UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS
DEPARTAMENTO DE ENGENHARIA DE TRANSPORTES

ADRIANO BELLETTI FELICIO

Identificação automática de motociclistas através de processamento de imagens de vídeo de tráfego

São Carlos

2019

ADRIANO BELLETTI FELICIO

Identificação automática de motociclistas através de processamento de imagens de vídeo de tráfego

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Transportes da Escola de Engenharia de São Carlos da Universidade de São Paulo como parte dos requisitos para a obtenção do título de “Mestre em Ciências”.

Programa de Pós-graduação: Engenharia de Transportes.

Área de concentração: Planejamento e Operação dos Sistemas de Transportes.

Orientador:
Prof. Dr. André Luiz Barbosa Nunes da Cunha

São Carlos

2019

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da EESC/USP com os dados inseridos pelo(a) autor(a).

FF314i
i

Felicio, Adriano Belletti
Identificação automática de motocicletas através de processamento de imagens de vídeo de tráfego / Adriano Belletti Felicio; orientador André Luiz Barbosa Nunes da Cunha. São Carlos, 2019.

Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia de Transportes e Área de Concentração em Planejamento e Operação de Sistemas de Transporte -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2019.

1. Classificador de Motocicleta. 2. Máquina de Vetores de Suporte (SVM). 3. Motocicletas. I. Título.

Eduardo Graziosi Silva - CRB - 8/8907

FOLHA DE JULGAMENTO

Candidato: Engenheiro **ADRIANO BELLETTI FELICIO.**

Título da dissertação: "Identificação automática e análise do comportamento de motociclistas através de sistema de processamento de imagens de vídeo".

Data da defesa: 28/06/2019.

Comissão Julgadora:	**Resultado:**
Prof. Dr. **André Luiz Barbosa Nunes da Cunha (Orientador)** (Escola de Engenharia de São Carlos/EESC)	APROVADO
Prof. Associado **Adilson Gonzaga** (Escola de Engenharia de São Carlos/EESC)	APROVADO
Prof. Dr. **Flávio José Craveiro Cunto** (Universidade Federal do Ceará/UFC)	APROVADO

Coordenadora do Programa de Pós-Graduação em Engenharia de Transportes:
Profa. Associada **Ana Paula Camargo Larocca**

Presidente da Comissão de Pós-Graduação:
Prof. Titular **Murilo Araujo Romero**

AGRADECIMENTOS

Agradeço a Deus.

A toda minha família, principalmente aos meus pais Ronaldo e Maria Regina, por estarem sempre ao meu lado, me apoiando e incentivando em todas as ocasiões. Sou muito grato também aos meus padrinhos, Carlos e Sonia, e ao meu primo Raphael por acreditarem no meu esforço e me confortarem com palavras de incentivo todas as vezes em que precisei.

Em especial ao Aylan Kener Meneghine pelo carinho e companheirismo demonstrados em todos os momentos, por todos os ensinamentos de vida, além da paciência durante esta etapa.

Ao CNPq pela bolsa de estudos concedida.

Ao professor Dr. André Luiz Barbosa Nunes da Cunha pelo aceite na orientação e, principalmente, pela confiança depositada na realização deste trabalho. Suas sugestões, opiniões e ensinamentos foram fundamentais para que este trabalho fosse, pouco a pouco, sendo moldado.

Aos professores e funcionários do Departamento de Transportes (STT) da EESC-USP pelo apoio, convívio e amizade.

Aos amigos do grupo de pesquisa: Elaine Rodrigues Ribeiro, Gabriel Jurado Martins de Oliveira e Mariana Marçal Thebit pela amizade e ótima convivência. As conversas e sugestões foram essenciais para o andamento deste trabalho. Em especial ao Leandro Arab Marcomini e Natália Ribeiro Panice pela amizade e principalmente pela ajuda nas coletas em campo e na incrível jornada que foi a construção dos códigos em Python.

Aos grandes amigos do STT, especialmente a Bruna Kuramoto que além de fazer parte do grupo de pesquisa também compartilhou comigo do começo ao fim essa montanha russa emocional que foi o mestrado.

Aos amigos do grupo Convescote, que sempre me apoiaram e me ajudaram a treinar minha apresentação nos momentos de interesse e de arrependimento (por parte deles) ao me perguntarem “como anda a pesquisa?”. Em especial a Marina Manduca e a Bruna Bolsoni, que sempre estiveram ao meu lado, até mesmo nas madrugadas de trabalho duro e desespero quanto aos prazos.

Agradeço também ao Francisco Gásparo, pelo apoio e por me ajudar a me manter no foco nas etapas finais de incrível jornada que foi o mestrado.

LISTA DE FIGURAS

Figura 1: Disposição da organização do texto. 24

Figura 2 : Crescimento da frota brasileira. 26

Figura 3: Representação padrão de uma imagem. 28

Figura 4 : Diferença de serrilhamento entre imagens com dpi distintos. 29

Figura 5: Representação do modelo de cores RGB de 24 bits. 30

Figura 6: Exemplos de processos morfológicos. 31

Figura 7: Exemplo da aplicação do descritor LBP. 33

Figura 8 : Exemplo da aplicação do descritor HOG. 34

Figura 9:Esquema do descritor SIFT. 35

Figura 10: Representação da detecção dos pontos de interesse utilizando o descritor SURF. 35

Figura 11: Exemplos usuais de representação de padrão. 36

Figura 12: Representação esquemática do funcionamento padrão de um perceptron, concebido por McCulloch e Pits. 39

Figura 13: Representação da estimativa do hiperplano ótimo de um classificador linear a partir das margens M_1 e M_2 das classes existentes. 41

Figura 14: Representação da estimativa do hiperplano ótimo de um classificador SVM de margens suaves. 42

Figura 15:Exemplo do esquema de particionamento e execução do método k-fold com k = 3. 44

Figura 16:Exemplos de modelos de classificação no espaço ROC. 44

Figura 17: Exemplos de imagens referentes a motocicletas e similares disponíveis nos repositórios citados anteriormente. 47

Figura 18: Visão aérea do local onde ocorreu o estudo do tráfego. 51

Figura 19: Imagem da passarela Vereador Elias Damus, na direção Oeste, tirada na Av. Padre Francisco Salles Culturato sentido Leste-Oeste. 52

Figura 20: Imagem ilustrando a alteração do tamanho dos veículos ao longo do percurso dentro da ROI. 53

Figura 21: Representação da malha projetada nos vídeos contendo 24 células de 120x210 pixels 54

Figura 22: Esquema simplificado do funcionamento do classificador SVM utilizado. 55

Figura 23: Os círculos representam as regiões das imagens consideradas interessantes (cantos) pela técnica SURF. 57

Figura 24: Parâmetros adotados para a função LBP. 57

Figura 25: Exemplos dos histogramas obtidos utilizando a técnica LBP. 58

Figura 26: Parâmetros da função LinearSVC do classificador SVM 59

Figura 27: Esquema das etapas empregadas aos subconjuntos criados. 61

Figura 28: Esquema das etapas empregadas na realização da validação cruzada. 61

Figura 29: Imagens na Avenida Padre Francisco Salles Culturato sentido Oeste-Leste, no dia 18/0/2017. 64

Figura 30: Fluxograma das etapas do sistema de Identificação Automática do Comportamento do Tráfego (IACT) proposto por MARCOMINI (2018). 79

Figura 31: Determinação da ROI utilizando o sistema proposto por MARCOMINI (2018) e um dos vídeos do tráfego coletados no Capítulo 2 deste trabalho. 80

Figura 32: Criação da Janela em Perspectiva a partir da ROI (retângulo em vermelho). 80

Figura 33: Frames resultantes da etapa de remoção de fundo. 81

Figura 34: Subdivisão da ROI. (I) Detecção, (II) Rastreamento e (III) Salvamento. 82

Figura 35: Exemplo de Relatório Resumido. 83

Figura 36: Exemplo de Relatório Detalhado. 84

Figura 37:Exemplo da saída Diagrama Espaço-Tempo. 84

Figura 38: A associação do sistema de Identificação Automática do Comportamento do Tráfego – IACT (MARCOMINI, 2018) com o módulo de Classificador de Motocicletas desenvolvido - CM. 85

Figura 39: Exemplo de veículos dentro e fora da região de classificação. 86

Figura 40: Ilustração da obtenção das coordenadas da imagem a ser extraída e classificada a partir das coordenadas do centroide dos veículos. 87

Figura 41: Etapa de classificação sendo realizada na janela em perspectiva, a imagem classificada como "M" em vermelho e a imagem classificada como "NM" em azul. 88

Figura 42: Exemplo de Relatório de Predição. _____ *89*

Figura 43: Imagens salvas pelo programa na etapa de classificação. _____ *89*

Figura 44: Exemplo do Relatório com a fusão dos dados, _____ *90*

Figura 45: Local utilizado para a realização do experimento. _____ *94*

Figura 46:Esquema do método adotado para realização da validação cruzada do sistema. *95*

Figura 47: Exemplo da aplicação da associação dos módulos do sistema proposto como um todo para a amostra 1. _____ *96*

Figura 48: Exemplo do Relatório com a fusão dos dados. Em destaque a marcação "null" utilizada quando notada alguma discrepância entre os módulos para identificar e classificar algum veículo. _____ *97*

Figura 49: Arquivo com os dados fundidos e agregados utilizado como entrada para o programa SPSS. _____ *97*

Figura 50: Distribuição dos dados por classes. _____ *98*

Figura 51: Exemplo das trajetórias do primeiro vídeo coletado (11:02 até às11:14) representando 12 minutos de observação. _____ *98*

Figura 52: Exemplo das interações entre os veículos do primeiro vídeo coletado (10:49 até às11:02) representando 12 minutos de observação. _____ *101*

Figura 53: Exemplo das interações entre os veículos do terceiro vídeo coletado (11:14 até às11:27) representando 12 minutos de observação. _____ *102*

Figura 54: Exemplo das interações entre os veículos do sétimo vídeo coletado (12:05 até às12:18) representando 12 minutos de observação. _____ *103*

Figura 55: Exemplo das interações entre os veículos do nono vídeo coletado (12:31 até às12:43) representando 12 minutos de observação. _____ *104*

LISTA DE TABELAS

Tabela 1: Relatório estatístico envolvendo tamanho da frota e total de óbitos dos principais municípios das Regiões de Bauru, Campinas, Ribeirão Preto, São José do Rio Preto e da Região Central do estado de São Paulo _____ *49*

Tabela 2: Métricas convencionais usada para a avaliação do reconhecimento de objetos _ *60*

Tabela 3: Resumo dos dados obtidos na coleta piloto, categorizados por tipo de veículos e sentido das vias de acesso. _____ *63*

Tabela 4: Resumo dos dados coletados através das filmagens do tráfego, do dia 18/04/2017, na Avenida Padre Francisco Salles Culturato no sentido Oeste-Leste. _____ *63*

Tabela 5: Resultado da etapa de extração e classificação das subimagens. _____ *64*

Tabela 6: Conjunto de configurações dos parâmetros da função LinearSVC utilizados para o teste de calibração. _____ *65*

Tabela 7: Exemplo de uma das tabelas de contingência obtidas. Quantidade absoluta (número de imagens). _____ *66*

Tabela 8: Exemplo de uma das tabelas de contingência obtidas. Métricas utilizadas para avaliar o desempenho dos 20 Conjuntos de Cenários analisados. _____ *67*

Tabela 9: Média da quantidade absoluta (número de imagens) das cinco amostras para cada um dos 20 cenários analisados _____ *67*

Tabela 10: Médias das taxas de verdadeiro positivo e falso positivo das 5 amostras. _____ *68*

Tabela 11: Desempenho da métrica AUC das 5 amostras para cada uma das 20 configurações de parâmetros de calibração dos conjuntos testes e suas respectivas métricas usuais. _____ *69*

Tabela 12:Avaliação do módulo IACT. Comparação entre a contagem de veículos manual e a automática. _____ *105*

Tabela 13: Avaliação do módulo CM. Comparação entre a métrica de exatidão do classificador desenvolvido no capítulo 2 com classificador utilizado no experimento. ___ *106*

Tabela 14: Avaliação do sistema proposto como um todo (associação dos módulos IACT – CM). Tempo de processamento. _____ *106*

SUMÁRIO

AGRADECIMENTOS **7**

LISTA DE FIGURAS **9**

LISTA DE TABELAS **13**

SUMÁRIO **15**

RESUMO **17**

ABSTRACT **19**

1. INTRODUÇÃO **21**

1.1. CONTEXTUALIZAÇÃO DA PESQUISA **22**

1.2. JUSTIFICATIVA **23**

1.3. META E OBJETIVOS **23**

1.4. ESTRUTURA DO TEXTO **23**

2. DESENVOLVIMENTO DE UM CLASSIFICADOR DE MOTOCICLETAS UTILIZANDO IMAGENS **25**

2.1. CONTEXTUALIZAÇÃO **25**
2.1.1. PANORAMA GERAL DO TRÂNSITO NO BRASIL 25
2.1.2. IMAGEM DIGITAL 27
2.1.3. REPRESENTAÇÃO E DESCRIÇÃO 32
2.1.4. RECONHECIMENTO DE PADRÕES 36
2.1.5. APRENDIZADO DE MÁQUINAS 37
2.1.6. REDES NEURAIS ARTIFICIAIS 38
2.1.7. MÁQUINAS DE VETORES DE SUPORTE 40
2.1.8. VALIDAÇÃO CRUZADA 43

2.2. MATERIAIS **45**

2.3. MÉTODO PROPOSTO **47**
2.3.1. CRIAÇÃO DO BANCO DE IMAGENS DE VEÍCULOS (BIV) 47
2.3.2. DESENVOLVIMENTO DE UM CLASSIFICADOR DO TIPO SVM 54

2.4. RESULTADOS **62**
2.4.1. CRIAÇÃO DO BANCO DE IMAGENS DE VEÍCULOS (BIV) 62
2.4.2. DESENVOLVIMENTO DE UM CLASSIFICADOR DO TIPO SVM 65

2.5. CONCLUSÕES 71
2.5.1. CRIAÇÃO DO BANCO DE IMAGENS DE VEÍCULOS (BIV) 71
2.5.2. DESENVOLVIMENTO DE UM CLASSIFICADOR DO TIPO SVM 71

3. ASSOCIAÇÃO DO MÓDULO DE IDENTIFICAÇÃO AUTOMÁTICA DO COMPORTAMENTO DO TRÁFEGO COM O MÓDULO DE CLASSIFICAÇÃO DE MOTOCICLETAS 73

3.1. CONTEXTUALIZAÇÃO 73
3.1.1. SENSORES DE MONITORAMENTO DO TRÉFEGO 73
3.1.2. USO DE MOTOCICLETAS EM PERÍMETRO URBANO 74

3.2. MATERIAIS 77

3.3. MÉTODO PROPOSTO 78
3.3.1. EXTRAÇÃO DE DADOS VEICULARES 78
3.3.2. OBTENÇÃO DOS DADOS VEICULARES 82
3.3.3. CLASSIFICAÇÃO DOS VEÍCULOS 84
3.3.4. OBTENÇÃO DA CLASSIFICAÇÃO DOS VEÍCULOS 88

3.4. RESULTADOS 90
3.4.1. FUSÃO DOS DADOS OBTIDOS 90

3.5. CONCLUSÕES 91

4. ANÁLISE EXPLORATÓRIA DO COMPORTAMENTO DOS MOTOCICLISTAS UTILIZANDO VÍDEOS DO TRÁFEGO 93

4.1. ESTUDO DE CASO 93

4.2. MÉTODO PROPOSTO 94
4.2.1. SEPARAÇÃO DOS VÍDEOS EM CONJUNTOS DE TREINAMENTO E CONJUNTOS DE TESTE 94
4.2.2. TREINAMENTO DO CLASSIFICADOR DE MOTOCICLETAS 95
4.2.3. APLICAÇÃO DA ASSOCIAÇÃO DOS MÓDULOS IACT / CM AO CONJUNTO TESTE 96

4.3. DISCUSÃO E ANÁLISE DOS RESULTADOS OBTIDOS 96

5. CONSIDERAÇÕES FINAIS 107

5.1. CONCLUSÕES 107

5.2. Sugestões de Trabalhos Futuros 109

APÊNDICE A 111

REFERÊNCIAS BIBLIOGRÁFICAS 127

FELICIO, A. B. Identificação automática e Análise do comportamento de motociclistas através de sistema de processamento de imagens de vídeo. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2019.

RESUMO

Devido ao grande crescimento de motocicletas na frota urbana e o crescimento do estudo sobre seu comportamento e de como este veículo afeta o fluxo de tráfego torna-se necessário o desenvolvimento de ferramentas e técnicas diferentes das convencionais para identificar sua presença no trânsito e ser capaz de extrair suas informações. O trabalho em questão tenta contribuir para o estudo sobre este tipo de veículo, gerando um banco de imagens de motocicletas, desenvolvendo e calibrando um classificador de motocicleta e analisando o comportamento destes condutores através da utilização conjunta de um sistema de detecção automática de veículos em imagens de vídeos com o classificador desenvolvido. O classificador desenvolvido combina as técnicas de LBP para criar os vetores de características e a técnica de classificação LinearSVC para realizar as previsões. Desta forma o classificador de veículos do tipo motocicleta desenvolvido nesta pesquisa pode classificar as imagens de veículos extraídos de vídeos de monitoramento entre duas classes de motocicletas e não-motocicletas com uma precisão e uma exatidão superior a 0,9. A análise exploratória realizada nos dados obtidos da utilização conjunta de um sistema de detecção automática de veículos em imagens de vídeos com o classificador desenvolvido evidenciou uma preferência por parte das motocicletas em trafegarem na faixa da direita, mais precisamente no seu 1/3 de faixa mais próxima ao acostamento. O conjunto de dados obtidos permitiu observar algumas situações muito interessantes como as manobras de mudanças de faixa e manobras de ultrapassagem.

Palavras Chaves: Classificador de Motocicleta; Máquina de Vetores de Suporte (SVM); Motocicletas.

FELICIO, A. B. Evaluation of the behavior of motorcyclists through the video image processing system. Thesis (Master's Degree) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2019.

ABSTRACT

Due to the great movement of the camera, the performance control model and the current transmission mechanism is necessary and can be used for the selection of tools. extract your information. The evaluation exercise to evaluate the movement of a motorcycle image bank to develop and calibrate a driver analysis and evaluation system through the joint use of an automatic image monitoring system in video images with the developed classifier.

The classifier generated combinations as LBP techniques to create the characteristic and a LinearSVC classification technique to perform as predictions. In this way the media wallpaper sorter shot two classes of motorcycles and non-motorcycles with an accuracy and accuracy greater than 0.9.

An exploratory analysis performed on the joint use data of an automatic video image detection system with the developed classifier evidences a preference for motorcycling in the right lane, more precisely its range of 1/3 scale near the shoulder. The data set of a course of monitoring of some occurrences is much more interesting as maneuvers of group changes and maneuvers of overtaking.

Keywords: motorcycle classifier; Support Vector Machine (SVM); motorcycles.

1. INTRODUÇÃO

De modo geral, a obtenção e a análise de dados, além de ser essencial para o desenvolvimento de qualquer ramo da ciência, é também a primeira e principal etapa para uma tomada de decisão assertiva tanto dos cientistas, quanto administradores e operadores de qualquer tipo de sistema. No ramo da Engenharia de Transporte esse cenário não é diferente, revelando ser altamente dependente da aquisição de informações do sistema, os quais podem ser obtidos de duas formas, manual ou automática.

A forma manual para a aquisição de dados de tráfego geralmente é realizada em campo por pessoas monitorando e registrando o comportamento dos veículos para posterior análise desses comportamentos. Por sua vez, a coleta automática é realizada por três principais tipos de sensores, como: (i) sensores fixos instalados no pavimento, apresentam dificuldade na manutenção ou reposição, porém são capazes de detectar a situação de determinada região do tráfego; (ii) sensores remotos, instalados ou localizados próximos à via de tráfego, apresentam uma moderada manutenção ou reposição e são capazes de detectar a situação de várias regiões do tráfego; e (iii) sensores remotos instalados nos veículos, até o momento apresentam a maior facilidade de manutenção ou reposição, sendo capazes de monitorar quase todo o trajeto de cada veículo individualmente (KLEIN; MILLS; GIBSON, 2006).

A partir das informações coletadas, é possível analisar o comportamento dos veículos presentes na corrente de tráfego. Ribeiro (2017) realizou uma análise exploratória dos dados históricos do tráfego veicular obtidos através de sensores do tipo laço indutivo instalados no pavimento. O propósito desse trabalho foi detectar padrões e anomalias no comportamento de automóveis, caminhões e motocicletas. Os resultados mostraram que apenas para as motocicletas o método não foi capaz de identificar algum padrão de comportamento. Além disso, é importante ressaltar que apesar do laço indutivo ser comum em rodovias, os sensores instalados no pavimento são particularmente sensíveis à passagem de veículos pesados e necessitam de manutenção em um período menor do que os sensores remotos (KLEIN, 2001). Sendo assim, esses são alguns dos fatos que têm favorecido o crescimento da utilização dos sensores remotos, no qual se enquadram as técnicas de aquisição de imagens e posterior processamento.

O sistema de monitoramento por imagens de vídeo tanto no perímetro urbano como nas rodovias é exemplo de sensor remoto. Entretanto, o monitoramento em rodovias ainda depende de um observador humano, o qual assiste as imagens de vídeo com o objetivo de relatar situações de perigos e alertas, no caso mais particular do monitoramento do tráfego, identificar também anomalias ou registrar o comportamento dos veículos. Um exemplo da identificação de anomalias do tráfego de veículos é a partir

do conhecimento prévio do observador sobre as leis de trânsito e identificar algum comportamento que se diferencie do esperado, seja em relação à trajetória ou à velocidade em que os veículos demonstrem nas imagens.

A utilização do processamento de imagens no estudo da corrente de tráfego tem despertado grande interesse para os profissionais do setor de transportes, tanto no quesito da automação operacional, monitoramento e também no planejamento do tráfego veicular. Entretanto, esse método apresenta algumas restrições, tal como a dificuldade de controle de iluminação, devido principalmente às alterações das condições climáticas e à oclusão veicular, muito comum com altos volumes de tráfego, em especial de veículos pesados (KLEIN, 2001; KLEIN; MILLS; GIBSON, 2006).

As técnicas de processamento de imagens e visão computacional são ferramentas de grande valia na área de *ITS (Intelligent Transportation Systems)*, possibilitando a extração de dados do fluxo do tráfego utilizando somente câmeras de monitoramento, sendo possível a obtenção e a análise também em tempo real. Nesse aspecto, SILVA *et al*. (2013a) desenvolveram um método automático para detectar e classificar motocicletas em vias públicas no Brasil, além de identificar automaticamente se os motociclistas estão de fato utilizando o capacete, uma vez que o número de acidentes com motocicletas tem aumentado em várias partes do mundo e que a frota de motocicletas tem apresentado um crescimento significativo ao longo dos anos no Brasil. Já BARCELLOS *et al*. (2015) propuseram um sistema para detecção e contagem de veículos em vídeos de monitoramento do tráfego urbano brasileiro através de *loops* virtuais, detectando os veículos no vídeo e rastreando as suas trajetórias durante a sequência de imagens.

1.1. CONTEXTUALIZAÇÃO DA PESQUISA

Este trabalho está inserido em uma linha de pesquisa sobre Sistemas Inteligentes de Transportes (*ITS*), desenvolvida no Departamento de Engenharia de Transportes da Escola de Engenharia de São Carlos (EESC-USP) sob orientação do professor Doutor André Luiz Barbosa Nunes da Cunha. O propósito desta pesquisa é utilizar técnicas de visão computacional para o monitoramento automático de tráfego veicular em trechos viários de forma a identificar e classificar os tipos de veículos, possibilitando à análise do comportamento dos condutores, mais precisamente, dos motociclistas brasileiros. Desta forma, a pesquisa em questão buscou a integração entre as áreas de Engenharia de Tráfego e Ciência da Computação.

1.2. JUSTIFICATIVA

A principal justificativa desta pesquisa se baseia na escassez de pesquisas na literatura que abordam o comportamento dos motociclistas e como estes interferem nos parâmetros da corrente de tráfego. Outra justificativa importante é devido ao crescimento da frota de motocicletas nas cidades brasileiras, seja pelas vantagens de mobilidade ou custo de manutenção, apesar da inerente fragilidade nos quesitos de segurança deste veículo em particular.

Desta forma, esta pesquisa baseia-se na necessidade de se obter dados da corrente de tráfego de forma fácil, barata e eficiente, podendo auxiliar os estudos sobre o comportamento do tráfego veicular, como por exemplo a simulação e calibração de modelos de interação dos veículos que compõem a corrente de tráfego, entre outros. Além disso, baseia-se também, no baixo custo dos equipamentos necessários para a captura de vídeos e para o processamento das imagens, como computadores e bibliotecas com códigos abertos que auxiliam no processamento de imagens.

1.3. META E OBJETIVOS

Esta pesquisa está baseada na hipótese de que é possível identificar comportamentos de motociclistas em uma corrente de tráfego usando visão computacional. Desta forma, a meta desta pesquisa de mestrado é identificar motociclistas através de um sistema de detecção automática em imagens de vídeos possibilitando a análise do comportamento desses condutores. Sendo que o objetivo principal é propor um método de pesquisa veicular com o foco nas motocicletas que permita de forma automática a sua contagem e o rastreamento ao longo de segmentos viários urbanos. Para isto foram propostas as seguintes etapas principais:

(a) Criar um Banco de Imagens de Veículos para utilizar em um Classificador de Imagens;
(b) Desenvolver um Classificador de motocicletas a partir de imagens de veículos;
(c) Identificar motocicletas em vídeos de tráfego utilizando um Classificador de Imagens;
(d) Coletar informações de vídeos de tráfego como velocidade, aceleração e tipo de veículos (moto ou não moto) presentes no trânsito e
(e) Analisar o comportamento dos motociclistas a partir dos dados veiculares obtidos.

1.4. ESTRUTURA DO TEXTO

Este texto está organizado em cinco capítulos. O *Capítulo 1* consiste nesta introdução que apresenta de maneira resumida o contexto, os objetivos e as justificativas para a realização desta pesquisa.

