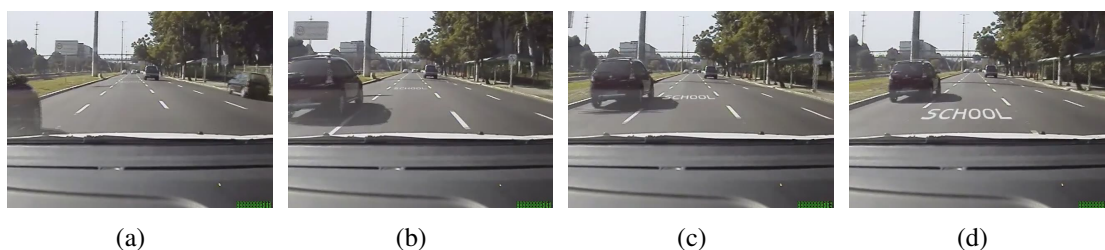


Figura A.2 – Quadros ilustrando a inserção de sinalização horizontal (inscrição no pavimento - legenda), indicando “região escolar”.



Figura A.3 – Quadros ilustrando a inserção de sinalização horizontal (inscrição no pavimento - legenda), indicando “velocidade regulamentada/permitida”.

Por outro lado, a sinalização vertical possui uma orientação ortogonal com o plano do chão, geralmente posicionada no lado esquerdo e/ou direito da via no sentido do fluxo de tráfego que devem regulamentar, transmitindo aos usuários as condições, proibições, obrigações ou restrições no uso das vias. Um exemplo de sinalização vertical virtual para a sequência de vídeo *Cambará* é mostrada Figura A.4. Na sequência de vídeo *Germany*, não há os valores de *ground truth* disponíveis para avaliar quantitativamente a qualidade dos parâmetros extrínsecos

estimados, mas a coerência visual dos objetos virtuais inseridos mostram que os parâmetros são acurados o suficiente para aplicações de AR.

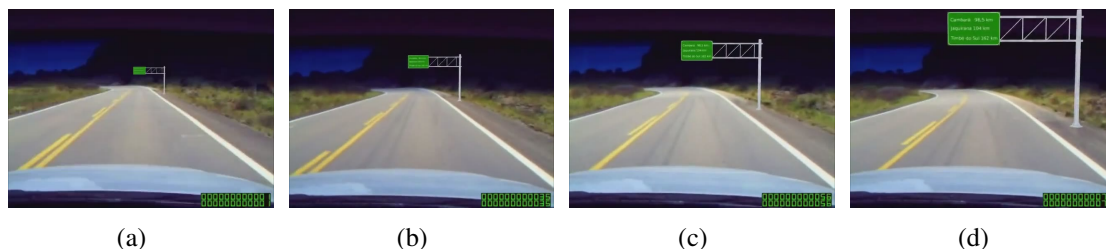


Figura A.4 – Quadros ilustrando a inserção de sinalização vertical, indicando “orientação de destino”.

Por fim, uma última aplicação é apresentada na Figura A.5. Neste exemplo, um sinal horizontal e cones sintéticos 3D são virtualmente inseridos em alguns quadros da sequência *BR116*, indicando uma advertência de “interferência no tráfego devido à existência de obras/manutenção”. Em particular, este tipo de sinalização de controle de tráfego pode ser muito útil para indicar as condições não permanentes/temporárias da estrada.

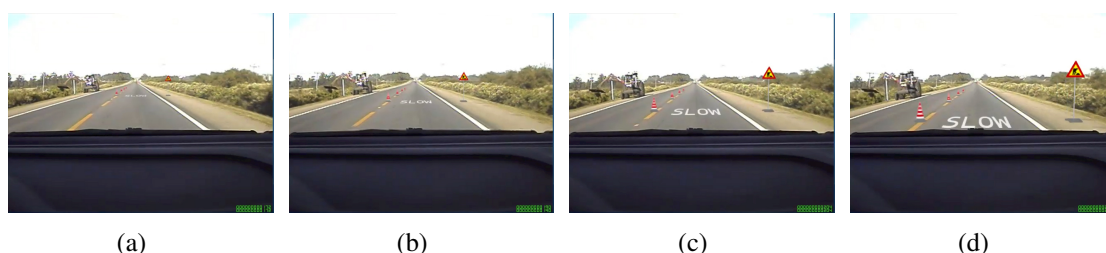


Figura A.5 – Quadros ilustrando a inserção de sinalização horizontal e vertical, indicando “interferência no tráfego devido à existência de obras/manutenção”.

Para as aplicações nesta seção, além das bibliotecas citadas na Seção 3.7, utilizou-se o *toolkit* OpenSceneGraph 3.0.1¹.

¹<<http://www.openscenegraph.org>>

APÊNDICE B — LISTA DE PUBLICAÇÕES

Aceitos/Publicados

1. Real-Time Detection and Classification of Road Lane Markings: artigo publicado no 26º SIBGRAPI - 2013 (PAULA; JUNG, 2013).
DOI: <<http://dx.doi.org/10.1109/SIBGRAPI.2013.21>>
2. Automatic On-the-fly Extrinsic Camera Calibration of Onboard Vehicular Cameras: artigo publicado na revista Expert Systems with Applications em 2014 (PAULA; JUNG; SILVEIRA JR., 2014).
DOI: <<http://dx.doi.org/10.1016/j.eswa.2013.08.096>>
3. Automatic Detection and Classification of Road Lane Markings using Onboard Vehicular Cameras: artigo publicado na revista IEEE Transactions On Intelligent Transportation Systems em 2015 (PAULA; JUNG, 2015).
DOI: <<http://dx.doi.org/10.1109/TITS.2015.2438714>>

Em preparação

1. Automatic Detection and Classification of Traffic Sign using calibrated cameras: artigo em preparação.