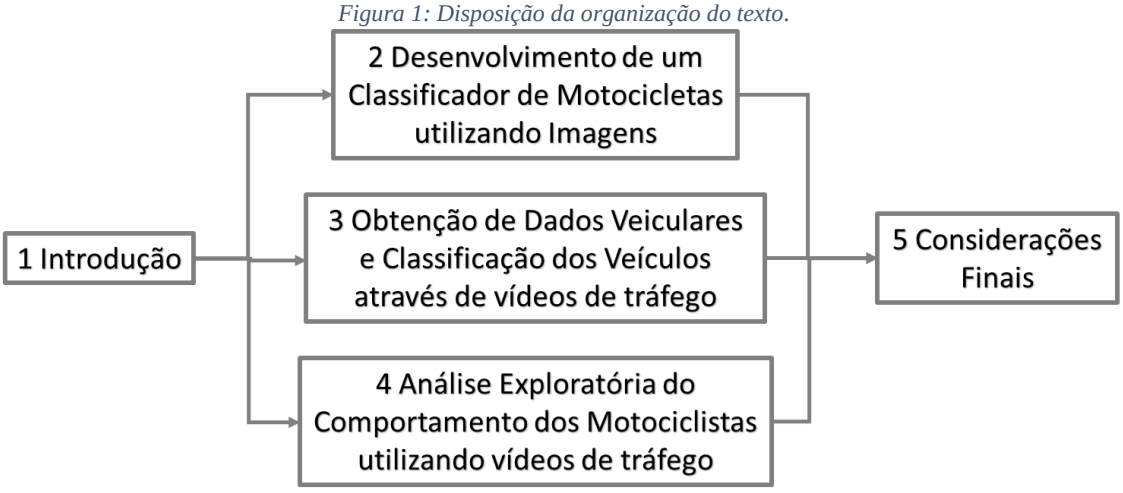
Os três capítulos seguintes constituem os experimentos e resultados da pesquisa. Optou-se por dividir a pesquisa em subpartes sendo cada uma delas um capítulo pois cada um destes capítulos pode ser entendido e aplicado de maneira individual. Desta forma cada um destes três capítulos experimentais apresenta método, resultados e conclusões.

A revisão bibliográfica realizada neste trabalho é apresentada como contextualização e se encontra no *Capítulo 2*, *Capítulo 3*, nos quais abordando alguns dos principais temas da pesquisa como o comportamento dos motociclistas e como estes interferem nos parâmetros da corrente de tráfego, tanto no âmbito mundial como no âmbito nacional, além de conceitos gerais e as técnicas de processamento de imagens, visão computacional e aprendizado de máquinas empregadas na área de engenharia de tráfego.

O *Capítulo 4* constituí na análise dos dados obtidos e por sua vez o *Capítulo 5* constituí as conclusões obtidas do trabalho e sugestões para trabalhos futuros.

Desta forma, a estrutura do texto está organizada conforme mostrado na Figura 1.

Figura 1: Disposição da organização do texto.



Fonte: O autor.

2. DESENVOLVIMENTO DE UM CLASSIFICADOR DE MOTOCICLETAS UTILIZANDO IMAGENS

Este *Capítulo 2* está organizado em três seções centrais que possuem objetivos, materiais, métodos e resultados próprios. O objetivo central deste capítulo é desenvolver um classificador de motocicletas utilizando um banco de imagens de veículos extraídos a partir de vídeos de tráfego.

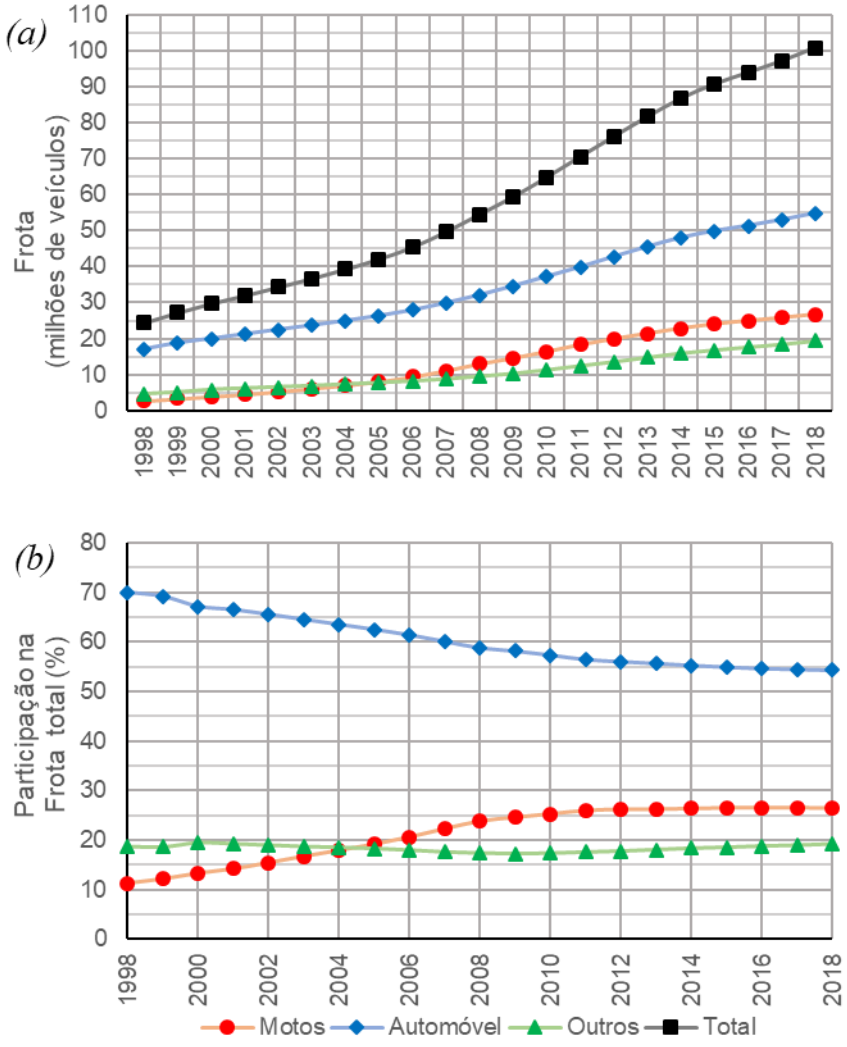
2.1. CONTEXTUALIZAÇÃO

Neste capítulo são apresentados o panorama geral da situação do trânsito brasileiro e os aspectos sobre a utilização da motocicleta como modo de transporte em áreas urbanas, além disso como esta pesquisa envolve temas da área da Ciência da Computação e de Engenharia de Tráfego. neste capítulo também serão apresentados alguns conceitos básicos da área da Ciência da Computação que podem ser desconhecidos para os interessados da área de Engenharia de Tráfego.

2.1.1. PANORAMA GERAL DO TRÂNSITO NO BRASIL

Embora a frota de veículos no Brasil tenha apresentado um crescimento nas últimas duas décadas, a parcela de automóveis vem apresentando uma redução, enquanto que a participação de motocicletas cresce gradativamente (DENATRAN, 2018). A Figura 2 apresenta o crescimento da frota brasileira e a participação dos principais veículos que a compõem nos últimos anos a partir do Registro Nacional de Veículo (RENAVAM), disponíveis no site do Departamento Nacional de Trânsito – DENATRAN.

Figura 2 : Crescimento da frota brasileira.
(a) em números absolutos, (b) frequência relativa por tipo de veículo).



Fonte: O autor sobre os dados do DENATRAN (2018).

Dessa figura, é possível notar que, apesar da redução na participação da frota, os automóveis ainda compõem mais da metade da frota total e o crescimento das motocicletas as colocam em segundo lugar em quantidade de veículos no Brasil. É possível notar também que as alterações foram proporcionais entre esses dois veículos, ou seja, a redução de 15% da frota de automóvel (de 70% em 1988 para 55% em 2018) é muito próxima do aumento de motocicletas (de 10% em 1988 para 27% em 2018) de aproximadamente 17%. Dessa forma, o conjunto, automóveis e motocicletas, representam mais de 80% da frota total.

Outro fato a ser notado é que em apenas vinte anos a frota nacional praticamente quadruplicou de tamanho, passando de aproximadamente 25,8 milhões de veículos em 1998 para aproximadamente 100,8 milhões de veículos em 2018. Esse aumento no volume de veículos trafegando proporciona um aumento na interação entre eles e consequentemente maiores serão as chances de ocorrerem acidentes de trânsito

tanto com ou sem fatalidades. E com esse pensamento e devido ao grande número de veículos registrados somente no estado de São Paulo (mais de 21 milhões de veículos representando 31% da frota nacional em 2011) foi implantado neste mesmo ano o Movimento Paulista de Segurança no Trânsito que, baseado na "Década de Ação Pela Segurança Viária", promovido pela Organização das Nações Unidas com o intuito de discutir a questão da violência no trânsito, tem como objetivo reduzir pela metade as vítimas fatais nos acidentes de trânsito até 2020 através do auxílio do sistema de Informações Gerenciais de Acidentes de Trânsito do Estado de São Paulo (INFOSIGA SP), um banco de dados que consolida as informações dos acidentes de trânsito de diversas fontes (INFOSIGA SP, 2016).

Além da composição da frota, os relatórios (DENATRAN, 2018) detalham que a região com maior frota total é a região Sudeste, sendo o estado de São Paulo com mais de 29% dos veículos registrados no mês de dezembro no ano de 2018 (últimos dados disponibilizados até a data de conclusão deste trabalho).

2.1.2. IMAGEM DIGITAL

Uma imagem normalmente é representada como uma função bidimensional $f\ (x, y)$, de forma que suas coordenadas no plano são representadas por x e y e o valor da função f para qualquer ponto no plano representa a intensidade ou nível de cinza deste ponto. É importante ressaltar que o plano usado para representar uma imagem é muito similar ao plano cartesiano usual, porém difere pela orientação e pela posição da origem de seus eixos, Figura 3a.

Como valores de x, y e de f assumem valores discretos e finitos, tem-se então uma imagem digital. Usualmente no campo do processamento de imagens, uma imagem real é representada digitalmente na forma de uma matriz, um *bitmap*, no qual cada par de coordenadas (x, y) corresponde a um elemento $f(x, y)$ da matriz. Esses elementos da matriz ou da imagem digital são comumente denominados de *pixels* (GONZALEZ; WOODS, 2010). A Figura 3b ilustra o padrão de representação de imagens digitais.

Figura 3: Representação padrão de uma imagem.
(a) Convenção dos eixos em imagens digitais. (b) Representação da matriz da imagem.

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & \dots & f(1,N-1) \\ \vdots & \vdots & \ddots & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1,N-1) \end{bmatrix}$$

(a) (b)

Fonte: O autor.

Cada pixel representa o brilho do ponto, de forma que os valores mais próximos do limite inferior representam pontos mais escuros (próximos a cor preta) e os valores mais próximos do limite superior representam pontos mais claros (próximos a cor branca), variando dentro do intervalo definido e em escala de cinza, sendo que o mais usual é um intervalo discreto composto por números interior variando de 0 até 255. Para imagens coloridas a representação ocorre pela sobreposição de três imagens diferentes que reproduzem os tons de uma única cor cada, no mesmo intervalo e na mesma escala (imagens monocromáticas).

Dessa maneira, para se representar uma imagem digital, são necessárias duas informações: (i) as coordenadas de cada pixel e (ii) as informações sobre os brilhos (escala de cinza) ou as cores (imagens coloridas) que cada pixel traz. Assim o processo de digitalizar os valores das coordenadas de cada pixel é denominado de Amostragem e o processo de digitalizar os valores de amplitude referente a cada pixel é denominado de Quantização. O conceito de resolução espacial utiliza uma medida para definir o tamanho do menor elemento a compor uma imagem (denominado de *pixel*) pois estipula o tamanho total da imagem e por consequência ajuda a definir a sua qualidade (GONZALEZ; WOODS, 2010). Quanto maior a quantidade de *pixels* utilizados para representar uma imagem, mais nítida será a imagem e menos perceptível será o efeito de serrilhamento presente nela. Dessa forma, quanto mais dpi (*dots per inch*) uma imagem for composta maior será sua resolução espacial, Figura 4, porém demandará maior espaço de armazenamento sendo importante combinar o tamanho com a qualidade da imagem.

Figura 4 : Diferença de serrilhamento entre imagens com dpi distintos.

Fonte: https://www.impressorajato.com.br/blog/qual-a-resolucao-ideal-para-a-sua-impressao.

O processo de quantização define a resolução de escala de cinza (ou de contraste) da imagem e implica diretamente na quantidade de memória para armazenamento da imagem, como os computadores processam elementos binários, logo tudo que for representado computacionalmente deve estar em potência de 2. Geralmente a quantidade de *bits* necessária para representar cada pixel determina o máximo de combinações de cores que uma imagem é capaz de possuir. Com o intuito de exemplificar essa afirmação, para uma imagem de 100×100 pixels em escala de cinza as seguintes características são possíveis:

- Com 1 *bit* por pixel de resolução de escala de cinza (ou de contraste), é possível representar $2^1 = 2$ níveis de cinza, a cor preta (valor 0) e a cor branca (valor 1) e o espaço necessário para armazenar essa imagem é de 10.000 *bits* ($100\ x\ 100\ x\ 1$).
- Com 8 *bit* por pixel de resolução, é possível representar $2^8 = 256$ níveis de cinza e o espaço necessário para armazenar essa imagem é 80.000 *bits* ($100\ x\ 100\ x\ 8$).

Porém, uma imagem colorida, no espaço de cor RGB (Red, Green, Blue) com as mesmas dimensões apresenta as seguintes características:

- Com 8 *bit* por pixel de resolução, com 3 *bits* para representar as cores vermelha e verde e 2 *bits* para representar a cor azul, é possível representar $2^8 = 256$ cores de modo geral, e o espaço necessário para armazenar essa imagem será de 80.000 *bits* ($100\ x\ 100\ x\ 8$).
- Com 24 *bit* por pixel de resolução, com 8 *bits* para representar as cores vermelha, verde e azul, é possível representar $2^{24} = 16.777.216$ cores de modo geral, e o espaço necessário para armazenar essa imagem será de 240.000 *bits* ($100\ x\ 100\ x\ 24$).

Por sua vez, a resolução de cores é uma medida que define a quantidade total de cores de uma imagem. Portanto um único pixel é capaz de armazenar várias componentes baseado no sistema de cor que segue, como caracterizado no exemplo anterior. Os modelos de cores são representados em um sistema de coordenadas no qual cada ponto dentro desse sistema representa uma cor definida, sendo que os modelos de cores mais comuns são: RGB o qual utiliza as cores primárias da luz, vermelho (*Red*), verde (*Green*) e

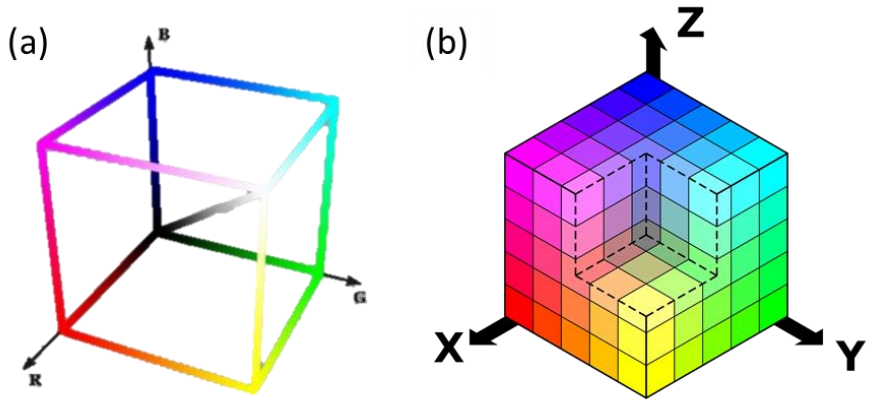
azul (*Blue*), para compor seu sistema e os modelos de cores CMY e CMYK que são baseados nas cores secundárias da luz, ciano (*Cyan*), magenta (*Magenta*), amarelo (*Yellow*) e preto (*Black*).

No caso do modelo de cores RGB cada pixel armazena três componentes de cores, de maneira que a combinação desses três componentes básicos permite a reprodução de todas as demais cores (sistema aditivo de cores).

Este modelo de cores pode ser demonstrado por um cubo, no qual as três cores primárias são representadas por três vértices, de maneira que, na aresta em comum é localizada a origem do sistema e é o local no qual é atribuída a cor preta (0,0,0). As três cores secundárias são atribuídas a os outros três vértices, de forma que para a outra aresta em comum é atribuída a cor branca (1,1,1). Desta forma a escala de cinza é representada pela diagonal principal do cubo, Figura 5 a.

As nuanças e tons das cores nesse modelo são representadas por pontos dentro do cubo ou nas faces do mesmo e podem ser interpretados como vetores que partem da origem. A Figura 5 ilustra o modelo de cores RGB.

Figura 5: Representação do modelo de cores RGB de 24 bits.
(a) Variação das cores primárias, secundárias e da escala de cinza representadas nos eixos e na diagonal principal do cubo.
(b) Reprodução do espaço de cores no cubo totalmente preenchido.



Fonte: O autor sobre imagens da Wikipedia.

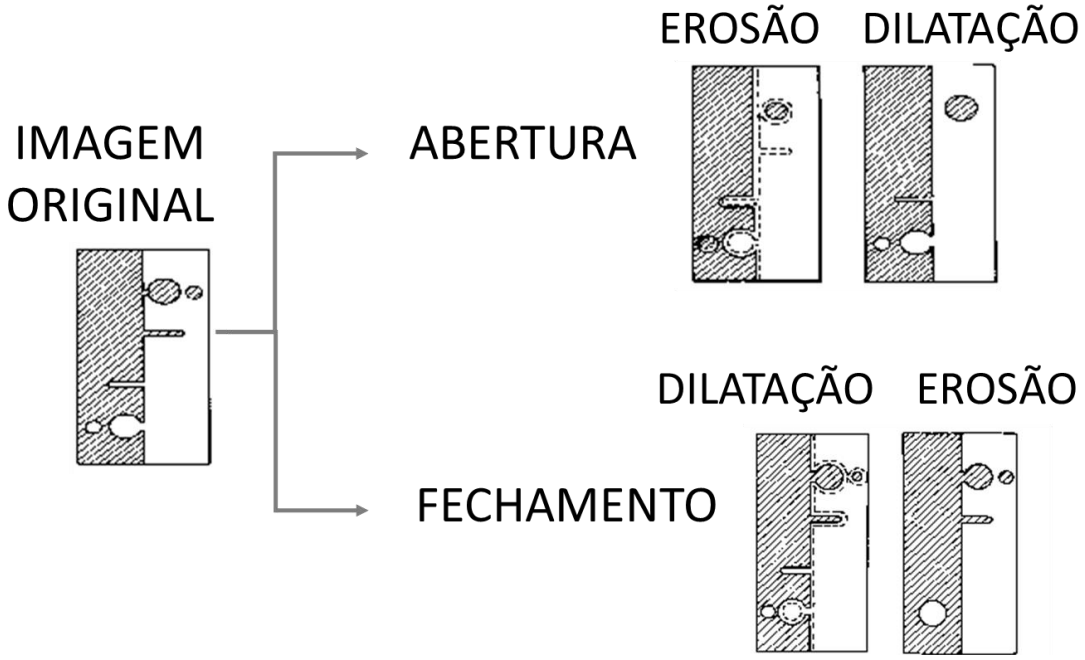
Em processamento de imagens digitais, as técnicas de segmentação possuem como objetivo subdividir uma imagem em regiões ou objetos de interesse e seu processo deve ser interrompido quando esses forem detectados, o que depende do nível de detalhamento do problema em questão (YU et al., 2003; SILVA et al., 2013b). O contexto "morfologia matemática" foi introduzido como uma ferramenta para a extração de componentes de uma região da imagem com o objetivo de facilitar a representação e descrição da forma desta região. Dessa forma, as técnicas relacionadas e os conceitos de morfologia se tornaram a pedra angular para a extração de "significado" de uma imagem.(GONZALEZ; WOODS, 2010).

De maneira geral, na segmentação cada pixel deve fazer parte de uma região, sendo que os pixels de uma mesma região devem estar conectados e possuírem propriedades similares, enquanto as regiões

diferentes devem apresentar propriedades diferentes. Esse agrupamento dos pixels pode ocorrer por: (i) detecção de descontinuidade, (ii) técnicas de limiarização, (iii) métodos baseados em região ou (iv) por métodos híbridos (GONZALEZ; WOODS, 2010). Algumas das técnicas comumente utilizadas para segmentação são: detecção de pontos isolados, detecção de linhas, modelos de bordas, transformada de Hough, crescimento de região, divisão e fusão de região, conceito de bacia hidrográfica (*watershed*), entre outras. Dentre essas, a mais usual na literatura é a transformada de Hough, que tenta detectar formas geométricas como retas , círculos e elipses de forma a parametrizá-las e que foi apresentada pela primeira fez em Duda e Hart, (1972).

No caso dos processos morfológicos, suas operações básicas são erosão e dilatação. A erosão é uma operação que usualmente reduz o tamanho das regiões da imagem, diminuindo os pontos do objeto de interesse (pontos escuros) e ressaltando os pontos do fundo da imagem (pontos claros). Enquanto a dilatação ressalta os pontos do objeto de interesse (pontos escuros) e diminui os pontos do fundo da imagem (pontos claros) Figura 6.

Figura 6: Exemplos de processos morfológicos.



Fonte: O autor sobre imagens do site http://www.dpi.inpe.br/spring/teoria/filtrage/filtragem.htm

É importante salientar que esses processos são operações duais e não inversos, de forma a permitir operações mais avançadas a partir da combinação destas como é o caso da abertura e do fechamento. A abertura (erosão seguida de dilatação) usualmente suaviza o contorno e elimina protuberâncias, suprimindo os detalhes do objetos de interesse; e o fechamento (dilatação seguida de erosão) suaviza contornos, porém usualmente une partes, eliminando buracos e preenchendo o contorno, suprimindo os detalhes do fundo.(GONZALEZ; WOODS, 2010). Dessa forma, as operações morfológicas permitem a

remoção de imperfeições nas imagens já segmentadas e a obtenção de informações sobre sua forma e estrutura.

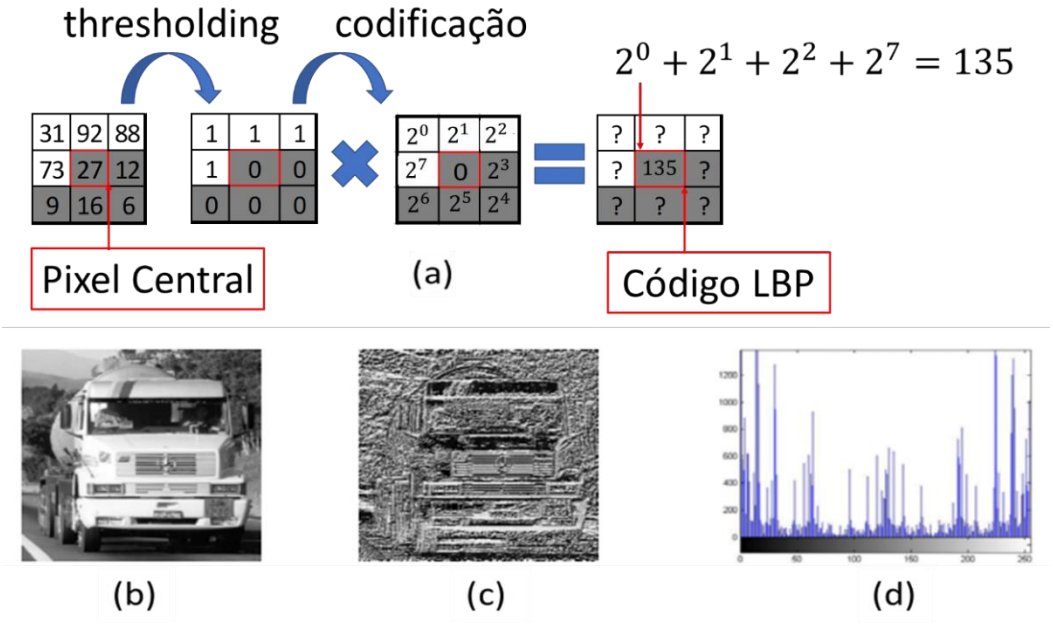
2.1.3. REPRESENTAÇÃO E DESCRIÇÃO

Após a segmentação de objetos na imagem é possível representar e descrever essas regiões de tal maneira que seja capaz de extrair algum conhecimento desses objetos. De maneira simplificada, representar uma região da imagem pode ocorrer de duas maneiras: (i) pelas características externas, analisando os pixels que compõem sua fronteira, ou (ii) pelas características internas, analisando os pixels que compõem a região em si (GONZALEZ; WOODS, 2010).

Após a escolha do modo de representação da região é necessário optar pela forma como será descrita. De forma geral, para se descrever uma região por cor ou textura, usualmente é escolhido a representação interna, e quando se deseja descrever a região por forma, normalmente é escolhido a representação externa (FORSSÉN; GRANLUND, 2003; DALAL; TRIGGS, 2005; HEIKKILA; PIETIKAINEN, 2006; GUOYING ZHAO *et al.*, 2012).

Através da revisão bibliográfica foi possível notar que alguns dos descritores mais utilizados na literatura atualmente para se trabalhar com imagens são: HOG, LBP e SURF (NIXON; AGUADO, 2002; FORSSÉN; GRANLUND, 2003; TAKALA; AHONEN; PIETIKÄINEN, 2005; DALAL; TRIGGS, 2005; BAY; TUYTELAARS; VAN GOOL, 2006; KIM; DAHYOT, 2008; GONZALEZ; WOODS, 2010; PIETIKÄINEN et al., 2011; GUOYING ZHAO et al., 2012; SILVA et al., 2013a; CUNHA, 2013; SILVA, 2014; BEYELER, 2015). O descritor LBP (*Local Binary Pattern*) é um descritor baseado em textura o qual é calculado a partir de uma varredura dos pixels da imagem analisando a variação dos brilhos entre o pixel de referência e os seus vizinhos. Esse descritor permite invariância a rotação realizando uma contagem entre a variação positiva e negativa entre os pixels analisados. Usualmente este descritor pode utilizar como critério de similaridade a distância euclidiana ou a distância *Manhattan* para reconhecer regiões semelhantes em imagens distintas (TAKALA; AHONEN; PIETIKÄINEN, 2005; PIETIKÄINEN et al., 2011; GUOYING ZHAO et al., 2012). Em sua versão original, sua vizinhança de análise era de 3×3, e os valores dos *pixels* eram comparados ao valor do pixel central produzindo um valor binário, ou seja, caso o valor do pixel da vizinhança, na escala de cinza fosse superior ao valor do pixel central, na mesma escala de cinza, o resultado seria 1, caso contrário o resultado seria 0. E para se obter o código do pixel central denominado de código LBP, cada um desses valores resultantes eram multiplicados por potência de dois e somados, como é ilustrado pela Figura 7.

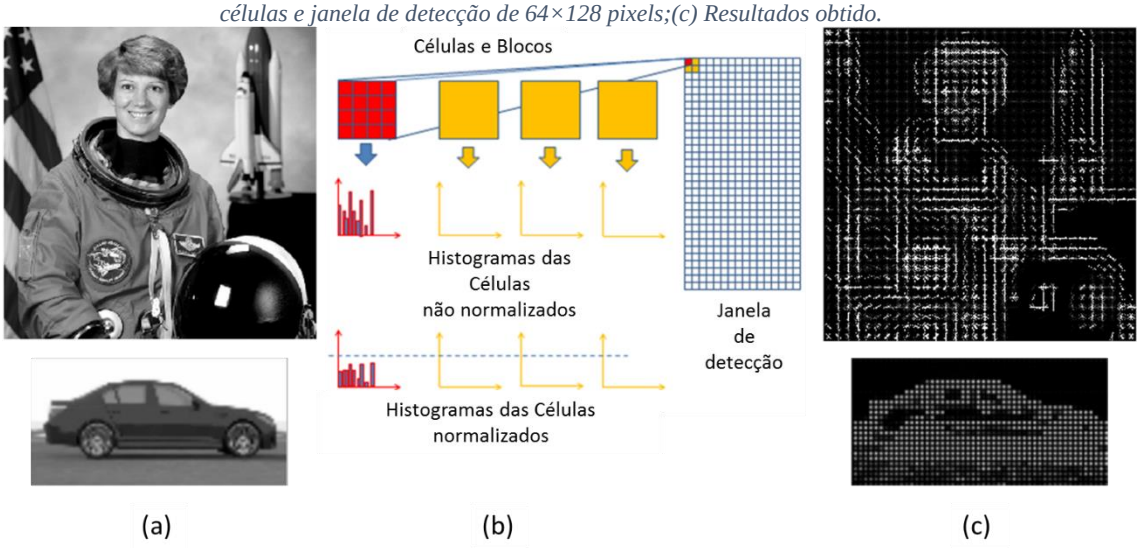
Figura 7: Exemplo da aplicação do descritor LBP.
(a) Procedimento utilizado pelo descritor. (b) Imagem de entrada. (c) Resultado obtido. (d) Histograma produzido pelo descritor.



Fonte: Adaptado de CUNHA (2013).

Já no descritor HOG (*Histogram of Oriented Gradients*) a ideia principal é descrever as formas locais e a aparência dos objetos presentes na imagem através de uma distribuição das direções de suas bordas. Para isso, a imagem é dividida em pequenas regiões chamadas de células (*cells*) que são conectadas formando os blocos (*blocks*) onde são calculados os histogramas de direções de inclinações, ou seja, os histogramas das direções das bordas, sendo assim o descritor é o resultado do agrupamento desses diferentes histogramas. Para se obter um desempenho melhor, é possível normalizar esses histogramas locais por contraste o que proporciona uma melhor invariância às mudanças de sombreamento e iluminação, Figura 8. De modo simplificado este descritor é basicamente uma lista dos histogramas normalizados de forma a permitir que o primeiro plano (objetos ou regiões de interesse) seja diferenciado do plano de fundo (DALAL; TRIGGS, 2005; BEYELER, 2015).

Figura 8 : Exemplo da aplicação do descritor HOG.
(a) Imagens de entrada;(b) Esquema do método utilizando os seguintes parâmetros: células de 4×4 pixels, blocos de 2×2 células e janela de detecção de 64×128 pixels;(c) Resultados obtido.



Fonte: Adaptado de https://software.intel.com/en-us/ipp-dev-reference-histogram-of-oriented-gradients-hog-descriptor e de "http://scikit-image.org/docs/dev/auto_examples/features_detection/plot_hog.html".

Com o avanço dos descritores, métodos surgiram com o intuito de localizar várias áreas de interesse em uma imagem, como por exemplo localizar regiões que apresentem grande variação de intensidade de direção (cantos).

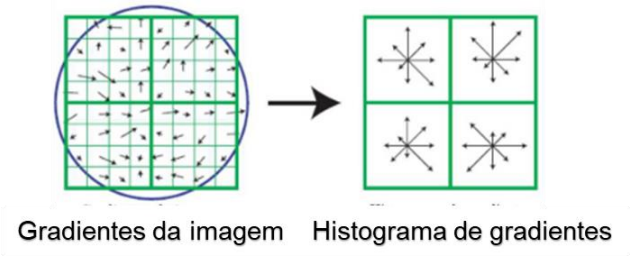
Porém, nem todos os cantos são fáceis de se identificar ou de rastrear ainda mais se a escala do objeto sofrer algum tipo de alteração, como muitas vezes ocorre em arquivos de vídeos que nada mais são do que sequência de imagens digitais (denominada *frames*) que usualmente apresentam uma taxa de atualização de 30 *frames* por segundo, ou seja, para cada segundo do vídeo, são mostradas 30 imagens. Essa taxa de atualização pode variar dependendo do tipo de vídeo: 16 *frames* por segundo, em filmes antigos (NEUMEYER, 2013) até 4,4 trilhões de *frames* por segundo, em gravações de câmera lenta (NAKAGAWA et al., 2014) .

Muitas vezes quando se trabalha com imagens obtidas de vídeos se torna necessário o uso de algumas técnicas de pré-processamento para minimizar imperfeições nas imagens e facilitar a diferenciação dos objetos nela presente , principalmente por diversas dificuldades encontradas quando se trabalha com imagens de objetos em movimento, como por exemplo veículos em velocidade elevada (interferindo na qualidade da imagem por alteração de foco), mudança de luminosidade, alteração do formato do objeto (redução de tamanho e contorno), obstrução e oclusão por interferência de outros objetos, entre outros.

Com essa ótica, o descritor SIFT (*Scale-Invariant Features Transform*) foi desenvolvido com o intuito de permitir localizar e descrever pontos chaves (*key points*) que não sofram com a mudança de tamanho ou com mudança de orientação tanto em imagens estáticas como em vídeos. Este descritor por ser um dos primeiros descritores locais baseado em gradiente é muito citado na literatura. Seu processo básico de

descrição, cálculo da escala dos gradientes seguido da formulação dos histogramas orientados para determinar a orientação principal de uma região envolta do ponto de interesse, está ilustrada na Figura 9.(LOWE, 1999; BAY; TUYTELAARS; VAN GOOL, 2006)

Figura 9:Esquema do descritor SIFT.



Fonte: Adaptado de BUENO (2011).

Posteriormente, o descritor SURF (*Speeded Up Robust Features*) surgiu permitindo a detecção e extração de características locais, mesmo a pequenas alterações de perspectiva, rotação, escala, mudanças de iluminação e ruído nas imagens. A principal característica que torna melhor o desempenho do descritor SURF perante o SIFT é que seu modelo aproximativo do espaço de escala é baseado na imagem integral (BAY; TUYTELAARS; VAN GOOL, 2006; KIM; DAHYOT, 2008). De modo simplificado, o algoritmo SURF pode ser separado em duas partes: (i) detectar os pontos de interesse e (ii) formular um descritor. Para que sua varredura por pontos de interesse seja realizada é necessário determinar um limite para se definir se o ponto deve ser considerado chave ou não, sendo que um valor elevado resulta em quantidade menor de pontos, porém teoricamente de maior importância, o que torna este método teoricamente mais rápido e com resultados de melhor qualidade (BEYELER, 2015). A Figura 10 mostra a utilização do descritor SURF em uma imagem em escala de cinza.

Figura 10: Representação da detecção dos pontos de interesse utilizando o descritor SURF.

Fonte: "https://www.mathworks.com/help/vision/ref/extractfeatures.html".

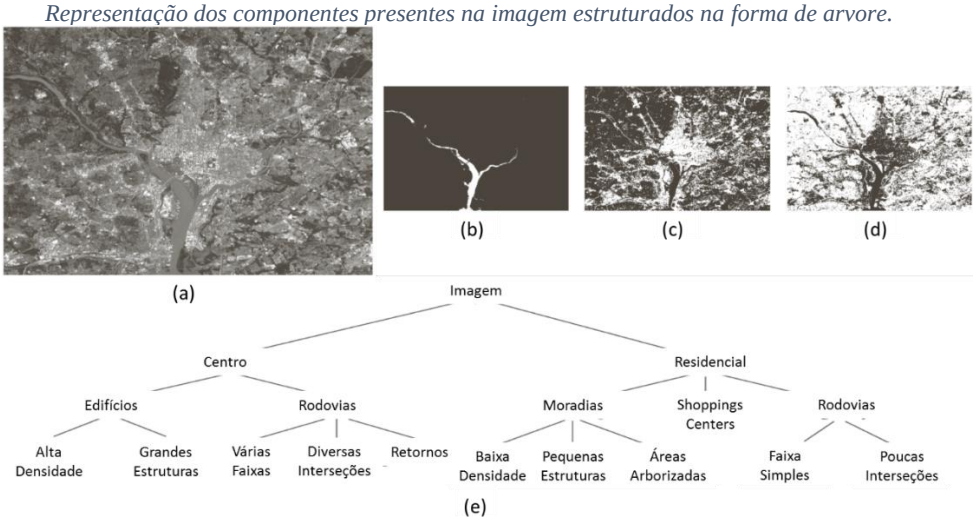
2.1.4. RECONHECIMENTO DE PADRÕES

O processo de representar e descrever uma imagem tem a finalidade de extrair um "significado" dessa imagem, ou seja, identificar objetos ou padrões. Dessa forma, um padrão é definido por um conjunto de descritores, sendo que cada descritor define de forma matemática (modela / descreve) uma característica específica. Ou seja, um padrão é representado por um conjunto de características.

Normalmente são utilizadas três formas de representar padrões básicos, vetores (para informações quantitativas), árvores (para informações estruturais) e *strings* (para informações textuais e ou de caracteres). É importante salientar que o reconhecimento de padrões é altamente dependente da escolha ou seleção das características utilizadas para tal, uma vez que essas *features* são a base para a criação do critério de exclusão ou inclusão nas classes (YU et al., 2003; KIM; DAHYOT, 2008; GONZALEZ; WOODS, 2010; SILVA et al., 2013b). (YU et al., 2003; KIM; DAHYOT, 2008; GONZALEZ; WOODS, 2010; SILVA et al., 2013b). A Figura 11 ilustra algumas formas usuais de se apresentar padrões em geral, não somente em imagens, mas utilizando dois dos tipos básicos comentados anteriormente.

Na Figura 11a após a realização de uma classificação bayesiana foi possível diferenciar as regiões presentes na imagem de forma que os pontos claros representem as regiões classificadas como amostras de água (Figura 11b), regiões urbanas (Figura 11c) e regiões com grande concentração de vegetações (Figura 11d). Na Figura 11e por sua vez é demonstrado os componentes estruturados na forma de árvore.

Figura 11: Exemplos usuais de representação de padrão.
(a) Imagem de satélite do centro da cidade de Washington, DC com seu elevado número de construções e suas zonas residenciais ao redor. Diferenciação das regiões presentes na imagem de forma que os pontos claros representam as regiões classificadas como;(b) amostras de água; (c) regiões urbanas e (d) regiões com grande concentração de vegetações. (e) Representação dos componentes presentes na imagem estruturados na forma de arvore.



Fonte: Adaptado de GONZALEZ; WOODS (2010).

Quando a abordagem escolhida para o reconhecimento de padrão se dá por decisão teórica, o critério é baseado em uma função discriminante de forma a encontrar um valor (ou uma fronteira) que separe as classes existentes no padrão em questão. Algumas das técnicas comumente utilizadas para essa abordagem

são combinação (*matching*), classificadores estatísticos ótimos, redes neurais artificiais (RNA), máquinas de vetores de suporte (*Support Vector Machines* – SVM), entre outras.

2.1.5. APRENDIZADO DE MÁQUINAS

O aprendizado de máquinas (*machine learning*) é uma subárea da ciência da computação que foi desenvolvida através dos estudos no campo do reconhecimento de padrões e da teoria do aprendizado computacional em inteligência artificial. Essa área investiga a construção de algoritmos capazes de "aprender" com seus erros e que conseguem fazer previsões sobre dados a partir da construção de modelos criados com dados de entrada (amostras) com o intuito de realizarem previsões ou decisões guiadas ao invés de seguir uma instrução pré-programada (LANGLEY, 1996; MITCHELL, 1997). Dentro do amplo campo da inteligência artificial (IA), na qual existem dois tipos de "raciocínio" (o indutivo e o dedutivo), a subárea do Aprendizado de Máquinas (AM) somente utiliza o aprendizado indutivo. As técnicas de AM mais utilizadas atualmente são o aprendizado supervisionado e o aprendizado não supervisionado, porém existem também as técnicas de aprendizado semi-supervisionado e por reforço.

Os algoritmos que utilizam o aprendizado supervisionado são treinados a partir de exemplos rotulados, ou seja, o dado de entrada possui previamente a resposta conhecida. A partir da comparação entre a saída conhecida e a do algoritmo, o modelo "aprende" com seus erros. Métodos como a classificação, previsão, regressão permitem que o aprendizado supervisionado consiga inferir, a partir de padrões, os rótulos para dados inéditos (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006; SILVA; SPATTI; FLAUZINO, 2010).

Por sua vez, o algoritmo de aprendizado não supervisionado é alimentado com dados não rotulados, dessa forma a saída do sistema não é conhecido antecipadamente. Sendo assim, por não existirem classes previamente estabelecidas o algoritmo deve por si só identificar algum padrão ou estrutura entre os dados, identificando subconjuntos (*clusters*) de maneira que os elementos de um mesmo grupo apresentem uma alta similaridade entre si (homogeneidade dentro do grupo) e que os elementos de grupos distintos apresentem uma baixa similaridade (heterogeneidade entre grupos). Esse tipo de aprendizado é muito utilizado para explorar os dados e encontrar uma estrutura dentro deles (SILVA; SPATTI; FLAUZINO, 2010; SCHMIDHUBER, 2015).

Os algoritmos de aprendizado "semi-supervisionados" são aplicados similarmente aos do aprendizado supervisionado, com a diferença da existência de alguns dados não rotulados na sua fase de treinamento. Isso ocorre pelo elevado esforço para a aquisição de dados rotulados, fazendo com que nos algoritmos de

aprendizado semi-supervisionados, sua fase de treinamento é composta por uma pequena quantia de dados rotulados e uma grande quantia de dados não rotulados.

No caso dos algoritmos de aprendizado por reforço, o sistema deve descobrir por tentativa e erro quais ações geram os maiores benefícios. Normalmente este sistema é considerado uma variação do aprendizado supervisionado e é composto por três integrantes principais, o indivíduo (tomador de decisões), o ambiente (toda e qualquer interação realizada pelo indivíduo) e as ações (conjunto de reações permitidas ao indivíduo). O objetivo deste tipo de aprendizado é descobrir ou "aprender" qual a melhor política de conduta a ser adotada, uma vez que a única resposta aceitável é a resposta correta de forma que o algoritmo revisa os parâmetros utilizados para produzir a resposta correta reforçando ou beneficiando esses padrões que tiveram um resultado aceitável (SILVA; SPATTI; FLAUZINO, 2010; KIM et al., 2015).

2.1.6. REDES NEURAIS ARTIFICIAIS

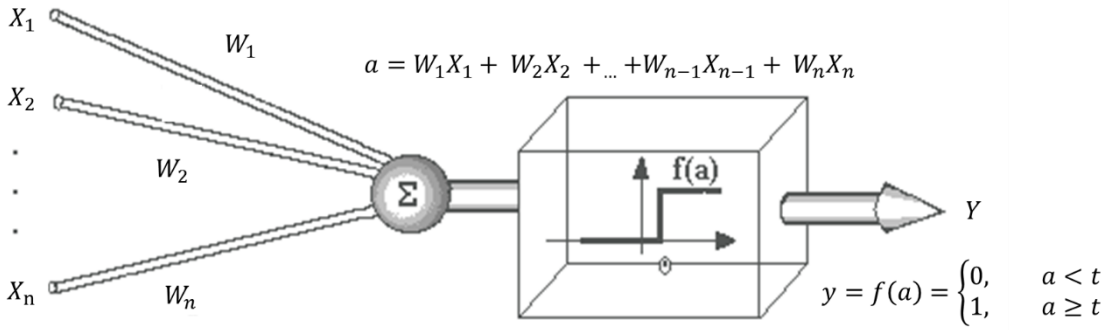
As redes neurais artificiais (RNA) são baseadas no funcionamento do sistema nervoso dos seres vivos, mais precisamente, tentam reproduzir o funcionamento do cérebro humano. Desta forma a rede neural é composta por diversos neurônios (*perceptron*) os quais são organizados em várias camadas. O desenvolvimento deste classificador linear *perceptron* foi realizado inicialmente por McCulloch e Pits em 1943, que de maneira simplificada, permite separar os dados através de uma função linear utilizando um aprendizado supervisionado (MCCULLOCH; PITTS, 1943).

Entre suas principais vantagens mais relevantes se destaca sua capacidade de trabalhar com dados discretos e contínuos, seu elevado grau de generalização mesmo com dados faltantes e/ou informações imprecisas, além de elevado grau de paralelismo. Suas desvantagens incluem a necessidade de um grupo de treinamento previamente rotulado, o fato de seus processos para conseguir a convergência esperada, normalmente serem exaustivos e demandarem um grande número de parâmetros (número de neurônios ou camadas ocultas, pesos e funções de ativação), além da existência de um aprendizado por caixa preta (o conhecimento adquirido pelo sistema não é explicito), entre outros (DUDA; HART; STORK, 2001; GONZALEZ; WOODS, 2010).

Basicamente sua forma de operação ocorre pela multiplicação do conjunto de dados (sinais) por um conjunto de pesos, de forma que a somatória ponderada produza um nível de ativação fazendo o neurônio produzir uma saída, ou seja, caso se tenha um conjunto com "n" dados (conjunto de sinais de entrada – $X_1, X_2, \dots, X_{n-1}, X_n$), deverá se ter o mesmo número "n" de pesos (conjunto de ponderações – $W_1, W_2, \dots, W_{n-1}, W_n$) e uma única função de ativação, no caso de um único neurônio. Desta forma, o

nível de ativação "a" é a somatória da multiplicação dos dados pelos pesos, que deve ser comparado com o limitante "t" presente na função de ativação, resultando no valor da saída y (BRAGA; CARVALHO; LUDERMIR, 2007), como ilustrado pela Figura 12.

Figura 12: Representação esquemática do funcionamento padrão de um perceptron, concebido por McCulloch e Pits.



Fonte: Adaptado de CARVALHO (2009).

Existem muitos tipos de algoritmos de aprendizado que utilizam o modelo de redes neurais, sendo diferenciados pelo tipo de hierarquia em que são organizados (número de neurônios e pela quantidade de camadas) e pelo modo como os pesos são modificados.

Quando o problema em questão não pode ser classificado linearmente por uma única fronteira linear a opção é adicionar mais neurônios em camadas extras (camadas escondidas), de forma que essas camadas adicionais trabalhem como uma espécie de extrator de características fornecendo para o sistema informação complementar permitindo a classificação, sendo configurada nesse caso, uma rede com múltiplas camadas (*multi-layer perceptron – MLP*) (SILVA; SPATTI; FLAUZINO, 2010).

Por se tratar de uma rede com muitos parâmetros atualizados automaticamente (muitos pesos), este sistema tende a se especializar (decorar) os exemplos do conjunto de treinamento, sendo necessário determinar um critério de parada do aprendizado, já que o objetivo real é que o sistema tenha a capacidade de generalizar. Normalmente esse critério é determinado pelo erro observado entre a resposta dada pelo sistema e a resposta esperada. Enquanto o sistema está aprendendo, o valor do erro tende a diminuir tanto no conjunto de treinamento como no conjunto de teste, porém quando o sistema começa a se especializar, o erro observado no conjunto de teste começa a subir. A atualização dos pesos pode ocorrer em dois momentos, a cada apresentação de um padrão ao sistema (versão *on-line* ou padrão) ou ao final do teste (versão *batch*). Na versão *on-line* a correção dos pesos é baseada somente no erro obtido em relação ao exemplo em questão, ocorrendo "n" correções dos pesos. Já na versão *batch* todos os pesos são somados para se determinar um erro médio e só então os pesos são corrigidos a partir deste erro médio calculado (BRAGA; CARVALHO; LUDERMIR, 2007; SILVA; SPATTI; FLAUZINO, 2010; SCHMIDHUBER, 2015).

A função de ativação determina qual a resposta produzida pelo sistema. As funções de ativação mais utilizadas são: (i) a função sinal, na qual o neurônio dispara se o nível de ativação for maior que um limitante "t", (ii) a função identidade, a qual não se utiliza função de ativação, sendo a saída o próprio nível de ativação "a" e (iii) a função *sigmoid*, que pode ser integrada por apresentar máximo de 1 e mínimo de 0 e dentro deste intervalo apresentar comportamento quase linear. Porém, existem diversas funções de ativação na literatura especializada.

Para exemplificar a utilização de classificadores lineares em imagens SILVA et al., (2013b) realizaram a comparação de alguns métodos implementados para a detecção automática de motocicletas em via pública utilizando imagens extraídas de câmeras de monitoramento do tráfego brasileiro. Com o intuito de extrair as informações das imagens, foram utilizados os descritores SURF, HOG, LBP entre outros e para classificar as imagens foram utilizados os classificadores MLP, SVM entre outros. A pesquisa mostrou que o classificador SVM implementado utilizando as funções *Kernel* linear e polinomial tiveram um excelente desempenho (exatidão superior a 0,97 e precisão superior a 0,94) tanto utilizando o descritor LBP como o descritor SURF.

2.1.7. MÁQUINAS DE VETORES DE SUPORTE

Umas das técnicas de AM de bastante notoriedade é a Máquina de Vetores de Suporte (*Support Vector Machines* – SVM), a qual tem apresentado resultados similares ou até superiores aos demais algoritmos de aprendizado como as RNAs (SILVA et al., 2013b; SILVA; AIRES; VERAS, 2018). As SVMs foram desenvolvidas em 1979 por Vladmir Vapnik e são consideradas classificadores lineares binários não probabilísticos baseado no aprendizado supervisionado e na Teoria do Aprendizado Estatístico (do inglês, *Statistical Learning Theory*). Para um conjunto de dados de entrada, o sistema prediz qual das duas possíveis classes cada um dos dados pertence, construindo um modelo na forma $y = f(x)$ (LORENA; CARVALHO, 2007; KIM et al., 2015; SILVA, 2016).

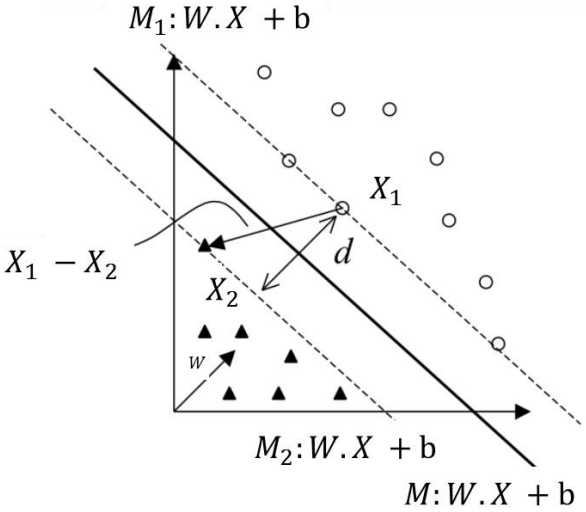
Suas principais características são: (i) boa capacidade de generalização (apresenta grande eficiência na classificação de dados inéditos presentes no conjunto de teste, evitando a especialização nos dados do conjunto de treinamento – *overfitting*), (ii) robustez em dados de grandes dimensões como nos casos de imagens, (iii) processos menos exaustivos para a convergência da função objetivo (uma vez a função de decisão tende a otimizar uma função quadrática, ou seja um único mínimo global) e (iv) baseada em teoria matemática e estatística bem definidas (LORENA; CARVALHO, 2007).

As SVMs encontram uma fronteira linear (hiperplano) de separação dos dados, porém sua principal diferença em relação à RNA está na maximização das distâncias das margens das classes do conjunto de

treinamento, ou seja, as SVMs tentam encontrar o hiperplano ótimo, no qual as distâncias das margens, tanto dos exemplos positivos (rotulados com o valor $+1$) como dos exemplos negativos (rotulados como o valor -1) até o hiperplano sejam iguais. Para isso, os dados do conjunto de treinamento mais próximos do hiperplano são denominados vetores de suporte (*Support Vector* – SV), os quais são utilizados para a definição do hiperplano, ou seja, somente os dados que pertencem às margens dos grupos são utilizados.

A Figura 13 ilustra a representação de um classificador linear por um hiperplano ótimo que utiliza a função sinal como critério de decisão, na qual o hiperplano é descrito por $f(x) = w \cdot x + b = 0$, sendo que w é o vetor normal ao hiperplano e $\frac{b}{\|w\|}$ é a distância do hiperplano em relação à origem (LORENA; CARVALHO, 2007).

Figura 13: Representação da estimativa do hiperplano ótimo de um classificador linear a partir das margens M_1 e M_2 das classes existentes.



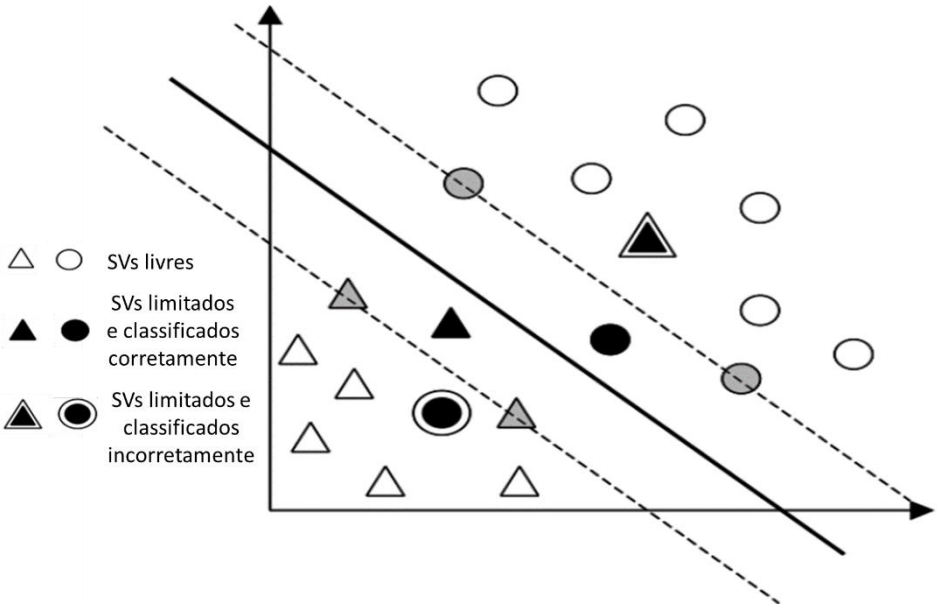
Adaptado de LORENA; CARVALHO (2007).

Quando a maximização da margem, obtida pela minimização de $\|w\|$, garante que dentro da região entre as margens não exista nenhum dado do conjunto de treinamento o sistema é denominado de SVM de margens rígidas, caso contrário, o sistema é denominado de SVM de margens suaves. Normalmente, é difícil encontrar situações reais em que os dados sejam separados linearmente, principalmente pela existência de ruídos e *outliers* (dados inconsistentes ou ambíguos), necessitando assim alterar as restrições da maximização das margens, permitindo a existências de alguns dados entre as margens e acarretando na classificação incorreta de alguns dados.

Esse procedimento é possível através da inclusão de uma constante "c", que determina a quantidade de *outliers* a serem descartados para determinar o hiperplano ótimo, e quanto maior o valor "c", mais dados são permitidos entre as margens. A Figura 14 exemplifica o procedimento de suavização das margens pela introdução de uma variável de folga. Os dados em cinza contribuem para a estimação do hiperplano ótimo,

denominados de SVs livres, já os na cor preta representam os SVs limitados, ou seja, dados que estão entre as margens das classes ou dados do conjunto de treinamento classificados incorretamente.

Figura 14: Representação da estimativa do hiperplano ótimo de um classificador SVM de margens suaves.



Adaptado de LORENA; CARVALHO (2007).

Nos casos em que os dados não são separáveis de maneira satisfatória por um hiperplano, a opção é mapear o conjunto de dados em seu espaço original e representar esses em um novo espaço de maior dimensão, denominado espaço de características (*feature space*) no qual seja possível encontrar um hiperplano para realizar a classificação (LORENA; CARVALHO, 2007; SILVA, 2016).

Como exemplo da utilização do SVM em imagens, a pesquisa de Kim e Dahyot (2008) um detector e classificador de partes do rosto humano foi desenvolvida utilizando o descritor SURF e a técnica SVM para conseguir diferenciar os olhos, nariz e boca presentes em imagens. A técnica de SVM foi utilizada em dois momentos durante a pesquisa, o primeiro para diferenciar quais pontos pertenciam ou não a face da pessoa, em seguida para diferenciar os pontos nas quatro classes procuradas (olho esquerdo, olho direito, boca e nariz). Segundo os autores, a combinação dessas técnicas apresentou uma taxa de detecção e classificação correta a ponto de sugerirem que esta combinação de técnicas deva apresentar bom desempenho mesmo aplicada em vídeo ou na detecção de outros tipos de objetos, como veículos em movimento.

Silva; Aires; Veras (2017) acataram tal sugestão de combinação das técnicas e compararam a combinação de tipos de descritores (SURF, HOG, LBP, entre outros) com diferentes classificadores (MLP, SVM, *Radial Basis Function Network*, entre outros). O objetivo desse trabalho foi detectar motocicletas em vídeos de monitoramento do tráfego e identificar se os usuários de fato utilizavam ou não o capacete de segurança. Essa pesquisa evidenciou que a combinação do uso do descritor SURF com o

classificador SVM apresenta excelente desempenho nos quesitos precisão e exatidão. É importante ressaltar que em ambas pesquisas citadas, por utilizarem o classificador SVM, as imagens utilizadas para treinamento foram previamente padronizadas para que suas dimensões fossem sempre as mesmas, tanto para as imagens dos exemplos positivos como para as imagens dos exemplos negativos, com o intuito de acelerar o processo de aprendizagem.

2.1.8. VALIDAÇÃO CRUZADA

A técnica Validação Cruzada (*Cross Validation*) é amplamente utilizada em pesquisas que tem como objetivo avaliar a capacidade de generalização de um modelo de predição utilizando um conjunto de dados. O conceito central desta técnica baseia-se em estimar na prática a verdadeira precisão do modelo estudado investigando o desempenho deste em um conjunto de dados inéditos (NIXON; AGUADO, 2002; SANTOS, 2002).

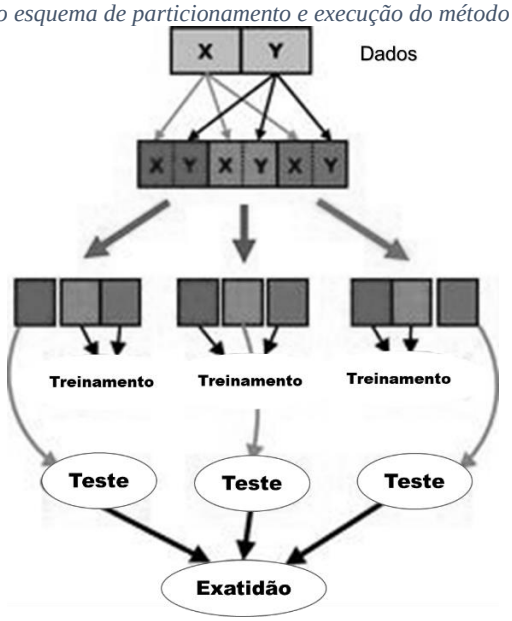
Seu principal procedimento é a divisão do conjunto de dados original em subconjuntos sendo que parte desses subconjuntos são utilizados para estimar os parâmetros do modelo (treinamento) e a outra parte dos subconjuntos são utilizados para validar o modelo (teste).

O critério adotado para realizar o particionamento do conjunto de dados original nos subconjuntos é o principal diferencial dos métodos existentes de validação cruzada.

O método *K-Fold* se caracteriza por subdividir o conjunto de dados total entre k grupos (usualmente entre cinco até dez grupos) de mesmo tamanho. Em seguida é escolhido de maneira aleatória um dos grupos para ser o grupo de teste ou de validação. Os demais grupos são utilizados para se treinar o modelo e serem aplicados ao grupo de teste permitindo uma comparação dos resultados e produzindo uma taxa de erro. O processo de treinamento deve ser repetido até que todos os grupos de treinamento tenham passado pelo grupo de validação fazendo com que todos os grupos apresentem uma taxa de erro. Como última etapa do método é calculado o erro médio obtido do modelo para aqueles k grupos (SILVA; AIRES; VERAS, 2018).

No método *Hold-Out* o conjunto de dados total é subdividido de maneira aleatória em dois grupos de tamanhos diferentes, um com 70% do conjunto de dados total que será usado para treinar o modelo e o outro com 30% do conjunto de dados total que será usado para testar o desempenho do modelo (HE et al., 2015; KIM et al., 2015).

Figura 15:Exemplo do esquema de particionamento e execução do método k-fold com k = 3.

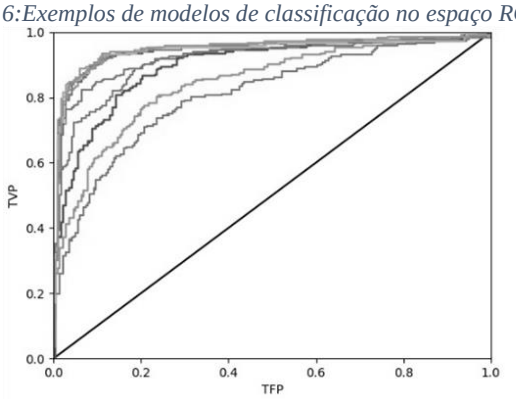


Fonte: Adaptado do Wikipedia.

Uma forma de complementar a validação cruzada permitindo avaliar a predição de um algoritmo de aprendizado é a análise da curva ROC (Receiver Operating Characteristic). Esse método considera a taxa de verdadeiro positivo (TVP) que representa a quantidade de elementos corretamente classificados, eixo das abscissas, e na taxa de falso positivo (TFP) que representa a quantidade de elementos erroneamente classificados, eixo das ordenadas, para representar o modelo de classificação / predição por um ponto no espaço ROC.

Uma vez tendo representado os pontos no espaço ROC de um modelo em questão é possível analisar a métrica AUC (Area Under the ROC Curve) deste modelo. A métrica AUC tem ganhado destaque como medida de avaliação de modelos de mineração de dados e de aprendizado de maquinas por avaliar a qualidade da predição geral do modelo em questão (FAWCETT, 2006; PRATI; BATISTA; MONARD, 2008; XIAO et al., 2014). A Figura 16 ilustra modelos de classificação no espaço ROC.

Figura 16:Exemplos de modelos de classificação no espaço ROC.



A linha diagonal ascendente (0,0 até 1,1) representa um modelo de comportamento aleatório (estocástico).
Fonte: O autor.

As curvas que pertencem ao triangulo superior esquerdo (0,0 ; 0,1 ; 1,1) fazem parte dos modelos com desempenho considerado melhor que o aleatório (0,5 < métrica AUC < 1,0) e as curvas que pertencem ao triangulo inferior direito (0,0 ; 1,0 ; 1,1) fazem parte dos modelos com desempenho considerado pior que o aleatório (0,0 < métrica AUC < 5,0) (FAWCETT, 2006; PRATI; BATISTA; MONARD, 2008; XIAO et al., 2014).

2.2. MATERIAIS

Neste tópico são apresentados os equipamentos utilizados para identificar e classificar motociclistas.

Atualmente, existem ferramentas computacionais disponíveis para o processamento de imagens, em várias linguagens de programação. Algumas dessas são adquiridas somente por licenças pagas e outras distribuídas gratuitamente (*Open Source*). De maneira geral, todas as ferramentas destinadas ao processamento de imagens oferecem pelo menos o pacote básico para a leitura e a gravação de imagens e vídeos, além do processamento no domínio espacial e de frequência.

A revisão bibliográfica, além de auxiliar na escolha das técnicas a serem utilizadas, também permitiu aprender uma linguagem de programação, *Python*, e a oportunidade de se familiarizar com os pacotes e funções existentes nas bibliotecas OpenCV, scikit-image e scikit-learn.

A linguagem de programação *Python*, além de ser livre, tem compatibilidade para vários sistemas operacionais sendo considerada de alto nível. Ela apresenta uma hierarquia estrutural forte, com pacotes, módulos, funções e blocos de códigos aninhados, facilitando assim a escrita dos códigos de maneira lógica e clara.

Foi desenvolvida por Guido van Rossum, em 1991, e atualmente possui um modelo de desenvolvimento comunitário gerenciado pela *Python Software Foundation*, uma organização sem fins lucrativos, que disponibiliza a linguagem no seu site oficial. Outra característica é que essa linguagem se aproxima mais da linguagem do programador do que da linguagem da máquina e como o código fonte é executado por interpretador e em seguida executado pelo processador ou sistema operacional não necessita de um compilador. Por ser uma linguagem interpretativa, sua utilização é bem difundida, seja por programadores particulares ou por grandes empresas, sendo aplicada em diversas áreas e possuindo um número de bibliotecas bem significativo (DEVMEDIA,).

Atualmente, existem duas versões oficiais e distintas da linguagem Python, ambas em produção e atualização constantes, a versão *Python* 3 e a versão *Python* 2. Os pacotes da versão 2 continuam sendo atualizados, tanto para adição de funções novas como para a correção de problemas. Por sua vez, a versão

3, mais recente, apresenta principalmente uma nova reestruturação na sintaxe da linguagem com o objetivo de padronizar algumas inconsistências que ocorriam na versão anterior (*Python* 2) (SOUZA,).

A biblioteca de código aberto *OpenCV* (*Open Source Computer Vision Library*) é composta por centenas de funções de Visão Computacional sendo totalmente livre para o uso comercial e acadêmico, o que tem possibilitado o desenvolvimento de inúmeros programas nessa área desde seu lançamento oficial (HOWSE, 2013; BEYELER, 2015).

Apresenta em seu acervo módulos de processamento de imagens e vídeos, álgebra linear, interface gráfica ao usuário básica (*Graphical User Interface* – GUI), estrutura de dados, além de algoritmos de visão computacional. O que possibilita a aplicação desta biblioteca para o reconhecimento de movimentos, reconhecimento de faces, interação Humano-Computador, gravações de vídeos, identificação de objetos, reconstrução 3D, realidade virtual, entre outras (HOWSE, 2013; BEYELER, 2015).

A biblioteca *scikt-image* trata de uma coletânea de algoritmos desenvolvidos para o processamento de imagens digitais disponíveis integralmente sem nenhum tipo de restrição escritos na linguagem de programação *Python* com códigos de alta qualidade e desenvolvidos por uma comunidade de voluntários ativa. Por sua vez, a biblioteca *scikit-learn* é uma ferramenta simples e de grande eficiência para a mineração de dados, análise de dados e aprendizado de máquinas. Como as demais, trata-se de uma biblioteca de código aberto com licença comercial do tipo BSD (*Berkeley Software Distribution*) e desenvolvida na linguagem de programação *Python* ("SCIKIT - IMAGE",).

Para realizar o delineamento do tema abordado, primeiramente foi necessário realizar uma revisão bibliográfica para analisar os trabalhos e as técnicas utilizadas atualmente, buscando as que tiveram bom desempenho mais próximos da meta desta pesquisa de mestrado.

É importante ressaltar que por existir uma gama enorme de técnicas para cada etapa necessária para o desenvolvimento de um sistema de detecção automática de veículos a partir de imagens, torna-se quase que impossível revisar todos os métodos disponíveis. Porém, a realização de uma revisão bibliográfica permite uma tomada de decisão mais assertiva no processo de escolha das técnicas a serem utilizadas na pesquisa, sendo que a busca teórica foi concentrada principalmente nos livros (GONZALEZ; WOODS, 2010), (DUDA; HART; STORK, 2001), nos artigos (SILVA et al., 2013b), (LORENA; CARVALHO, 2007) e também na dissertação de mestrado de (SANTOS, 2002) e na tese de doutorado de (CUNHA, 2013).

De forma geral, os materiais utilizados para a realização desta pesquisa foram: câmera, tripé e computador. As câmeras foram configuradas para a resolução de captura de vídeo digitais de 1920×1080 pixels. O computador utilizado para o processamento dos algoritmos foi um Core i7, com processador de 2,40 GHz,

memória RAM de 8 GB e sistema operacional Windows 8. Os códigos foram escritos em linguagem *Python 2*, utilizando a plataforma Anaconda versão 2.7.13[1], em conjunto com as bibliotecas de funções de processamento de imagens *OpenCV*[2] versão 3.0.0, *scikit-image*[3] e a biblioteca de funções de aprendizado de máquinas *scikit-learn* [4].

2.3. MÉTODO PROPOSTO

A estrutura deste *Capítulo 2* está dividida em duas seções principais: 1) Criação do Banco de Imagens de Veículos (BIV) e 2) Desenvolvimento de um Classificador do tipo SVM.

2.3.1. CRIAÇÃO DA BASE DE IMAGENS DE VEÍCULOS (BIV)

A utilização de uma Base de Dados (BD) facilita a análise e a tomada de decisão de maneira mais assertiva, por se tratar de um conjunto de dados ou de informações organizadas com o intuito de facilitar a interpretação e extração de padrões sobre um tema específico.

Apesar de existir vários repositórios de imagens disponíveis na internet, tais como: *Imagenet*[5], *Pixabay*[6], *Compfight*[7], as imagens de motocicletas e similares nessas bases não se enquadram nas necessidades desta pesquisa, pois as imagens disponíveis não estão inseridas em trechos urbanos ou na corrente de tráfego, na verdade, esses repositórios trazem imagens de motocicletas de exposição ou de acrobacias. A Figura 17 exemplifica o tipo de imagens encontradas nesses repositórios. Dessa forma, optou-se pela criação de uma Base de Imagem de Veículos (BIV) extraídas de maneira manual, diretamente das filmagens obtidas através da coleta de dados.

Figura 17: Exemplos de imagens referentes a motocicletas e similares disponíveis nos repositórios citados anteriormente.

Fonte: Imagenet, Pixabay e Compfight.

1 Plataforma Anaconda disponível no site <https://www.continuum.io/downloads>
2 OpenCV disponível no site <http://opencv.org>
3 Scikit-image disponível no site <https://scikit-image.org/>
4 Scikt-learn disponível no site <http://scikit-learn.org/stable/install.html>
5 Imagenet disponível no site <https://image-net.org>
6 Pixabay disponível no site <https://pixabay.com>
7 Compfight disponível no site <https://compfight.com>

O principal objetivo desta seção, criação de um BIV, é obter imagens de motocicletas (exemplos positivos) e imagens dos demais veículos e da paisagem de fundo (exemplos negativos). Com o intuito de que a tomada de decisão em relação aos parâmetros adotados para a criação desse BIV fosse a mais assertiva possível, esta primeira parte do método proposto foi subdivida em 2 etapas:1.1. Coleta dos Vídeos do Tráfego; 1.2. Extração das Imagens dos Vídeos do Tráfego e Separação da Imagens nas Classes desejadas.

Etapa 1.1. Coleta dos Vídeos do Tráfego

Para que a coleta dos vídeos do tráfego fosse realizada antes foi preciso escolher o melhor local para se realizar essa coleta. Como comentado anteriormente na seção 2.1.1, os dados do DENATRAN(2018) mostram que no decorrer dos anos houve um aumento da frota de motocicletas no país. O motivo desse aumento pode ser por diversos motivos, uma vez que há fatores sócio-político-econômicos envolvidos nessa questão. A motocicleta proporciona deslocamento de forma mais prática e eficiente dentro da malha urbana, além de ser econômica tanto em questões de manutenção como de aquisição o que estimula a população optar por este tipo de veículo em detrimento dos demais.

Outro fator também comentado na mesma seção é que com o aumento do volume de veículos maiores são as chances de ocorrerem acidentes do transito e como o estado de São Paulo possui a maior parcela da frota nacional, foi realizado uma análise da relação do tamanho da frota com o total de óbitos no trânsito dos principais municípios das principais macrorregiões do estado com o intuito de escolher a melhor cidade a realizar a coleta.

Na Tabela 1 estão demostrados os cenários dos principais municípios das macrorregiões de Bauru, Campinas, Ribeirão Preto, São José do Rio Preto e da região Central pertencentes ao estado de São Paulo. Analisando os cenários desses municípios foi possível evidenciar que a cidade de Araraquara apresenta uma população e um volume de frota considerável além de apresentar um dos maiores índices de óbito ficando entre as quatro cidades com maior número de acidentes fatais da região, sendo que uma parcela considerável dessas vítimas se locomovia através de motocicletas (39% dos acidentes fatais envolviam motociclistas).

Tabela 1: Relatório estatístico envolvendo tamanho da frota e total de óbitos dos principais municípios das Regiões de Bauru, Campinas, Ribeirão Preto, São José do Rio Preto e da Região Central do estado de São Paulo

Município	Frota de veículos (2016) *					IDH **			Informações do município ***			
	Total	Automóvel		Motocicleta		2010	Ranking	População	Total óbitos (2016)	Locomoção da vítima (motocicleta)		Índice (óbito/100 mil habitantes)
		Absoluto	% do total	Absoluto	% do total					Absoluto	% do total	
Limeira	204.092	122.669	60,1	43.662	21,4	0,775	90	286.882	63	16	25,4	22,0
Jundiaí	308.524	201.842	65,4	49.356	16,0	0,822	4	391.040	85	23	27,1	21,7
Catanduva	102.022	52.465	51,4	30.173	29,6	0,785	50	115.288	25	3	12,0	21,7
Araraquara	**171.139**	**100.778**	**58,9**	**39.829**	**23,3**	**0,815**	**7**	**219.631**	**41**	**16**	**39,0**	**18,7**
Bragança Paulista	118.264	66.887	56,6	31.982	27,0	0,776	83	156.241	29	10	34,5	18,6
Votuporanga	79.554	38.544	48,5	27.609	34,7	0,790	35	88.477	16	5	31,3	18,1
São José do Rio Preto	369.620	203.147	55,0	96.955	26,2	0,797	28	430.312	74	25	33,8	17,2
Monte alto	38.362	17.913	46,7	13.579	35,4	0,768	121	47.664	8	2	25,0	16,8
Jaú	98.817	55.589	56,3	25.901	26,2	0,778	75	139.844	23	10	43,5	16,4
Lins	56.805	29.871	52,6	15.212	26,8	0,786	47	73.484	12	2	16,7	16,3
Bauru	268.019	167.471	62,5	54.983	20,5	0,801	20	354.928	56	28	50,0	15,8
Campinas	864.782	589.772	68,2	124.125	14,4	0,805	17	1.134.546	170	54	31,8	15,0
Jales	42.400	20.234	47,7	14.066	33,2	0,776	83	47.170	7	1	14,3	14,8
Ribeirão Preto	505.903	286.891	56,7	131.951	26,1	0,800	22	647.862	82	26	31,7	12,7
Jaboticabal	54.410	29.127	53,5	13.819	25,4	0,778	75	73.091	9	1	11,1	12,3
Sertãozinho	84.806	42.728	50,4	24.037	28,3	0,761	160	116.832	14	4	28,6	12,0
Rio Claro	160.011	83.824	52,4	50.130	31,3	0,803	19	194.087	23	7	30,4	11,9
Piracicaba	290.409	174.610	60,1	57.932	19,9	0,785	50	378.185	42	14	33,3	11,1
São Carlos	171.988	109.916	63,9	33.327	19,4	0,805	18	233.249	25	4	16,0	10,7
São João da Boa Vista	66.496	37.323	56,1	17.209	25,9	0,797	28	85.695	8	2	25,0	9,3
Fernandópolis	55.315	27.915	50,5	16.811	30,4	0,797	28	65.552	4	1	25,0	6,1

** Segundo os Municípios da Federação*

*** Divulgado em 2010 pelo PNUD (Programa das Nações Unidas para o Desenvolvimento)*

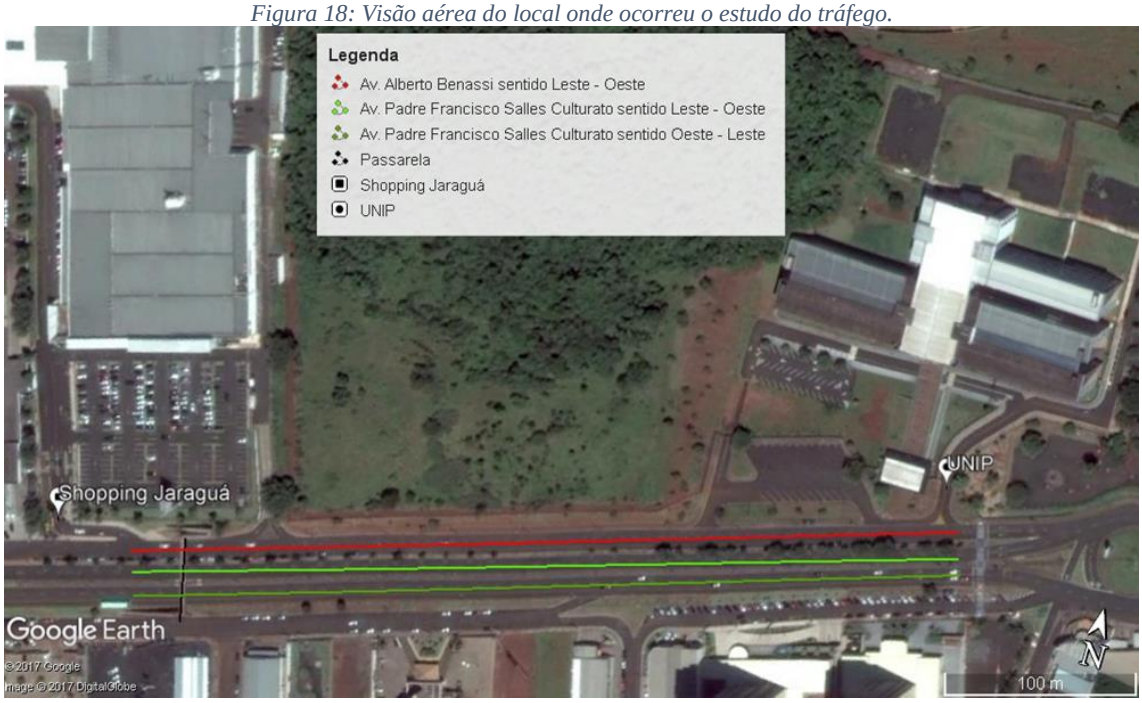
**** Segundo o INFOSIGA (Sistema de Informações de Acidentes de Trânsito do Estado de São Paulo*

Fonte: O autor sobre os dados do DENATRAN (2018).

Sendo assim, o município de Araraquara foi considerado um candidato a se realizar a coleta dos vídeos do tráfego, entretanto era necessário verificar se neste município existia algum local que se adequasse aos tipos de imagens que se desejava coletar, imagens dos veículos dentro da corrente de tráfego e capturando o veículo por cima, tentando reproduzir o mais fiel possível as imagens obtidas pelas câmeras de monitoramento do tráfego, normalmente instaladas nas rodovias. Para isso foi realizado um *estudo do tráfego* e uma *coleta piloto* de dados no município de Araraquara, antes de se realizar de fato a *coleta definitiva* dos vídeos.

O *estudo do tráfego* foi realizado nas Avenidas Alberto Benassi e na Avenida Padre Francisco Salles Culturato, entre a entrada da Universidade Paulista (UNIP) e a entrada do Shopping Jaraguá, em um trecho de aproximadamente 400 metros (Figura 18). Essas avenidas são principais interligações entre a cidade e a rodovia Washington Luís (BR-364 / SP-310 no acesso próximo ao km 276). O estudo foi realizado sobre uma passarela existente neste trecho, posicionando câmeras digitais apontadas para a direção Leste, permitindo a captura das imagens de uma posição superior ao sentido do fluxo de tráfego (Figura 19).

Figura 18: Visão aérea do local onde ocorreu o estudo do tráfego.



Fonte: O autor sobre imagem obtida do Google Earth.

A Avenida Alberto Benassi é uma via marginal que dá acesso às entradas da universidade e do shopping. A captura das imagens desta avenida foi obtida no sentido contrário ao fluxo – sentido Leste-Oeste, obtendo dessa forma imagens frontais e superiores dos veículos. Já a Avenida Padre Francisco Salles Culturato é uma via principal controlada e a captura das imagens do tráfego foi obtida em ambos os sentidos – Oeste-Leste (a favor do fluxo) registrando imagens traseiras e superiores dos veículos, e no sentido Leste-Oeste (contrário ao fluxo) capturando imagens frontais e superiores dos veículos.

Figura 19: Imagem da passarela Vereador Elias Damus, na direção Oeste, tirada na Av. Padre Francisco Salles Culturato sentido Leste-Oeste.

Fonte: Google Earth.

É importante ressaltar que esse estudo auxiliou a verificar se o fluxo de veículos e a posição das câmeras a serem instaladas condiziam com os interesses desta pesquisa e que uma vez comprovado que o local se mostrou adequado, uma *coleta piloto* foi realizada para coletar os dados sobre o fluxo de veículos das três vias existentes e poder determinar qual delas seria a mais adequada (fluxo absoluto de veículos maior e maior participação de motocicletas na corrente de tráfego) para a realização da *coleta definitiva* dos vídeos do tráfego.

A *coleta piloto* dos dados por sua vez foi realizada no dia 4 de abril de 2017 em dois períodos da manhã, próximos aos horários de entrada e saída da UNIP. O primeiro período ocorreu das 08:00 horas até as 08:40 compreendendo o início das aulas da manhã da universidade (8:25 horas) e o segundo período ocorreu das 10:45 horas até as 11:40 horas, cobrindo a saída das aulas matutinas (11:15 horas).

O propósito desta *coleta piloto* foi observar o volume e a taxa de veículos em cada sentido das vias e a qualidade das imagens, para verificar se a posição das câmeras e o fluxo nestas vias seriam capazes de representar a situação real encontrada nos trechos urbanos brasileiros.

A fim de obter uma amostra maior para compor o BIV a *coleta definitiva* de dados foi realizada no dia 18 de abril de 2017, na Avenida Padre Francisco Salles Culturato no sentido Oeste-Leste, com o auxílio de duas câmeras de vídeo, as quais foram posicionadas, uma a favor do fluxo e a outro contrária ao fluxo. O período de filmagem ocorreu da 10:50 às 14:10, totalizando 3 horas e 20 minutos de gravação.

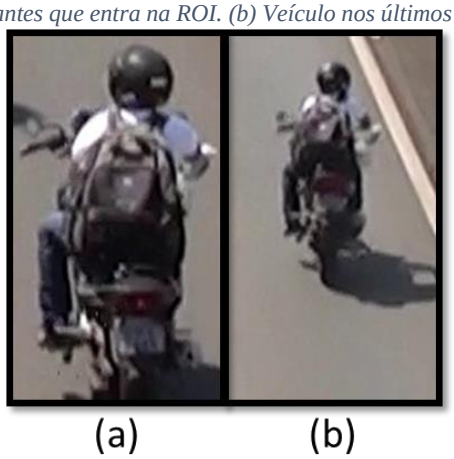
Ao todo foram coletados 15 vídeos de aproximadamente 12 minutos cada e 1 arquivo de aproximadamente 7 minutos para cada câmera (ao total 32 vídeos), sendo que todas as gravações foram realizadas em Full HD (1920 x 1080) com 30 fps.

Etapa 1.2. Extração de Imagens dos Veículos dos Vídeos do Tráfego e Separação das Imagens nas Classes Desejadas

Uma vez coletados os vídeos de tráfego, o passo seguinte é extrair as imagens dos tipos de veículos presentes na corrente de tráfego para treinamento do algoritmo de classificação.

Vale ressaltar que, nos vídeos obtidos no sentido a favor da corrente de tráfego, os veículos inicialmente aparecem com grandes dimensões e conforme seguem seu percurso o seu tamanho é reduzido (Figura 20). Por outro lado, os vídeos no sentido contrário a corrente de tráfego, os veículos surgem com dimensões reduzidas e ao se aproximar da câmera suas dimensões são ampliadas. Por este motivo optou-se por utilizar somente os vídeos a favor do fluxo para a extração de imagens de veículos. Desta forma de todos os 32 vídeos coletados somente metade foram utilizados nesta pesquisa.

Figura 20: Imagem ilustrando a alteração do tamanho dos veículos ao longo do percurso dentro da ROI. (a) Veículo nos primeiros instantes que entra na ROI. (b) Veículo nos últimos instantes antes de sair da ROI.



Fonte: O autor.

Seguindo a mesma abordagem de KIM; DAHYOT (2008) e SILVA; AIRES; VERAS(2017), optou-se por criar um banco de imagens próprio, de tamanhos pré-determinados e devidamente rotuladas, com o propósito de ter amostras verdadeiros positivos (motos) e verdadeiros negativos (não motos). Para isso, *a priori* foi determinado uma região de interesse (*ROI*) nos vídeos para extrair as imagens do banco de dados de forma automática.

Após análise das imagens e compreensão da redução do tamanho dos veículos durante o percurso gravado, foi definido para a imagem padrão as dimensões de 210x120 pixels, dimensões estas suficientes para enquadrar as motocicletas por completo nos momentos iniciais dos vídeos a favor da corrente de tráfego. O método empregado para extrair as imagens padronizadas foi aplicando uma amostragem de

subimagens (ou células) nas dimensões padronizadas adotadas (120x210) em todos os frames, resultando em uma malha de 3x8, como ilustra a Figura 21.

Durante a identificação manual da presença e ausência de veículos nos vídeos de monitoramento realizada pelo autor, somente quando os veículos presentes nos vídeos estavam os mais enquadrados possíveis dentro da célula, no caso das motocicletas, ou de mais de uma célula, no caso dos veículos que não eram motocicletas, que as subimagens eram extraídas. Desta forma para cada *frame* em que se verificava a presença de um veículo dentro da ROI e se conseguia enquadrar este veículo dentro de uma das vinte e quatro células que constituem a malha, todas as vinte e quatro subimagens deste *frame foram extraídas.* Ao todo foram extraídas mais de 60 mil subimagens.

Figura 21: Representação da malha projetada nos vídeos contendo 24 células de 120x210 pixels

Fonte: O autor.

Sendo assim, é possível destacar pela a Figura 21 que as imagens representam a real situação encontrada em uma corrente de tráfego urbano diferentemente do que ocorre com as imagens de exposição ou de acrobacias de motocicletas disponíveis nos repositórios online (vide Figura 17).

Tendo as subimagens de veículos foi possível realizar uma classificação manual das imagens (separação em arquivos diferentes) de acordo com as classes desejadas: (i) "M", contendo 721 imagens de motos e (ii) "NM", contendo 13.393 imagens de exemplos de qualquer outro tipo de veículo ou parte de veículo.

Desta forma o banco de imagens de veículos foi criado e rotulado, possibilitando o desenvolvimento do classificador de veículos do tipo motocicletas utilizando imagens extraídas de vídeos de monitoramento do tráfego

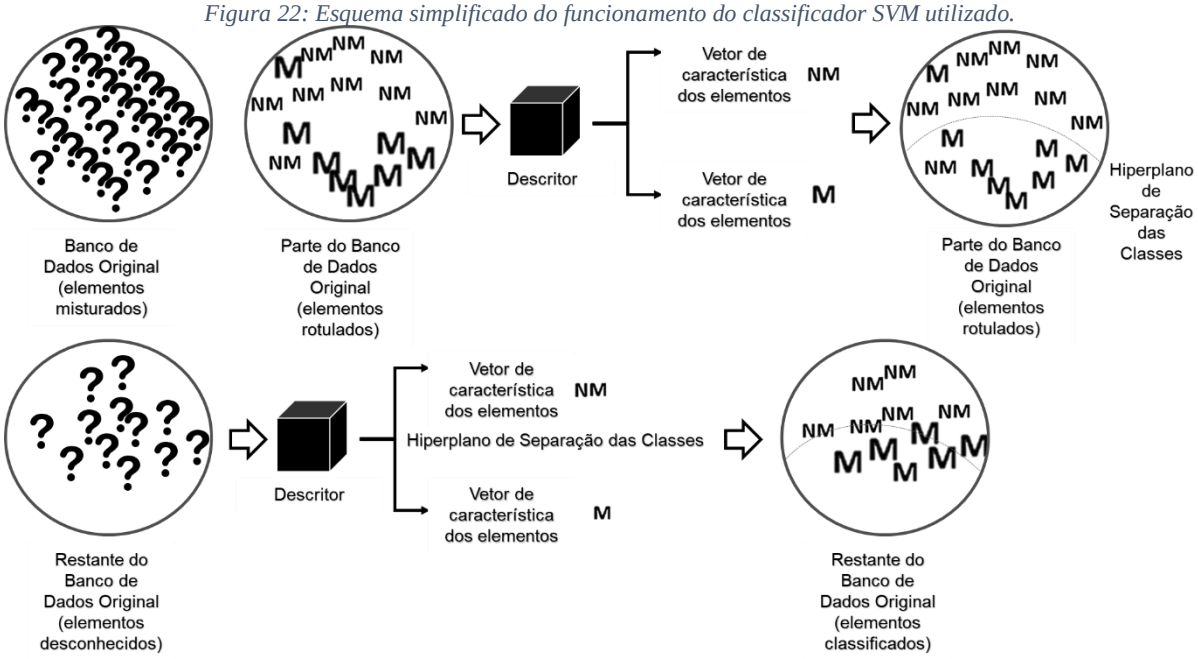
2.3.2. TREINAMENTO DE UM CLASSIFICADOR DO TIPO SVM

Um classificador tem a capacidade de aprender a partir das informações extraídas por um descritor e identificar um tipo ou classe de objetos presentes em uma imagem.

Optou-se por utilizar o classificador SVM (*Support Vector Machine*) que é uma técnica de AM baseada no aprendizado estatístico e de forma supervisionada que tem apresentado bom desempenho no quesito classificação de imagens, classificação de veículos em geral, especialmente motocicletas mesmo quando comparado com outras técnicas (XIAO et al., 2014; HE et al., 2015; CAO et al., 2016; FU et al., 2016; SILVA; AIRES; VERAS, 2018). Além disso, como seu acesso é disponibilizado por funções pré-estabelecidas nas bibliotecas do OpenCV e Scikit-learn existem diversos materiais e fóruns de discussão na internet tornando a assistência e suporte a esse tipo de classificador mais acessível.

O classificador SVM é uma técnica que utiliza um conjunto de dados de entrada denominado conjunto de treinamento, o qual pode ser composto por um conjunto de imagens de veículos que podem ou não ser rotuladas. Neste trabalho foram usados os rótulos "M" para motos e "NM" para não motos. A partir da extração de informações significantes em forma de vetores de características (realizado por descritores) e da rotulação das imagens deste conjunto de treinamento, o algoritmo do classificador SVM modela um hiperplano de separação das classes existentes possibilitando a classificação de imagens inéditas em uma das classes definidas (KIM; DAHYOT, 2008; XIAO et al., 2014; CAO et al., 2016; FU et al., 2016). A Figura 22 ilustra o funcionamento de um classificador SVM.

Figura 22: Esquema simplificado do funcionamento do classificador SVM utilizado.



Fonte: O autor.

O principal objetivo desta seção, é calibrar os parâmetros de um classificador do tipo SVM utilizando o banco de imagens de veículos (BIV) rotulados permitindo que este seja capaz de identificar veículos do tipo motocicletas em imagens extraídas de vídeos do tráfego. Com o intuito de que a tomada de decisão em relação a calibração dos parâmetros adotados ocorresse da maneira mais assertiva possível, esta

segunda parte do método proposto foi subdivida em quatro etapas: 2.1. Geração dos Vetores de Características (VD); 2.2. Definição de Cenários a se utilizar na Calibração dos Parâmetros do Classificador SVM; 2.3. Definição das Amostras e Subconjuntos de Treinamento e Subconjuntos de Teste e 2.4. Validação Cruzada.

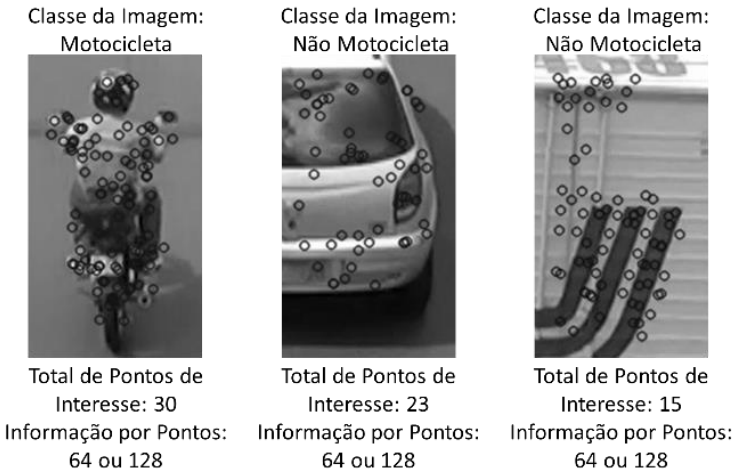
Etapa 2.1. Geração dos Vetores de Características

Para que o classificador do tipo SVM consiga classificar um dado inédito dentre as classes existentes, a técnica utiliza a criação de vetores de características para determinar quais elementos pertencem a cada classe disponível. Desta forma o vetor de características a ser adotado deve sempre apresentar a mesma quantidade de informação para qualquer elemento inédito ou não que a ele seja apresentado (CORINNA; VAPNIK, 1995; PONTIL; VERRI, 1998), estipulando que o vetor características adotado tenha um padrão muito bem definido para facilitar e padronizar a classificação realizada. Através da revisão na literatura especializada e na realização de testes, foi possível comparar os descritores LBP e SURF, e verificar qual destes seria capaz de produzir um vetor de características apropriado levando em consideração os objetivos desta pesquisa.

A aplicação do descritor SURF na imagem resulta em um conjunto de pontos de interesse com quantidade variável de elementos, sendo que cada um desses elementos também apresentam uma quantidade distinta de informações, 64 ou 128 dimensões por ponto de interesse encontrado.

Para se produzir um vetor características padronizado utilizando este tipo de descritor viu-se a necessidade de que para cada ponto de interesse encontrado por imagem fosse produzido um histograma com as 64 ou 128 dimensões / informações disponíveis. Este procedimento se mostrou impraticável devido à grande quantidade de pontos de interesse encontrados pela técnica para cada imagem e pelo número de imagem a serem utilizadas. Dessa forma esta técnica se mostrou inviável para a criação de vetores de características a serem utilizados neste trabalho. A Figura 23 mostra as respostas da função SURF em linguagem Python da biblioteca OpenCV.

Figura 23: Os círculos representam as regiões das imagens consideradas interessantes (cantos) pela técnica SURF.



Fonte: O autor.

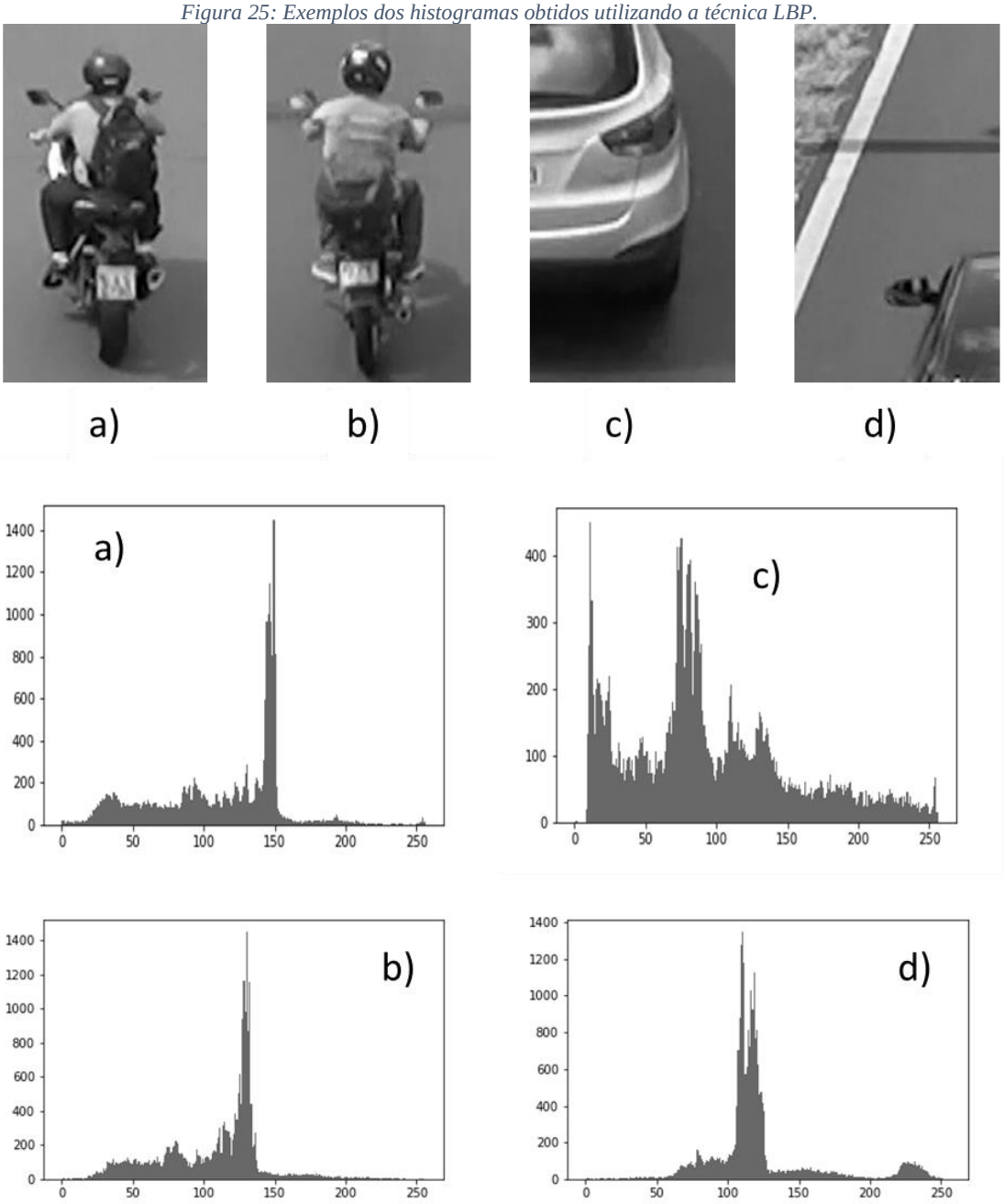
No caso do descritor de textura LBP, o resultado é uma imagem com códigos LBP que é representada por um histograma normalizado, ou seja, o vetor de característica teria sempre informação padronizada. Desta maneira optou-se por utilizar essa técnica como a geradora do vetor de característica a ser fornecido ao classificador SVM. A Figura 24 mostra a implementação da função LBP em linguagem Python disponível na biblioteca scikit-image com os respectivos parâmetros adotados para a produção do histograma em escala de cinza normalizado e a Figura 25 mostra alguns exemplos das imagens utilizadas e os respectivos histogramas fornecidos.

Figura 24: Parâmetros adotados para a função LBP.

```
skimage.feature.local_binary_pattern(
                                     image, "Graylevel image."
                                     P = 3, "Number of circularly symmetric neighbour set points."
                                     R = 8 * P, "Radius of circle (spatial resolution of the operator)"
                                     method = "uniform" {"default", "ror", "uniform", "var"}, "Method to determine the pattern."
                                              )
```

Fonte: O autor.

Figura 25: Exemplos dos histogramas obtidos utilizando a técnica LBP.



Fonte: O autor.

Etapa 2.2. Definição de Cenários a se utilizar na Calibração dos Parâmetros do Classificador SVM

A função LinearSVC escolhida para desenvolver o classificador do tipo SVM está disponível na biblioteca scikit-learn e realiza a separação das classes existentes por um limiar linear utilizando 12 parâmetros a serem configurados para cada situação em particular em que o classificador for utilizado, Figura 26. Entretanto como não se encontrou informações a respeito dos intervalos dos parâmetros para situações envolvendo o tráfego a solução foi realizar uma análise de sensibilidade.

Após uma análise prévia de sensibilidade dos parâmetros verificou-se que apenas metade destes parâmetros produziam uma mudança significativa nos resultados obtidos (a classificação realizada pelo algoritmo para algumas imagens sofria alteração conforme o valor de determinados parâmetros eram alterados). Por esse motivo apenas metade desses parâmetros foram realmente calibrados por serem capazes de causar alguma interferência no tipo de classificação desejada, os outros parâmetros foram mantidos em sua configuração padrão.

Figura 26: Parâmetros da função LinearSVC do classificador SVM

```
sklearn.svm.LinearSVC(
                        penalty='l2', {"l1", "l2"} "Specifies the norm used in the penalization."
                        loss='squared_hinge', {"hinge", "squared_hinge"} "Specifies the loss function."
                        dual=True, {"True", "False"} "Select the algorithm to either solve the dual or primal optimization problem."
                        tol=0.0001, "Tolerance for stopping criteria."
                        C=1.0, "Penalty parameter C of the error term."
                        multi_class='ovr', {"ovr","crammer_singer"}"Determines the multi-class strategy if y contains more than two classes."
                        fit_intercept=True,
                        intercept_scaling=1,
                        class_weight=None,
                        verbose=0,
                        random_state=None,
                        max_iter=1000
                        )
```

A análise de sensibilidade dos 6 principais parâmetros foi realizada através do desenvolvimento de 20 cenários distintos, ou seja, 20 combinações diferentes dos conjuntos de parâmetros possíveis. Para avaliar o desempenho de cada um destes cenários utilizados, a cada um dos conjuntos de predição obtidos pelos grupos de teste realizou-se uma análise das métricas : (i) a taxa de verdadeiro positivo (TVP), que representa a quantidade de imagens corretamente classificadas como motocicletas, (ii) a taxa de falso positivo (TFP), que representa a quantidade de imagens erroneamente classificadas como motocicletas, (iii) a taxa de verdadeiro negativo (TVN), que representa a quantidade de imagens corretamente classificadas como não motos, (iv) a precisão, (v) a exatidão, como mostrado nas Tabela 2 para complementar o critério de avaliação calculou-se também (vi) o valor da métrica AUC, que avalia o desempenho geral de predição do classificador em aleatório (valor próximo de 0,5) ou não aleatório (quanto mais próximo o valor for de 1,0, melhor o seu desempenho geral) (FAWCETT, 2006).

Tabela 2: Métricas convencionais usadas para a avaliação do reconhecimento de objetos

Referência (Rotulação)	Resultado (Classificação)		Total (T)
	Positivos (Motocicletas)	Negativos (Não-Motocicletas)	
Positivos (Motocicletas)	Verdadeiro Positivo (VP)	Falso Negativo (FN)	Positivos (P)
Negativo (Não-Motocicletas)	Falso Positivo (FP)	Verdadeiro Negativo (VN)	Negativos (N)
Taxa de Verdadeiro Positivo $= \frac{VP}{P}$		Taxa de Falso Positivo $= \frac{FP}{N}$	
Precisão $= \frac{VP}{VP+FP}$		Exatidão $= \frac{VP+VN}{P+N}$	

Fonte: Adaptado de CUNHA (2013).

Etapa 2.3. Definição das Amostras e Subconjuntos de Treinamento e Subconjuntos de Teste para a realização da Validação Cruzada

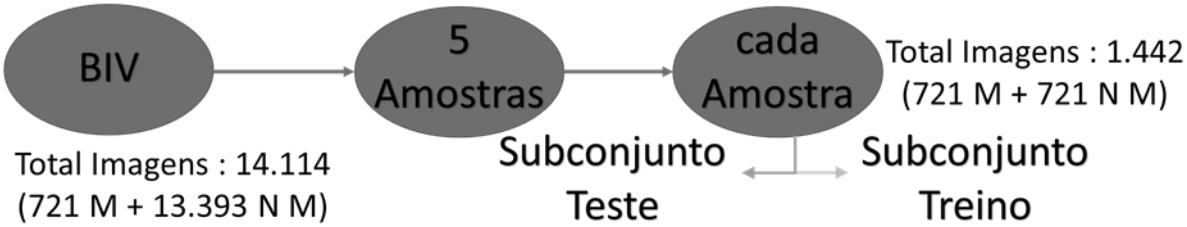
Como a calibração do classificador desenvolvido gerou um conjunto considerável de possibilidades, optou-se por particionar o BIV criado em cinco amostras de forma a analisar cada um dos 20 cenários de maneira mais efetiva permitindo escolher o cenário com a melhor média de acertos de classificação das cinco amostras testadas, de maneira similar ao método de validação cruzada *K-Fold* com $k = 5$.

Entretanto como a quantidade de imagens disponíveis no BIV é diferente para cada uma das classes desejadas (721 imagens de motos e 13.393 imagens de exemplos de não moto) e o intuito da pesquisa é identificar imagens de veículos do tipo motocicletas, utilizou-se a quantidade máxima disponível para esta classe como padrão para geração das amostras, descartando assim algumas imagens da classe não motos e preservando todas as imagens disponíveis da classe motocicletas na geração das amostras.

Consequentemente cada uma das amostras era constituída de 1.442 imagens de veículos sendo 721 imagens de cada uma das duas classes. Para a criação de cada uma das amostras foi realizada uma seleção aleatória dentre as 13.393 imagens de exemplos de não motos para definir quais seriam as 721 imagens escolhidas. Porem a cada criação de uma nova amostra, as imagens eram repostas, possibilitando com que um mesmo exemplos de não motos estivesse em mais de uma amostra. Por este motivo foi realizado uma análise de similaridade entre os elementos da classe não moto entre as amostras e verificou-se que em todas elas existe menos do que cinco por cento de imagens repetidas do tipo não moto. Entretanto todas as 721 imagens do tipo moto são iguais em todas as cinco amostras.

Por sua vez cada uma das 5 amostras foi dividida em dois subconjuntos: (i) treinamento com 1.010 imagens (70% do total de imagens da amostra) sendo 505 de cada uma das duas classes e (ii) teste com 432 imagens (30% do total de imagens da amostra) sendo 216 de cada classe, de maneira similar ao método de validação cruzada *Hold-Out*. A Figura 27 ilustra o particionamento utilizado pelos métodos de validação cruzada adotados.

Figura 27: Esquema das etapas empregadas aos subconjuntos criados.



Fonte: O autor.

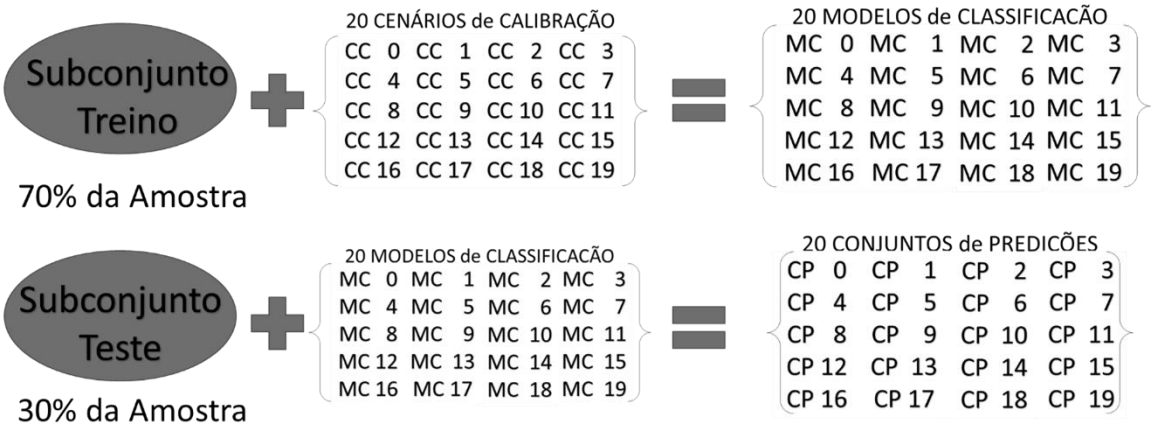
Etapa 2.4. Validação Cruzada

A cada um dos 5 subconjuntos de treinamentos gerados (amostras) foi aplicado os 20 cenários possíveis, configurações dos parâmetros de calibração da função *LinearSVC*, produzindo uma coleção de 20 modelos de classificação distintos por subconjunto de treinamento, ou seja, 100 classificadores diferentes ao total.

Cada uma destas coleções de modelos de classificação fora aplicada ao seu respectivo subconjunto de teste produzindo um aglomerado de 20 cenários de predições distintos por subconjunto de teste, ou seja, 100 cenários de predições diferentes ao total.

Para auxiliar a compreender essa etapa de avaliação a figura 28 exemplifica o processo utilizado à cada uma das 5 amostras.

Figura 28: Esquema das etapas empregadas na realização da validação cruzada.



Fonte: O Autor.

O desempenho do classificador do tipo SVM foi avaliado pela sua exatidão em classificar os exemplos positivos como motocicletas, considerando que os rótulos definidos pelo autor estão corretos.

Para avaliar o desempenho de cada um dos 100 cenários de predição, foram utilizadas as medidas da tabela de contingência (vide Tabela 2) e a AUC. A taxa de verdadeiro positivo (TVP) representa a quantidade de imagens corretamente classificadas como motocicletas; a taxa de falso positivo (TFP) descreve a quantidade errada de imagens classificadas como motocicletas; a taxa de verdadeiro negativo (TVN) representa as imagens corretamente classificadas como não motos; (FAWCETT, 2006). Por sua vez, para avaliar o desempenho do sistema como um todo foi comparado com a contagem manual realizada durante a coleta dos dados além de se verificar se o limitante de tempo será cumprindo, ou seja, todo o processamento de detecção, classificação, rastreio e armazenamento das informações deverá ocorrer no período entre *frames*.

2.4. RESULTADOS

2.4.1. CRIAÇÃO DO BANCO DE IMAGENS DE VEÍCULOS (BIV)

ESTUDO DO TRÁFEGO

Este estudo permitiu definir que a coleta de imagens noturnas do tráfego ficaria muito comprometida uma vez que a luz emitida pelos faróis dos veículos nas proximidades dificulta a visualização e identificação dos veículos presentes no fluxo estudado, optando desta forma que a coleta fosse realizada no período diurno.

COLETA PILOTO

A Tabela 3 apresenta um resumo dos dados obtidos a partir da observação do tráfego realizado na *coleta piloto*.

Tabela 3: Resumo dos dados obtidos na coleta piloto, categorizados por tipo de veículos e sentido das vias de acesso.

VIAS DE ACESSO	PERÍODO	VEÍCULO			
		MOTOCICLETA		AUTOMÓVEL	
		TOTAL	TAXA (veículo / h.)	TOTAL	TAXA (veículo / h.)
Avenida Alberto Benassi (Leste-Oeste)	08:00 – 08:40	19	28,5	99	148,5
	10:45 – 11:40	73	79,6	236	257,4
Avenida Padre Francisco Salles Culturato (Leste-Oeste)	08:00 – 08:40	39	58,5	274	411
	10:45 – 11:40	52	56,7	271	295,6
Avenida Padre Francisco Salles Culturato (Oeste-Leste)	08:00 – 08:40	68	102	336	504
	10:45 – 11:40	108	117,8	443	483,3

Fonte: O autor.

Os dados obtidos pela *coleta piloto* evidenciaram que a Avenida Padre Francisco Salles Culturato no sentido Oeste-Leste no segundo período apresenta volume maior de motocicletas, resultando na escolha deste sentido para a realização da *coleta definitiva* dos vídeos do tráfego.

COLETA DEFINITIVA

A Tabela 4 apresenta um resumo dos dados obtidos na *coleta definitiva* e a Figura 29 ilustra exemplos das imagens obtidas.

Tabela 4: Resumo dos dados coletados através das filmagens do tráfego, do dia 18/04/2017, na Avenida Padre Francisco Salles Culturato no sentido Oeste-Leste.

PERÍODO	VEÍCULO				
	MOTOCICLETA	AUTOMÓVEL	ÔNIBUS	CAMINHÃO	**TOTAL**
10:50 – 11:00	15	80	0	8	**103**
11:00 – 12:00	133	584	4	42	**763**
12:00 – 13:00	146	643	5	19	**813**
13:00 – 14:00	129	634	4	17	**784**
14:00 – 14:10	12	74	1	2	**89**
TOTAL	**435**	**2.015**	**14**	**88**	**2.552**

Fonte: O autor.

Figura 29: Imagens na Avenida Padre Francisco Salles Culturato sentido Oeste-Leste, no dia 18/04/2017. (a) contrária ao fluxo de tráfego e (b) a favor do fluxo.

(a) (b)

Fonte: O autor.

EXTRAÇÃO DE IMAGENS DOS VEÍCULOS DOS VÍDEOS DO TRÁFEGO E SEPARAÇÃO DAS IMAGENS NAS CLASSES DESEJADAS

Após a realização da *coleta definitiva* dos vídeos do tráfego foi possível realizar a extração das imagens dos tipos de veículos presentes. Tendo as subimagens de veículos foi possível realizar uma classificação manual das imagens (separação em arquivos diferentes) de acordo com as classes desejadas: (i) "M", contendo 721 imagens de motos e (ii) "NM", contendo 13.393 imagens de exemplos de qualquer outro tipo de veículo ou parte de veículo. As subimagens que não continham nenhum tipo de veículo ou parte deles também foram classificadas como "F" referenciando ao fundo das imagens, entretanto não foram utilizadas para desenvolver o classificador nesta pesquisa. A Tabela 5 apresenta a quantidade de subimagens obtidas por vídeo analisado e suas respectivas distribuições dentro das classes mencionadas.

Tabela 5: Resultado da etapa de extração e classificação das subimagens.

ID dos Vídeos	Duração dos Vídeos		Nº Subimagens [absoluto]			
	Minutos	Segundos	M	NM	F	**TOTAL**
1	12	42	30	725	2.509	**3.264**
2	12	41	35	1.056	3.277	**4.368**
3	12	41	49	957	3.218	**4.224**
4	12	41	67	787	3.010	**3.864**
5	12	41	45	804	2.919	**3.768**
6	12	48	56	881	3.335	**4.272**
7	12	45	32	954	2.998	**3.984**
8	12	43	80	980	3.668	**4.728**
9	12	44	65	921	3.550	**4.536**
10	12	51	41	832	4.130	**5.003**
11	13	04	28	784	3.172	**3.984**
12	13	03	50	645	2.929	**3.624**
13	12	53	40	863	3.514	**4.417**
14	12	41	40	795	3.172	**4.007**
15	12	41	42	1.011	3.483	**4.536**
16	07	30	21	398	1.597	**2.016**
∑	**3h. 19m. e 30s.**		**721**	**13.393**	**50.481**	**64.595**

Fonte: O autor.

2.4.2. DESENVOLVIMENTO DE UM CLASSIFICADOR DO TIPO SVM

GERAÇÃO DOS VETORES DE CARACTERÍSTICAS (VD)

Uma vez tendo criado o BIV foi possível através de alguns testes simples realizados em um conjunto pequeno porem representativo das imagens obtidas, avaliar o tempo de processamento necessário pelo computador para a produção dos respectivos vetores e a dimensão dos vetores em si e assim analisar dentre os descritores mais usuais observados na literatura especializada como LBP e SURF, quais destas técnicas seriam capazes de produzir um vetor de características apropriado levando em consideração os objetivos desta pesquisa, . Sendo assim optou-se por utilizar os histogramas em escala de cinza fornecidos pelo descritor LBP.

DEFINIÇÃO DE CENÁRIOS A SE UTILIZAR NA CALIBRAÇÃO DOS PARÂMETROS DO CLASSIFICADOR SVM

Após a análise de sensibilidade dos parâmetros verificou-se que apenas seis dos parâmetros produziam uma mudança significativa nos resultados obtidos sendo assim um conjunto de 20 cenários foi definido, contento conjuntos de parâmetros variados para calibrar e determinar qual o melhor conjunto de parâmetros para o classificador proposto nesta pesquisa. A Tabela 6 descreve os 20 cenários e seus respectivos parâmetros da função.

Tabela 6: Conjunto de configurações dos parâmetros da função LinearSVC utilizados para o teste de calibração.

Cenários / Parâmetros	C	Dual	Loss	Multi_Class	Penalty	Tol
CC 0	1	True	squared_hinge	ovr	I2	0.0001
CC 1	1	True	squared_hinge	crammer_singer	I2	0.0001
CC 2	1	True	hinge	ovr	I2	0.0001
CC 3	1	False	squared_hinge	ovr	I2	0.0001
CC 4	1	False	squared_hinge	ovr	I1	0.0001
CC 5	150	True	squared_hinge	ovr	I2	0.0001
CC 6	150	True	squared_hinge	crammer_singer	I2	0.0001
CC 7	150	True	hinge	ovr	I2	0.0001
CC 8	150	False	squared_hinge	ovr	I2	0.0001
CC 9	150	False	squared_hinge	ovr	I1	0.0001
CC 10	1	True	squared_hinge	ovr	I2	0.01
CC 11	1	True	squared_hinge	crammer_singer	I2	0.01
CC 12	1	True	hinge	ovr	I2	0.01
CC 13	1	False	squared_hinge	ovr	I2	0.01
CC 14	1	False	squared_hinge	ovr	I1	0.01
CC 15	150	True	squared_hinge	ovr	I2	0.01
CC 16	150	True	squared_hinge	crammer_singer	I2	0.01
CC 17	150	True	hinge	ovr	I2	0.01
CC 18	150	False	squared_hinge	ovr	I2	0.01
CC 19	150	False	squared_hinge	ovr	I1	0.01

Os parâmetros em vermelho representam a configuração diferente do padrão.
Fonte: O autor.

DEFINIÇÃO DAS AMOSTRAS E SUBCONJUNTOS DE TREINAMENTO E SUBCONJUNTOS DE TESTE

Como a quantidade de imagens disponíveis no BIV apresenta uma desigualdade para cada uma das classes desejadas (721 imagens para a classe "M" e 13.393 imagens de exemplos para a classe "NM") e o intuito da pesquisa é identificar imagens de veículos do tipo motocicletas, optou-se por descartar algumas imagens da classe não motos e preservando todas as imagens disponíveis da classe motocicletas na geração das amostras.

Consequentemente cada uma das cinco amostras criadas foram constituídas de 1.442 imagens de veículos sendo 721 imagens de cada uma das duas classes. Para a criação de cada uma das amostras foi realizada uma seleção aleatória dentre todas as imagens disponíveis de exemplos de não motos para definir quais seriam as imagens escolhidas.

Por sua vez cada uma das cinco amostras foi dividida em dois subconjuntos: (i) treinamento com 1.010 imagens (70% do total de imagens da amostra) sendo 505 de cada uma das duas classes e (ii) teste com 432 imagens (30% do total de imagens da amostra) sendo 216 de cada classe, de maneira similar ao método de validação cruzada *Hold-Out*.

VALIDAÇÃO CRUZADA

Com a análise da tabela de contingência (vide Tabela 2, página 63), foi possível obter uma média para cada tipo de teste de calibração (5 amostras para cada uma das 20 configurações dos parâmetros de calibração utilizados), e verificar que apenas os parâmetros *C* (penalidade por erro) e *Tol* (tolerância para o critério de parada) mostraram uma mudança significativa nas previsões obtidas. A Tabela 7 e a Tabela 8 ilustram as tabelas de contingência analisadas. A Tabela 9 apresenta a média das tabelas de contingência analisadas, apresentando distribuição das imagens para cada um dos cenários analisados.

Tabela 7: Exemplo de uma das tabelas de contingência obtidas. Quantidade absoluta (número de imagens).

	Conjunto de Predição 0					
Amostra	VP	FN	TP	FP	VN	TN
A1	197	19	216	70	146	216
A2	203	13	216	46	170	216
A3	194	22	216	61	155	216
A4	197	19	216	59	157	216
A5	198	18	216	59	157	216
Média	197.8	18.2	216	59	157	216

Fonte: O autor.

Tabela 8: Exemplo de uma das tabelas de contingência obtidas. Métricas utilizadas para avaliar o desempenho dos 20 Conjuntos de Cenários analisados.

Conjunto de Predição 0						
Amostra	TVP	TFP	PRECISÃO	EXATIDÃO	TVN	TVP
A1	0,91	0,32	0,74	0,79	0,68	0,91
A2	0,94	0,21	0,82	0,86	0,79	0,94
A3	0,90	0,28	0,76	0,81	0,72	0,90
A4	0,91	0,27	0,77	0,82	0,73	0,91
A5	0,92	0,27	0,77	0,82	0,73	0,92
Média	0,916	0,273	0,771	0,821	0,727	0,916

Fonte: O autor.

Tabela 9: Média da quantidade absoluta (número de imagens) das cinco amostras para cada um dos 20 cenários analisados

Média das 5 Amostras						
Conjunto Predição	VP	FN	TP	FP	VN	TN
0	197,8	18,2	216,0	59,0	157,0	216,0
1	201,6	14,4	216,0	96,2	119,8	216,0
2	201,6	14,4	216,0	96,2	119,8	216,0
3	197,8	18,2	216,0	59,0	157,0	216,0
4	200,2	15,8	216,0	34,0	182,0	216,0
5	202,4	13,6	216,0	21,2	194,8	216,0
6	202,8	13,2	216,0	22,2	193,8	216,0
7	200,8	15,2	216,0	23,4	192,6	216,0
8	201,6	14,4	216,0	21,0	195,0	216,0
9	198,0	18,0	216,0	18,4	197,6	216,0
10	197,8	18,2	216,0	59,0	157,0	216,0
11	201,6	14,4	216,0	96,0	120,0	216,0
12	204,2	11,8	216,0	116,0	100,0	216,0
13	197,8	18,2	216,0	59,0	157,0	216,0
14	199,6	16,4	216,0	33,8	182,2	216,0
15	200,6	15,4	216,0	19,8	196,2	216,0
16	202,8	13,2	216,0	22,0	194,0	216,0
17	201,6	14,4	216,0	25,8	190,2	216,0
18	201,8	14,2	216,0	21,2	194,8	216,0
19	199,2	16,8	216,0	19,2	196,8	216,0
Variação	5,0	6,4	-	75,2	97,6	-

Os valores em vermelho representam os limites superiores e inferiores para cada tipo de predição realizada.
Fonte: O autor.

Na Tabela 10. Os valores destacados representam as taxas de verdadeiro positivo (TVP) mais altas e as taxas de falso positivo (TFP) mais baixas.

Tabela 10: Médias das taxas de verdadeiro positivo e falso positivo das 5 amostras.

Calibração		TOL = $1x10^{-4}$		Calibração	TOL = $1x10^{-2}$	
		TVP	TFP		TVP	TFP
C = 1	**CP 0**	0,916	0,273	**CP 10**	0,916	0,272
	CP 1	0,933	0,445	**CP 11**	0,933	0,444
	CP 2	0,945	0,538	**CP 12**	0,945	0,537
	CP 3	0,916	0,273	**CP 13**	0,916	0,273
	CP 4	0,927	0,157	**CP 14**	0,924	0,156
C = 150	**CP 5**	0,937	0,098	**CP 15**	0,929	0,092
	CP 6	0,939	0,103	**CP 16**	0,939	0,102
	CP 7	0,930	0,108	**CP 17**	0,933	0,119
	CP 8	0,933	0,097	**CP 18**	0,934	0,098
	CP 9	0,917	0,085	**CP 19**	0,922	0,089

Fonte: O autor.

Pela a análise da tabela de contingência dos desempenhos dos 5 subconjuntos testes é possível verificar que para os 2 principais parâmetros de calibração da função em questão a variação apresentada nos testes é muito pequena o que evidencia a grande dificuldade em estimar os melhores valores dos parâmetros de um algoritmo, reforçando as informações da literatura (MINH; SANO; MATSUMOTO, 2012; ARRÓSPIDE; SALGADO, 2014; KIM et al., 2015; LUO et al., 2018). Entretanto, a literatura pesquisada sugere que o classificador pode ser considerado de bom desempenho, quando em sua maioria os valores de taxa de verdadeiro positivo (TVP) são superiores a 0,9 e taxa de falso positivo (TFP) estão próximos a 0,1.

Para que a escolha do melhor conjunto de parâmetros a ser adotada utilizou-se também a métrica AUC para auxiliar na escolha da melhor calibração do classificador. Na Tabela 11 é mostrada a média de desempenho de cada um dos subconjuntos teste tanto para a métrica AUC. Para a AUC o melhor desempenho obtido foi 0,96, o que caracteriza o desempenho do classificador como ótimo e sua predição geral não possui comportamento aleatório, segundo FAWCETT, (2006) e PRATI; BATISTA; MONARD (2008).

Tabela 11: Desempenho da métrica AUC das 5 amostras para cada uma das 20 configurações de parâmetros de calibração dos conjuntos testes e suas respectivas métricas usuais.

	AMOSTRA 1	**AMOSTRA 2**	**AMOSTRA 3**	**AMOSTRA 4**	**AMOSTRA 5**	**MÉDIA_ AUC**	TVP	TFP	PRECISÃO	EXATIDÃO	TVN
AUC_CP 0	0,894	0,936	0,892	0,896	0,896	0,903	0,916	0,273	0,771	0,821	0,727
AUC_CP 1	0,844	0,898	0,817	0,848	0,855	0,852	0,933	0,445	0,677	0,744	0,555
AUC_CP 2	0,807	0,867	0,770	0,804	0,821	0,814	**0,945**	0,538	0,637	0,704	0,462
AUC_CP 3	0,894	0,936	0,892	0,896	0,896	0,903	0,916	0,273	0,771	0,821	0,727
AUC_CP 4	0,910	0,962	0,955	0,936	0,938	0,940	0,927	0,157	0,855	0,885	0,843
AUC_CP 5	0,943	0,973	0,975	0,963	0,955	**0,962**	0,937	0,098	0,906	**0,919**	0,902
AUC_CP 6	0,945	0,974	0,974	0,962	0,954	0,962	0,939	0,103	0,902	0,918	0,897
AUC_CP 7	0,942	0,972	0,970	0,958	0,951	0,959	0,930	0,108	0,896	0,911	0,892
AUC_CP 8	0,943	0,973	0,975	0,963	0,955	**0,962**	0,933	0,097	0,906	0,918	0,903
AUC_CP 9	0,947	0,970	0,974	0,966	0,952	0,962	0,917	**0,085**	0,916	0,916	0,915
AUC_CP 10	0,894	0,936	0,892	0,896	0,896	0,903	0,916	0,272	0,771	0,822	0,728
AUC_CP 11	0,844	0,898	0,817	0,848	0,855	0,852	0,933	0,444	0,678	0,744	0,556
AUC_CP 12	0,807	0,868	0,770	0,804	0,821	0,814	**0,945**	0,537	0,638	0,704	0,463
AUC_CP 13	0,894	0,936	0,892	0,896	0,896	0,903	0,916	0,273	0,771	0,821	0,727
AUC_CP 14	0,924	0,962	0,954	0,939	0,938	0,944	0,924	0,156	0,855	0,884	0,844
AUC_CP 15	0,943	0,973	0,975	0,963	0,955	0,962	0,929	0,092	**0,911**	**0,919**	**0,908**
AUC_CP 16	0,945	0,974	0,974	0,962	0,954	0,962	0,939	0,102	0,902	**0,919**	0,898
AUC_CP 17	0,942	0,972	0,970	0,958	0,951	0,959	0,933	0,119	0,887	0,907	0,881
AUC_CP 18	0,944	0,973	0,975	0,963	0,955	**0,962**	0,934	0,098	0,905	0,918	0,902
AUC_CP 19	0,949	0,971	0,975	0,963	0,953	**0,962**	0,922	0,089	0,913	0,917	0,911
MAX	0,949	0,974	0,975	0,966	0,955						
MIN	0,807	0,867	0,770	0,804	0,821						

2.5. CONCLUSÕES

2.5.1. CRIAÇÃO DO BANCO DE IMAGENS DE VEÍCULOS (BIV)

Uma vez que o principal objetivo desta seção era a criação de um banco de imagens de veículos que possuísse imagens de motocicletas e imagens dos demais veículos e da paisagem de fundo é possível constatar que o trabalho atingiu o objetivo proposto dado que o banco de dados concebido é constituído de (i) 721 imagens de motos (classe "M"), (ii) 13.393 imagens de exemplos de qualquer outro tipo de veículo ou parte de veículo (classe "NM") e (iii) 50.481 imagens de exemplos da paisagem de fundo (classe "F").

2.5.2. TREINAMENTO DE UM CLASSIFICADOR DO TIPO SVM

Devido ao principal objetivo desta seção ser a calibração dos parâmetros de um classificador do tipo SVM utilizando o banco de imagens de veículos (BIV) devidamente rotulados possibilitando que este seja capaz de identificar veículos do tipo motocicletas em imagens extraídas de vídeos do tráfego é concebível afirmar que esta pesquisa alcançou o objetivo proposto visto que as análises de todas as métricas obtidas (tabela de contingência e métrica AUC), evidenciou o desenvolvimento de mais de um modelo de classificação utilizando vetores de características criados a partir de histogramas em escala de cinza pela função LBP (local_binary_pattern) que no geral apresentou desempenho muito bom, com valores de precisão e exatidão superiores a 0,91. Segundo a literatura (FAWCETT, 2006; PRATI; BATISTA; MONARD, 2008; MINH; SANO; MATSUMOTO, 2012; SILVA et al., 2013b; ARRÓSPIDE; SALGADO, 2014; LUO et al., 2018; SILVA; AIRES; VERAS, 2018), um classificador com valores de precisão superior a 0,94 e valores de exatidão superiores a 0,97 configuram um classificador de desempenho excelente. Por consequência, o Cenário de Classificação CC 19 foi considerado o mais indicado para ser utilizado como calibração para o Classificador de Motocicletas desenvolvido, pois apresentou uma das melhores métricas tanto para AUC (0,96), como para TVP (0,92), TFP (0,09), TVN (0,91), Precisão (0,91) e Exatidão (0,92).

3. ASSOCIAÇÃO DO MÓDULO DE IDENTIFICAÇÃO AUTOMÁTICA DO COMPORTAMENTO DO TRÁFEGO COM O MÓDULO DE CLASSIFICAÇÃO DE MOTOCICLETAS

Este *Capítulo 3* está organizado em cinco seções centrais que possuem objetivos, materiais, métodos e resultados próprios. O objetivo central deste capítulo é utilizar técnicas de visão computacional para o monitoramento automático de tráfego veicular em trechos urbanos de forma a identificar e classificar os tipos de veículos nos vídeos de monitoramento do tráfego, possibilitando a análise do comportamento dos condutores, mais precisamente, dos motociclistas brasileiros.

3.1. CONTEXTUALIZAÇÃO

Neste capítulo são apresentados os tipos de sensores de monitoramento do trânsito mais utilizados e os aspectos sobre a utilização da motocicleta como modo de transporte em áreas urbanas.

3.1.1. SENSORES DE MONITORAMENTO DO TRÁFEGO

Historicamente, o monitoramento do tráfego é desempenhado por duas abordagens: (i) Tecnologias baseadas na malha viária com aparelhos instalados próximos a via como sensores de radar ultrassónicos, sensores acústicos, sensores de micro-ondas, detectores de imagens de vídeo, detectores infravermelhos, ou aparelhos instalados no leito viário como detectores piezoeléctricos, tubos pneumáticos, detectores do tipo laço indutivo ou magnetômetros e (ii) Tecnologias baseadas nos veículos como sistemas automáticos de identificação de veículos, sistemas automáticos de localização de veículo, veículos com sonda ou simplesmente a utilização de smartphones para obter a localização dos veículos (KLEIN, 2001; KLEIN; MILLS; GIBSON, 2006; MORGUL et al., 2014).

Estas duas tecnologias de vigilância têm várias características positivas e negativas. Ainda que as tecnologias baseadas em veículos apresentem uma medição mais precisa dos tempos de viagem real, sua utilização não acontece em larga escala essencialmente pelo elevado custos de implementação (MORGUL et al., 2014).

Os sensores de tráfego podem ser classificados em três grupos: detectores de ponto único, detectores de multipontos e sensores móveis de vasta abrangência. A coleta de dados de tráfego de vasta abrangência usualmente realiza a vigilância do tráfego instalando diversas unidades ao longo da extremidade da via e

devido a este fato esse tipo de sensor apresenta custos maiores de implementação e manutenção (KLEIN, 2001; KLEIN; MILLS; GIBSON, 2006; MORGUL et al., 2014).

As alternativas de baixo custo, como sensores sem fio, transponders de identificação por rádio frequência e Bluetooth normalmente apresentam problemas devido a dificuldade em se identificar os veículos que se encontram em movimento e também ao escasso número de veículos que possuem essa tecnologia instalada e portanto disponibilizam esse tipo de dados em rotas de interesse (KLEIN; MILLS; GIBSON, 2006; MORGUL et al., 2014).

Veículos de sonda equipados com aparelhos de GPS (*Global Positioning System* – Sistema de Posicionamento Global) como ônibus municipais, táxis, ou veículos comerciais, disponibilizam uma grande quantidade de dados de elevada qualidade para monitoramento de tráfego e com custos adicionais quase inexistentes. Embora alguns destes dados coletados não sejam disponibilizados publicamente por razões de segurança ou confidencialidade, é possível encontrar fontes de dados com serviços abertos que fornecem informações de viagem em tempo real, como o NextBus[8] (MORGUL et al., 2014)

3.1.2. USO DE MOTOCICLETAS EM PERÍMETRO URBANO

Em qualquer área urbana, seja mundialmente ou principalmente em território nacional, é possível perceber que o modo motocicleta quase sempre apresenta uma participação significativa na composição da frota de veículos, uma vez que este tipo de veículo apresenta um fácil deslocamento urbano na cidade. O aumento de motocicletas na frota pode ser explicada principalmente por: (i) sua agilidade no trânsito urbano, pois permite ao seu condutor se deslocar na malha viária de forma fácil e ágil, devido às suas configurações físicas (dimensões e peso menores do que os demais veículos); (ii) por ser um veículo porta-a-porta, permitindo um controle mais eficiente do tempo de viagem, pois não há necessidade da utilização de nenhum outro modo de transporte complementar e (iii) por apresentar um reduzido valor tanto na aquisição quanto na sua manutenção (FERRAZ; TORRES, 2004). Entretanto, as mesmas características supracitadas desse modo de transporte, também o tornam um modo de alto risco de acidentes, seja pela fragilidade na proteção do condutor e, principalmente, pela indisciplina de condução veicular, tanto do motociclista nas leis de trânsito como do desrespeito dos demais veículos para com os motociclistas (FERRAZ *et al*., 2012).

Babu; Vortisch; Mathew (2015) destacam o número limitado de pesquisas com foco de estudo no comportamento de veículos leves de duas rodas e como esse comportamento interfere nos parâmetros de desempenho do tráfego, tanto em fluxo misto como em fluxo heterogêneo. Em países onde a participação de motocicletas é bem substancial, estudos tentam preencher essa lacuna do conhecimento (KOV; YAI,

8 NextBus disponível no site <http://nexbus.com>

2009; MUNIGETY; VICRAMAN; MATHEW, 2014; NGUYEN, 2016). A participação de motocicletas em algumas cidades é tamanha que o termo "cidade dependente de motocicleta" (*Motorcycle Dependent City – MDC*) vem sendo aplicado para determinar os níveis de dependências, definidos a partir de três indicadores: (i) número de proprietários de veículos em geral, (ii) disponibilidade de transportes alternativos e (iii) o número de motocicletas (NGUYEN, 2016).

Nguyen; Hanaoka; Kawaski(2014) e Nguyen (2016) retratam o cenário de *Ho Chi Minh City*, uma cidade Vietnamita cuja frota é constituída por 90% de motocicletas. Nguyen; Hanaoka; Kawasaki (2014) desenvolveram um modelo integrado para avaliar o conflito de tráfego de motocicletas em condições congestionadas. O modelo proposto inclui um conceito de espaço de segurança para descrever os movimentos irregulares normalmente realizados por motociclistas em situações de múltiplas e consecutivas ultrapassagens, além de fornecer novos recursos para a avaliação de conflitos de trânsito. Alguns dos parâmetros presentes no modelo utilizavam dados da aceleração e desaceleração para determinar quais eram os critérios utilizados pelos motociclistas para elegerem um veículo líder (veículo cujo seu comportamento e trajetória seria reproduzido pelos demais), sendo que os dados utilizados para a calibração desses parâmetros foram extraídos de vídeo de monitoramento em dois segmentos rodoviários em *Ho Chi Minh City*.

Os resultados obtidos na pesquisa salientaram que o aumento da densidade da corrente de fluxo no cenário estudado é proporcional aos conflitos associados à frenagem súbita. Isso ocorre em situação de fluxo normal quando o condutor de motocicletas, para manter seu padrão de velocidade, reage e é influenciado a somente um veículo por vez. Porém, em situação de fluxo congestionado o mesmo condutor por não alterar seu padrão de velocidade é forçado a alterar seu padrão de deslocamento, sendo obrigado a interagir de forma mais dinâmica com diversos veículos o que ocasiona uma redução em seu tempo de reação.

Por outro aspecto, a pesquisa de NGUYEN (2016) descreve os resultados de uma análise abrangente sobre alguns dos parâmetros do tráfego como a taxa de saturação em intersecções sinalizadas em *MDC*, considerando as condições específicas do comportamento do condutor e do fluxo de tráfego. Essa pesquisa indicou que a descarga da fila ocasionada pela interseção sinalizada, para o tráfego específico de motocicletas, tem comportamento diferente do encontrado na literatura. Além disso, descreveram que a trajetória escolhida pelos condutores, as configurações físicas das interseções e a composição da frota nas aproximações apresentam impactos diferente na taxa de saturação do tráfego. Dessa forma, através de vídeos de monitoramento do fluxo do tráfego e de um método de contagem para extração dos dados, o estudo foi capaz de produzir um modelo de taxa de fluxo de saturação usando um método de regressão, além de fornecer um procedimento próprio para calcular a taxa de saturação em condições de tráfego

homogêneo de motocicletas introduzindo desta forma o termo “unidade de motocicleta” (*MCU – Motorcycle Unit*).

De forma similar, MUNIGETY; VICRAMAN; MATHEW (2014) e BABU; VORTISCH; MATHEW (2015) retrataram o cenário da Índia, um país no qual o tráfego também é considerado heterogêneo apresentando vários tipos de veículos na composição de sua frota, além da pouca disciplina no trânsito.. MUNIGETY; VICRAMAN; MATHEW (2014) elucidam o desenvolvimento de uma ferramenta semiautomática, um extrator de dados de tráfego (*Traffic Data Extractor – TDE*) na plataforma *Visual Basic*, para extrair os detalhes dos movimentos dos veículos. O método empregado por essa ferramenta utiliza técnicas de calibração de câmera baseada em ponto de fuga para obter as coordenadas do mundo real a partir de coordenadas de imagem bidimensional extraindo assim dados sobre os comportamentos dos veículos a partir de vídeos de monitoramento do tráfego. As principais vantagens dessa ferramenta são: (i) a capacidade de monitorar a trajetória de diversos tipos de veículos simultaneamente, obtendo vários parâmetros mesmo em um tráfego heterogêneo, (ii) a facilidade na utilização e a grande compatibilidade com os arquivos de vídeos em diversos formatos. Entretanto, essa ferramenta é semiautomática, uma vez que a identificação e seleção dos veículos a serem monitorados deve ocorrer por parte do utilizador, em outras palavras, essa ferramenta é capaz de rastrear os veículos e extrair seus dados de comportamento, porém não é capaz de detectar e nem de classificar os mesmos.

Por sua vez, BABU; VORTISCH; MATHEW (2015) propuseram um modelo de simulação microscópico que permite a reprodução do movimento de motocicletas em um ambiente de tráfego heterogêneo. O desempenho do modelo é avaliado usando a comparação de trajetória individual entre simulação e observação em campo, obtida pela gravação do tráfego de Mumbai, Índia, e a capacidade de simular qualitativamente o comportamento naturalista de motocicletas através de um cenário teste. As teorias do modelo de força social (*Social Force Approach – SFA*), normalmente utilizadas para representar o comportamento de pedestres e o modelo de motorista inteligente (*Intelligent Driver Model – IDM*) foram adaptadas e empregadas para o desenvolvimento do modelo em questão, e que pelos resultados apontados na pesquisa foram capazes de representar visualmente o comportamento das motocicletas estudadas.

KOV e YAI (2009) ressaltaram o cenário observado em Phnom Penh, capital do Camboja, onde 70% da frota é constituída por motocicletas e os condutores apresentam o mesmo padrão de indisciplina no trânsito retratado também por MUNIGETY; VICRAMAN; MATHEW (2014) e BABU; VORTISCH; MATHEW (2015). Por se tratar de uma *MDC*, a pesquisa investigou primeiramente o desempenho do tráfego em via urbana dominada por motocicletas e como respondem a diferentes proporções de veículos leves na composição deste tráfego. Para a primeira parte da pesquisa, os dados foram coletados usando imagens de vídeos do tráfego e na etapa seguinte foram realizadas entrevistas com os condutores, buscando analisar a percepção dos motociclistas sobre a questão da segurança do trânsito e o grau em que

esses envolvem-se em acidentes de trânsito. Dessa forma, a pesquisa revelou que o aumento da participação de automóveis no fluxo de tráfego reduz consideravelmente a velocidade da corrente e o grau de liberdade de movimento lateral das motocicletas. Essas reduções alteram o comportamento dos motociclistas, acarretando concentração de motos nas faixas internas da via.

Sob a ótica de segurança no trânsito, SILVA *et al*. (2013a) ressaltaram que o número de acidentes com motocicletas tem aumentado em várias partes do mundo ao longo dos anos, sendo que o principal equipamento de segurança dos motociclistas – o capacete – é muitas vezes negligenciado. Sendo assim, os autores desenvolveram um método automático para detectar e classificar motocicletas em vias públicas no Brasil, além de um sistema automático que verifica se os usuários estão utilizando ou não o capacete. O método desenvolvido utiliza uma técnica de segmentação de objetos para identificar os veículos, um descritor híbrido (conjunto de descritores composto pelo *Local Binary Pattern* e pelo *Histogram of Oriented Gradients*), a técnica Transformada Circular de *Hough* para reconhecer as características dos veículos detectados e também utiliza o classificador *Support Vector Machines* para conseguir classificar os veículos em "motocicletas" ou "não-motocicletas" e posteriormente classificar as motocicletas em "com-capacete" ou "sem-capacete".

Através dessas pesquisas é possível evidenciar que o modo de transporte motocicleta, além de apresentar um grande potencial para a mobilidade na rede urbana, requer medidas que permitam à fiscalização e monitoramento do comportamento de motociclistas. Assim, um estudo aprofundado sobre o comportamento dos condutores de motocicletas e de como ocorre a interação deste modo de transporte com os demais veículos na corrente de tráfego viabiliza compreender como esse modo de transporte em particular se aplica dentro do contexto de segurança viária. A compreensão do comportamento dos motociclistas também permite a sua modelagem e representação em modelos de simulação de tráfego, permitindo amplo planejamento de medidas e soluções para melhorar a complexa coexistência entre os veículos no tráfego urbano.

3.2. MATERIAIS

O trabalho desenvolvido em questão está inserido em uma linha de pesquisa sobre Sistemas Inteligentes de Transportes (*ITS*), desenvolvida no Departamento de Engenharia de Transportes da Escola de Engenharia de São Carlos (EESC-USP) O propósito desta pesquisa é utilizar técnicas de visão computacional para o monitoramento automático de tráfego veicular em trechos urbanos de forma a identificar e classificar os tipos de veículos, possibilitando à análise do comportamento dos condutores, mais precisamente, dos motociclistas brasileiros. Sendo assim o grupo de pesquisa que desenvolve o estudo de *ITS* por sua vez já possuía a disposição um sistema de Identificação Automática do

Comportamento do Tráfego a Partir de Imagens de Vídeo (MARCOMINI, 2018), que fora desenvolvido em trabalhos anteriores. Possibilitando assim a utilização deste módulo de Identificação Automática do Comportamento do Tráfego - IACT em conjunto com o classificador de motocicletas desenvolvido nas partes anteriores desta pesquisa permitindo a extração de informações sobre o comportamento dos veículos, principalmente das motocicletas, foco desta pesquisa.

De forma geral, os materiais utilizados para a realização desta pesquisa foram: (i) um computador utilizado para o processamento dos algoritmos; (ii) o sistema de Identificação Automática do Comportamento do Tráfego a Partir de Imagens de Vídeo proposto por (MARCOMINI, 2018); (iii) o Classificador de Motocicletas desenvolvido no *Capítulo 2* desta pesquisa e (iv) os 16 vídeos do tráfego coletados e utilizados para produzir o classificador de motocicletas do *Capítulo 2*.

O computador utilizado foi um Core i7, com processador de 2,40 GHz, memória RAM de 8 GB e sistema operacional Windows 8. Os códigos, tanto do sistema proposto por MARCOMINI (2018) como do classificador de motocicletas desenvolvido, foram escritos em linguagem *Python 2*, utilizando a plataforma Anaconda versão 2.7.13[9], em conjunto com as bibliotecas de funções de processamento de imagens *OpenCV*[10] versão 3.0.0, *scikit-image*[11] e a biblioteca de funções de aprendizado de máquinas *scikit-learn 12*. Os vídeos do tráfego possuem resolução Full HD (1920 x 1080) com 30 fps.

3.3. MÉTODO PROPOSTO

A estrutura deste *Capítulo 2* está dividida em quatro seções principais: 1) Extração de dados veiculares, 2) Obtenção dos dados veiculares, 3) Classificação dos veículos e 4) Obtenção da classificação dos veículos.

3.3.1. EXTRAÇÃO DE DADOS VEICULARES

Baseado no texto de dissertação de mestrado de MARCOMINI (2018) e de maneira similar ao estudado e apresentado por CUNHA (2013) o módulo IACT é capaz de detectar e separar os veículos (objetos em movimento) do fundo (objetos estáticos) presentes no vídeo. Esse processo de diferenciar ou separar os objetos de interesse do plano de fundo ocorre pela subtração entre o *frame* atual e um modelo do plano de fundo, sendo que a geração desse modelo de plano de fundo tem um grande impacto para melhorar a segmentação dos veículos presentes na corrente de tráfego.

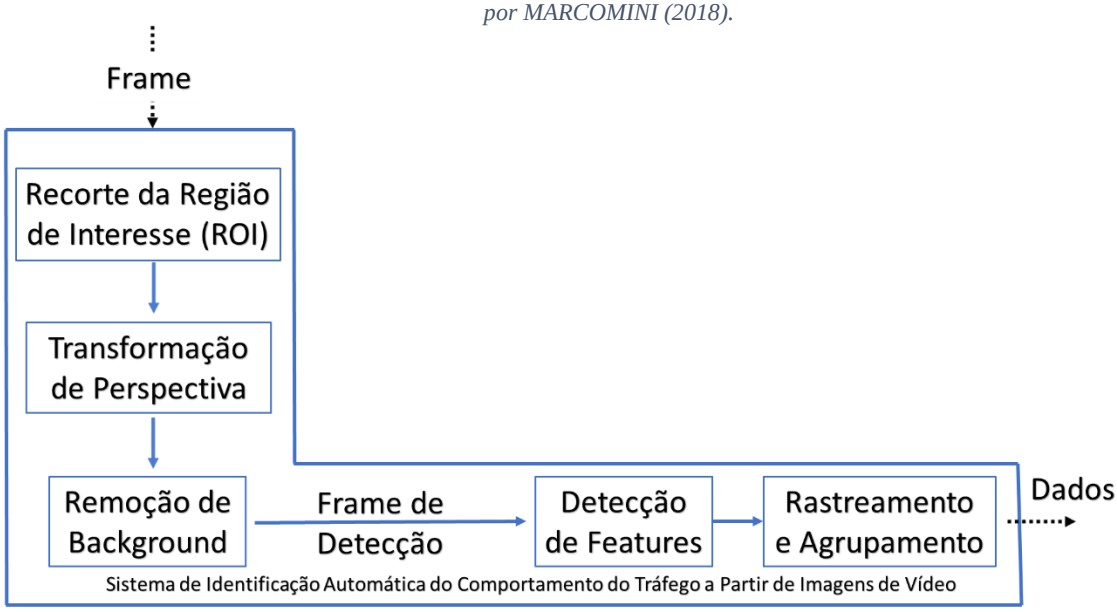
---

9 Plataforma Anaconda disponível no site <https://www.continuum.io/downloads>
10 OpenCV disponível no site <http://opencv.org>
11 Scikit-image disponível no site <https://scikit-image.org/>
12 Scikt-learn disponível no site <http://scikit-learn.org/stable/install.html>

Uma vez tendo os veículos identificados e separados do plano de fundo o sistema inicia a etapa de rastreamento (*tracking*) presente no algoritmo que deverá avaliar quais dos pixels que compõem os veículos em geral (objetos em movimento) serão importantes para que seja possível relacionar o mesmo veículo em diferentes *frames* sucessivos, ou seja, permitir que uma associação temporal seja realizada para um mesmo veículo durante todo o seu percurso. Essa associação temporal permite que o algoritmo colete e armazene cada uma das coordenadas de cada um dos veículos ao longo de cada um dos momentos, *frames*, de cada trajetória realizada, disponibilizando assim dados veiculares como posição ao longo do tempo (diagrama espaço tempo) e dados de velocidade, aceleração etc.

A Figura 30 mostra as principais etapas do módulo IACT proposto por MARCOMINI (2018).

Figura 30: Fluxograma das etapas do sistema de Identificação Automática do Comportamento do Tráfego (IACT) proposto por MARCOMINI (2018).

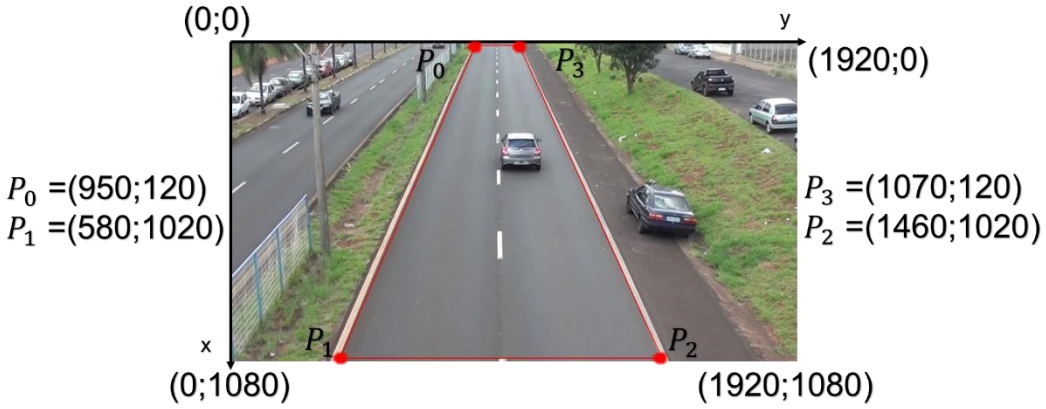


Fonte: Adaptado de MARCOMINI (2018).

Etapa 1.1. REGIÃO DE INTERESSE (ROI)

Ao se utilizar o sistema de MARCOMINI (2018), o usuário inicialmente delimita uma região de interesse (*Region of Interest* - ROI). Essa região é demarcada logo no primeiro *frame* da sequência de imagens, antes que haja movimentação dos veículos. Vale ressaltar algumas particularidades nessa etapa: (i) as coordenadas dos vértices da ROI são baseadas a partir do *frame* original do vídeo em questão; (ii) a ROI deve evitar regiões as quais a possibilidade da presença de veículos é bem baixa, como o caso do canteiro central e regiões vizinhas ao fluxo de veículos. Entretanto dependendo da perspectiva da filmagem é aconselhável que a ROI contenha parte do acostamento para facilitar a delimitação da região. A Figura 31 elucida os instantes iniciais da utilização do programa.

Figura 31: Determinação da ROI utilizando o sistema proposto por MARCOMINI (2018) e um dos vídeos do tráfego coletados no Capítulo 2 deste trabalho.



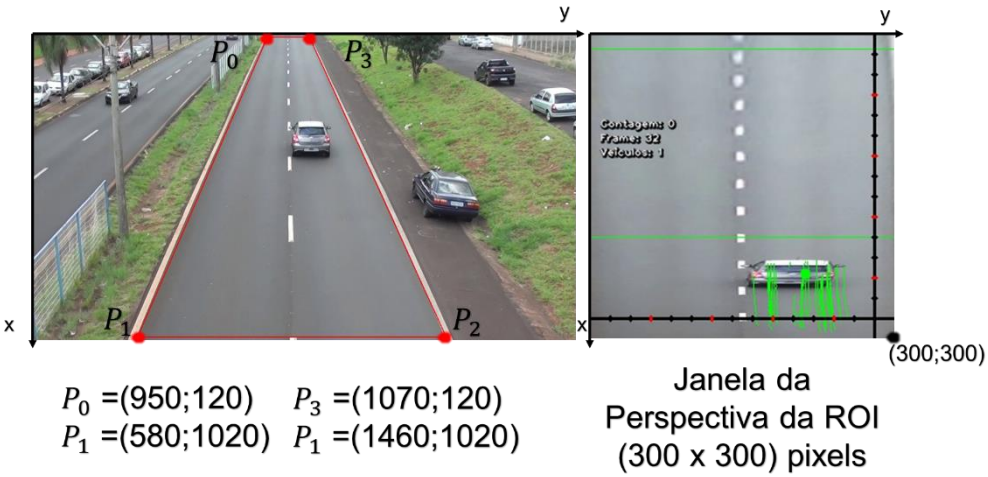
Region of Interest - *ROI determinada pelos pontos P_0 , P_1, P_2 e P_3.*
Fonte: O autor.

Nesta etapa é imprescindível que se conheça algumas dimensões no mundo real, contidas na ROI para que nesta nova perspectiva gerada seja possível relacionar a distância percorrida pelos veículos nesta nova janela em perspectiva (distância em pixels) com a distância percorrida no mundo real (distância em metros) possibilitando o cálculo de velocidade. Tanto as coordenadas dos quatro pontos da ROI como o valor da relação "distância mundo real / distância na janela em perspectiva" [metros / pixels] devem ser inseridas pelo usuário dentro do código fonte em *Python*. É importante salientar que todas as próximas etapas do sistema acontecem somente na ROI, o que melhora o desempenho do programa como um todo, uma vez, que todas as técnicas utilizadas daqui por diante acontecem em uma região restrita e menor do que o *frame* todo disponível.

Etapa 1.2. TRANSFORMAÇÃO DE PERSPECTIVA

Uma vez delimitada a ROI, o programa realiza a transformação de perspectiva, Figura 32, onde os pixels dentro da ROI são mapeados para um quadrado com dimensões de 300x300 pixels.

Figura 32: Criação da Janela em Perspectiva a partir da ROI (retângulo em vermelho).

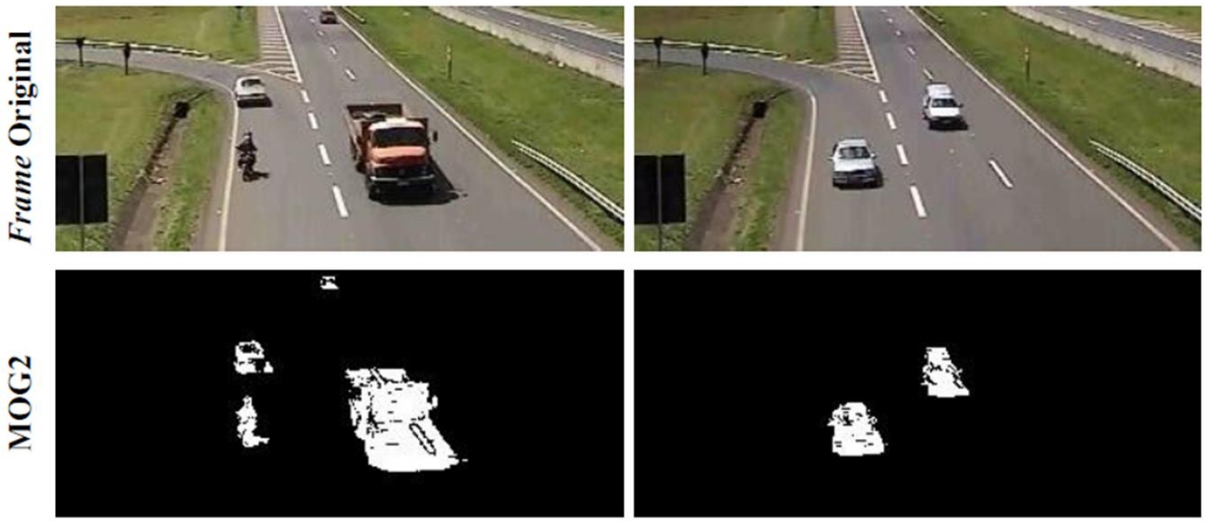


Fonte: O autor.

Etapa 1.3. REMOÇÃO DE FUNDO

Após a geração da perspectiva, inicia-se a etapa de remoção do fundo (*background*) e a detecção e separação dos veículos presentes no vídeo. Com a utilização do algoritmo MOG2 e de uma função *threshold* que aplica uma limiarização binaria para que os pixels assumam valores brancos ou pretos é possível segmentar, quase que sem ruído, as regiões dos veículos, das regiões de fundo da imagem. Esse processo de limiarização ocorre transformando os pixels com valores menores que o limite para preto, "0" e os demais são convertidos para branco, "255" A Figura 32 elucida o resultado obtido dessa etapa de remoção de fundo.

Figura 33: Frames resultantes da etapa de remoção de fundo.



O valor limite adotado para a função threshold foi de 60.
Fonte: Adaptado de MARCOMINI (2018).

Etapa 1.4. DETECÇÃO DA FEATURES E RASTREAMENTO DOS VEÍCULOS

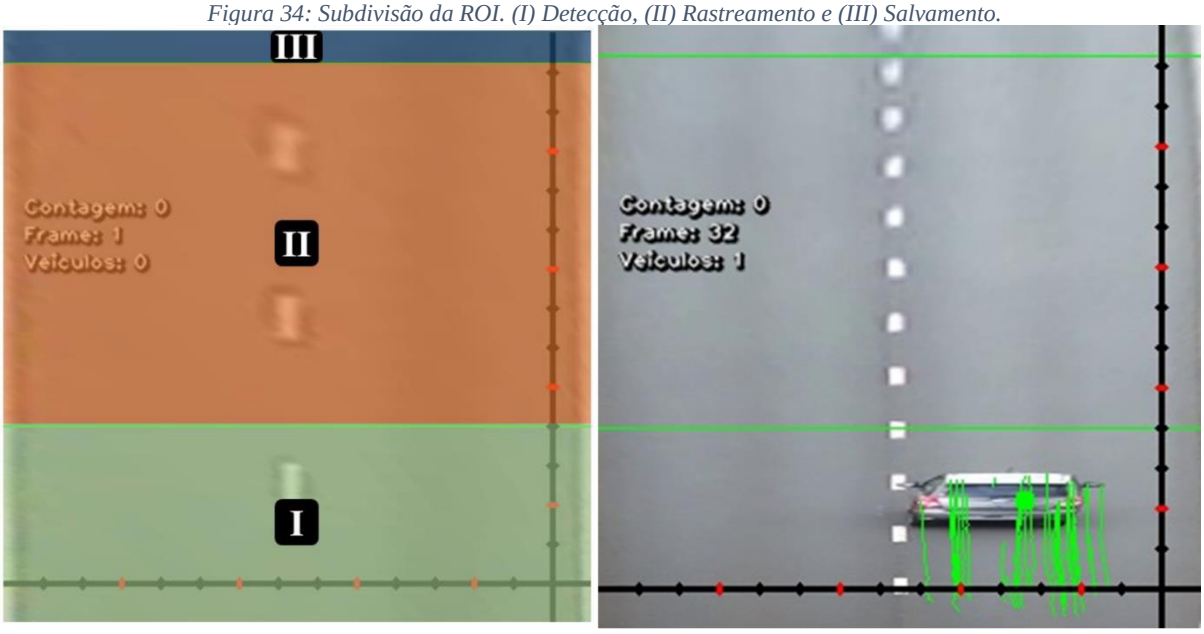
Tendo as regiões dos veículos (objetos em movimento) devidamente separados, a etapa de rastreamento (*tracking*) é iniciada. Os métodos utilizados nesta etapa foram: algoritmo de Shi-Tomasi para detectar as *features* (pontos interessantes a serem analisados) e algoritmo Lucas-Kanade que avalia e reconhece as *features* dos veículos em *frames* consecutivos permitindo que uma associação temporal seja realizada para um mesmo veículo durante todo o seu percurso.

Etapa 1.5. RASTREAMENTO E AGRUPAMENTO DOS VEÍCULOS

A etapa de rastreamento é dividida em três momentos, como mostrado pela Figura 34: (i) fase do movimento do veículo dentro da janela de perspectiva na qual ocorre a detecção de *features*, definida aproximadamente por 33,3% da ROI (100 pixels); (ii) fase do movimento do veículo dentro da janela de perspectiva na qual ocorre o rastreamento em si e região na qual a associação temporal acontece, definida aproximadamente por 61,6% da ROI (185 pixels) e (iii) fase do movimento do veículo dentro da janela

de perspectiva na qual é representada uma linha de final de percurso dos veículos, uma espécie de "Linha de Chegada" (LC), sendo que a definição dessa última área é de aproximadamente 5% da ROI (15 pixels).

Figura 34: Subdivisão da ROI. (I) Detecção, (II) Rastreamento e (III) Salvamento.



Fonte: Adaptado de MARCOMINI (2018).

3.3.2. OBTENÇÃO DOS DADOS VEICULARES

Com a utilização do módulo IACT proposto por MARCOMINI (2018) é possível ter acesso a vários tipos de dados, entre eles temos os (i) Relatórios Resumidos, (ii) Relatórios Detalhados e (iii) Diagrama Espaço-Tempo. Nessa seção será explicado e exemplificado cada um desses tipos de dados de saída fornecidos pelo sistema.

Relatório Resumido

O relatório resumido agrega a grande quantidade de informações coletadas durante a etapa de rastreamento permitindo uma rápida consulta aos dados, como velocidade média, o número da faixa e o valor do frame final que determinado veículo encerrou seu percurso, seu padrão de salvamento é "nome do vídeo_veículos_resumo.csv". Na Figura 35 é possível ver um exemplo de um dos relatórios resumidos disponibilizados pelo sistema.

Figura 35: Exemplo de Relatório Resumido.

```
201718041401_veiculos_resumo.csv - Bloco de notas
Arquivo  Editar  Formatar  Exibir  Ajuda
###### Dados Veiculares Resumidos ######
Tempo Total do Vídeo = 00:12:41.19
Padrão do arquivo:
ID, Velocidade Média, Frame Final, Faixa Final
##########################################

1,43,141,2
2,27,423,1
3,37,444,2
4,33,446,2
5,40,497,1
6,39,536,2
7,46,657,2
8,44,825,1
9,46,897,1
10,36,1084,2
11,44,1126,1
12,37,1390,2
13,33,1697,2
14,43,1722,1
```

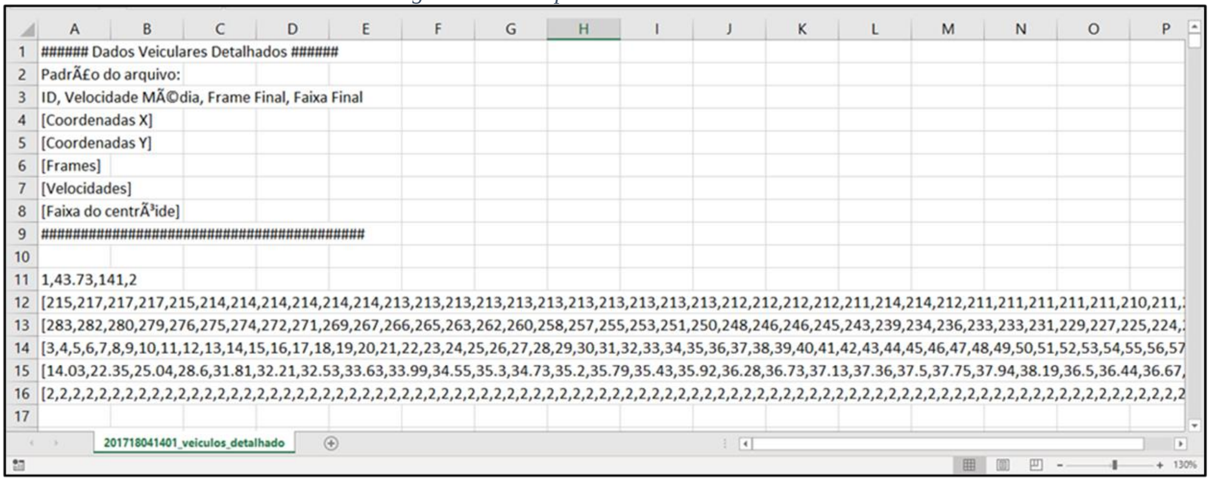
Fonte: O autor.

Relatório Detalhado

O relatório detalhado por sua vez apresenta de maneira completa todos os dados coletados durante a etapa de rastreamento permitindo uma investigação mais completa e profunda das condições do tráfego durante o período analisado. Como este arquivo contém uma quantidade considerável de dados, se torna vantajoso manuseá-lo somente após importá-lo em alguma ferramenta de análise de dados. A Figura 36 exemplifica o arquivo aberto em planilha Excel®. Neste arquivo as informações fornecidas são: (i) as Coordenadas X, neste caso representando o eixo horizontal na janela de Perspectiva (linha 12 do arquivo); (ii) as Coordenadas Y, representando o eixo vertical na janela de Perspectiva (linha 13); (iii) o número do *Frame*, representando a passagem do tempo (linha 14); (iv) as Velocidades Instantâneas do veículo (linha 15); (v) o número da faixa onde o centroide do veículo se encontra (1 para faixa esquerda 1 e 2 para faixa da direita) (linha 16); (vi) além do resumo da trajetória de cada veículo, similar ao apresentado no Relatório Resumido (linha 11). Seu padrão de salvamento é "nome do vídeo_veículos_detalhado.csv".

Vale destacar que os primeiros dados presentes nos vetores Velocidade Instantânea e Faixa do Centroide, sofrem grande variação devido ao começo da trajetória do veículo, pois o rastreamento está se adaptando às *features* que constituem cada veículo na cena.

Figura 36: Exemplo de Relatório Detalhado.

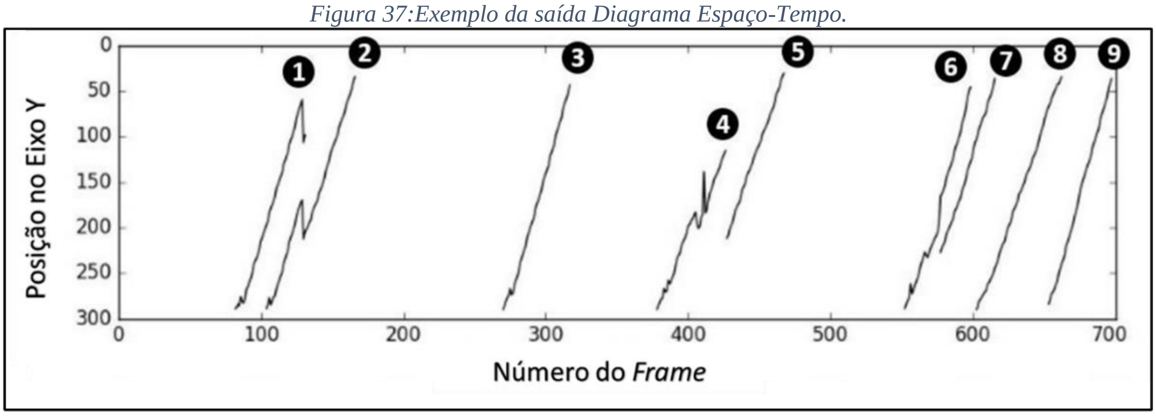


Fonte: O autor.

Diagrama Espaço-Tempo

O diagrama espaço-tempo é uma ferramenta gráfica muito utilizada na área da Engenharia de Transportes. No eixo horizontal são plotadas as informações referentes a passagem do tempo e no eixo vertical as informações referentes a posição do veículo no sentido da corrente de tráfego. Expor as trajetórias desta maneira permite que outras informações sejam derivadas, entretanto como este diagrama é uma saída do sistema proposto de MARCOMINI (2018), ele não apresenta nenhuma distinção entre os tipos de veículos presentes no diagrama, o que dificultou a utilização deste resultado nesta pesquisa. A Figura 37 retrata o diagrama espaço-tempo fornecido pelo sistema.

Figura 37:Exemplo da saída Diagrama Espaço-Tempo.



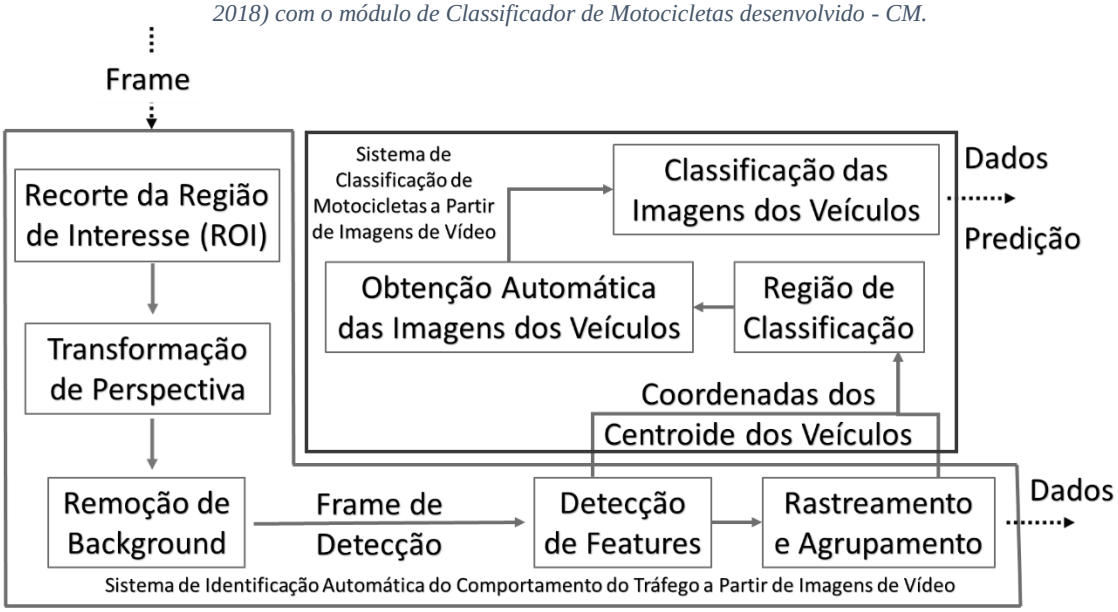
Fonte: MARCOMINI (2018).

3.3.3. CLASSIFICAÇÃO DOS VEÍCULOS

A associação do módulo IACT (MARCOMINI, 2018) com o Classificador de Motocicletas desenvolvido nesta pesquisa ocorre na etapa de detecção de features e rastreamento do sistema proposto por MARCOMINI (2018). No momento em que o centroide de um veículo é determinado, o processo de classificação deste veículo é iniciado.

É importante ressaltar que no desenvolvimento do classificador utilizado, (vide *Capítulo 2*), o procedimento para se obter as imagens dos veículos e classifica-los nas classes moto ou não moto ocorreu manualmente, entretanto como a associação com o sistema proposto por MARCOMINI (2018) esse procedimento pode ser automatizado. Sendo assim, uma rotina computacional foi codificada para extrair as imagens dos veículos no vídeo a partir das coordenadas dos centroides dos veículos. A Figura 38 mostra as etapas que constituem o módulo de Classificação de Motocicletas desenvolvido - CM e como este se associa ao módulo IACT proposto por MARCOMINI (2018).

Figura 38: A associação do sistema de Identificação Automática do Comportamento do Tráfego – IACT (MARCOMINI, 2018) com o módulo de Classificador de Motocicletas desenvolvido - CM.



Fonte: Adaptado de MARCOMINI (2018)

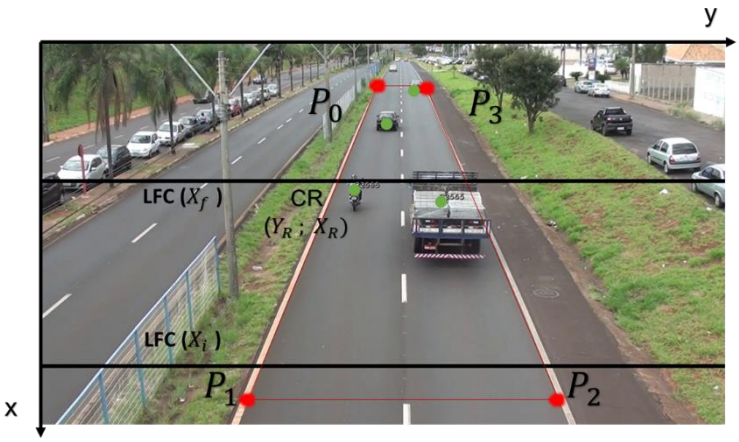
Etapa 1.1. REGIÃO DE CLASSIFICAÇÃO

Por motivo de desempenho do classificador e para que a adição do módulo CM acarretasse um impacto reduzido ao processamento geral do sistema proposto como um todo optou-se por criar uma região de classificação delimitada por uma Linha de Início de Classificação (LIC) e uma Linha de Final de Classificação (LFC).

A LIC deve utilizar o início da ROI (pontos P_1 e P_2) mais precisamente as coordenadas X (X_1 e X_2). O usuário deve verificar qual dos pontos possui o maior valor da coordenada X e deste valor subtrair metade da altura padrão das imagens utilizadas pelo classificador ($^{h}/_{2} = 105$ pixels). O resultado deste cálculo define a altura da LIC (X_i). No caso da estimativa da altura da LFC (X_f), deve-se levar em consideração as dimensões da malha projetada para extrair as imagens utilizada no desenvolvimento do classificado, (vide *Capítulo 2*), permitindo assim a obtenção de um conjunto de imagens razoável por veículo e evitando uma redução excessiva nas dimensões deste mantendo assim a eficiência na classificação. Desta forma a LFC deve estar a uma distância da LIC de no máximo três vezes a altura da

imagem padrão (630 pixels aproximadamente) ou seja, $X_f \leq X_i - 630$. Tanto a coordenada da altura da LIC (X_i) como a coordenada da altura da LFC (X_f) devem ser inseridas pelo usuário dentro do código fonte em *Python*. A Figura 39 mostra a projeção das duas linhas horizontais de obtenção e classificação das imagens no *frame* original.

Figura 39: Exemplo de veículos dentro e fora da região de classificação.



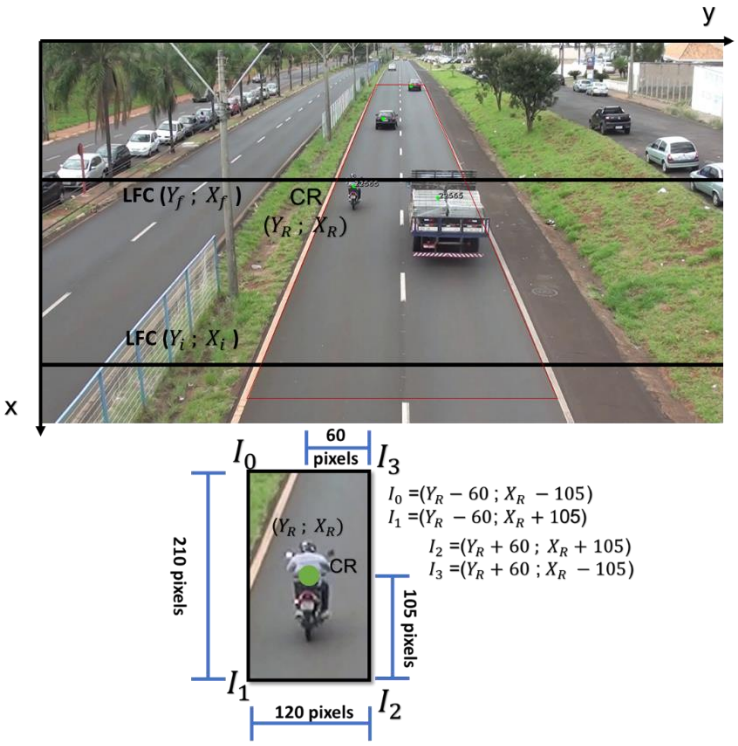
$P_0 = (Y_0 ; X_0)$ $P_3 = (Y_3 ; X_3)$
$P_1 = (Y_1 ; X_1)$ $P_2 = (Y_2 ; X_2)$

Os pontos que constituem a ROI são mostrados em vermelho e os centroides dos veículos são demostrados por pontos verdes. Fonte: O autor.

Etapa 1.2. OBTENÇÃO AUTOMÁTICA DAS IMAGENS DOS VEÍCULOS

Como as dimensões das imagens no modelo de classificação calibrado adotado eram de 120x210 pixels, as imagens a serem obtidas a partir do centroide do veículo também devem possuir essas mesmas dimensões. O sistema de classificação utiliza as coordenadas do centroide do veículo $CP = (Y_P; X_P)$, determinadas na janela de perspectiva (300x300 pixels), e converte para as coordenadas reais $CR = (Y_R; X_R)$, *frame* original (1920x1080 pixels). Uma vez tendo obtido a coordenada X_R do veículo o sistema verifica se esta coordenada se encontra entre as LIC e a LFC, caso a coordenada esteja dentro da região de classificação, $X_i \geq X_R \geq X_f$, sendo X_i a altura de LIC e X_f a altura da LFC, o sistema extrai uma imagem nas dimensões da imagem padrão como mostrado na figura 40.

Figura 40: Ilustração da obtenção das coordenadas da imagem a ser extraída e classificada a partir das coordenadas do centroide dos veículos.



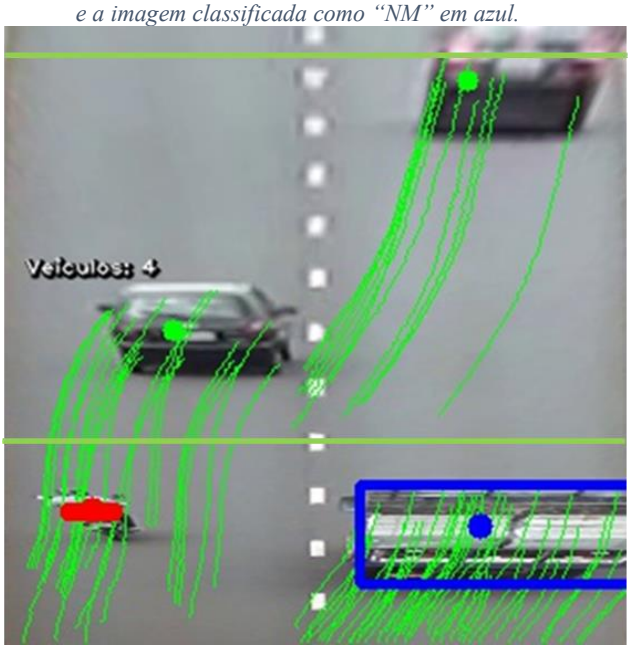
Fonte: O autor.

Etapa 1.3. CLASSIFICAÇÃO DAS IMAGENS DOS VEÍCULOS

Dado que o sistema tenha extraído uma imagem a partir de um centroide de um veículo a rotina computacional do classificador de motocicletas desenvolvido é acionado, gerando um histograma em escala de cinza normalizado (produzido pela função LBP) que é comparado com o vetor de característica e com o hiperplano ótimo, desenvolvido durante a etapa de treinamento do modelo, possibilitando a geração de uma predição / classificação desta imagem extraída em uma das classes "moto" ou "não moto".

A Figura 41 mostra um exemplo de como a classificação é mostrada na janela de perspectiva, quando a imagem obtida é classificada como Moto ("M") seu centroide e contorno são alterados para cor vermelha, entretanto quando a imagem obtida é classificada como Não-Moto ("NM") seu centroide e contorno são alterados para cor azul.

Figura 41: Etapa de classificação sendo realizada na janela em perspectiva, a imagem classificada como "M" em vermelho e a imagem classificada como "NM" em azul.



Fonte: O autor.

3.3.4. OBTENÇÃO DA CLASSIFICAÇÃO DOS VEÍCULOS

Com a associação do sistema de MARCOMINI (2018) com o Classificador de Motocicletas desenvolvido nesta pesquisa foi possível ter acesso a vários tipos de dados. Dos dados provenientes da etapa classificação temos os (i) Relatório de Predição, (ii) Imagens utilizadas para a Classificação, (iii) Imagens da janela em Perspectiva no momento em que os veículos cruzam a linha de chegada e (iv) Imagens do *frame* original no momento em que os veículos cruzam a linha de chegada.

Relatório de Predição

O relatório de predição contém a predição da classe realizada para cada imagem obtida e o valor do *Score* gerado pelo classificador (dado que é utilizado pelo modelo para realizar a predição) e a identificação do veículo ("nome do veículo; número do *frame* atual"). O padrão de salvamento é "Predictions_nome do vídeo.txt". A Figura 42 ilustra uma saída do Relatório de Predição.

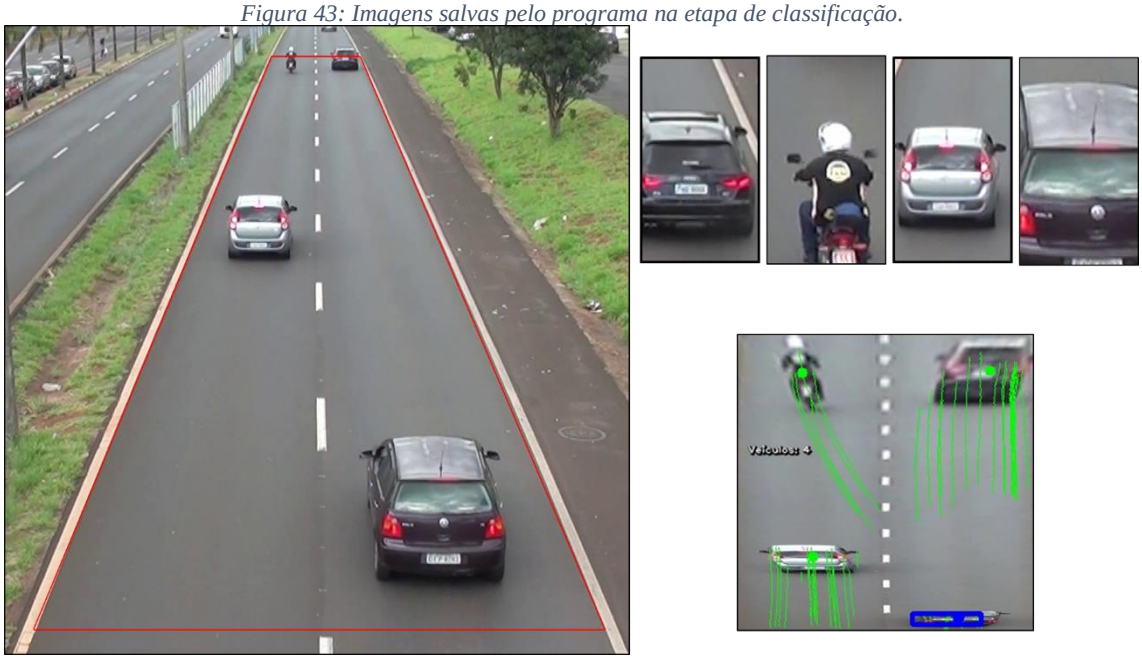
Figura 42: Exemplo de Relatório de Predição.

```
Imagem;Predição;Score
['Veic1;Frame30.jpg;N_M;0.611966578344']
['Veic1;Frame31.jpg;MOTO;-0.174043642681']
['Veic2;Frame31.jpg;MOTO;-0.922920043276']
['Veic1;Frame33.jpg;N_M;0.579005067392']
['Veic1;Frame55.jpg;N_M;1.37825535742']
['Veic1;Frame56.jpg;N_M;1.14289443428']
['Veic1;Frame57.jpg;N_M;1.50186023837']
['Veic1;Frame58.jpg;N_M;0.950273511795']
['Veic1;Frame59.jpg;N_M;1.37172412581']
['Veic1;Frame60.jpg;N_M;1.68222809427']
['Veic1;Frame61.jpg;MOTO;-0.264467838954']
['Veic1;Frame62.jpg;N_M;0.484175889654']
['Veic1;Frame262.jpg;N_M;1.66434296526']
['Veic1;Frame263.jpg;N_M;2.26581940446']
['Veic1;Frame264.jpg;N_M;2.35013727192']
['Veic1;Frame265.jpg;N_M;2.54606522031']
['Veic1;Frame266.jpg;N_M;2.39815913936']
['Veic1;Frame267.jpg;N_M;2.20448249578']
['Veic1;Frame268.jpg;N_M;2.46685402061']
```

Fonte: O autor.

As outras três saídas adicionadas ao programa foram os salvamentos das Imagens utilizadas para a Classificação, as Imagens da janela em Perspectiva no momento em que os veículos cruzam a linha de chegada e as Imagens do *frame* original no momento em que os veículos cruzam a linha de chegada, que estão exemplificadas na Figura 43. Com o auxílio delas foi possível acompanhar as principais situações que ocorreram no tráfego sem a necessidade de assistir mais uma vez aos vídeos de monitoramento.

Figura 43: Imagens salvas pelo programa na etapa de classificação.



A esquerda, o frame original quando o primeiro veículo ("NM") atravessam a linha de chegada, a direita superior, exemplos das imagens utilizadas para classificar os veículos e a direita inferior, a janela em Perspectiva quando o primeiro veículo ("NM") atravessam a linha de chegada.
Fonte: O autor.

3.4. RESULTADOS

3.4.1. FUSÃO DOS DADOS OBTIDOS

Uma vez tendo acesso aos dados fornecidos pelos dois módulos, modulo de identificação automática do comportamento do tráfego e módulo de classificação de motocicletas, se tornou necessário confrontá-los para que a fusão destes dados ocorresse. Visto que nos *relatórios detalhados* as trajetórias dos veículos são agrupadas e referenciadas a cada respectivo veículo devido a associação temporal realizada na etapa de rastreio, no *relatório de predição* não se possui esse mesmo agrupamento ou referenciamento por veículo e sim uma predição / classificação por imagens.

Desta forma utilizando os IDs dos veículos e seus *Frames* finais obtidos no Relatórios Detalhados juntamente com as imagens salvas do *frame* original e da janela em Perspectiva quando os veículos cruzam a linha de chegada foi possível identificar os veículos pelos seus IDs. Em seguida foi possível confrontar as imagens destes veículos identificados com as respectivas predições / classificações atribuídas a essas imagens e determinar para cada veículo qual classe este mais se enquadrou, segundo a maioria das predições realizadas as imagens referentes a ele. Esta abordagem permitiu gerar manualmente um novo relatório com a fusão dos dados (Figura 44).

O padrão deste arquivo segue a seguinte sequência: na primeira coluna é mostrado os dados agregados por veículos juntamente com sua predição (fornecida pelo classificador) e sua classificação (fornecida pelo autor), em seguida nas próximas seis colunas são mostrados os dados detalhados por veículo, suas Coordenadas no eixo X, suas Coordenadas no eixo Y, o número do frame, a Velocidade instantânea, o número da Faixa na qual o centroide do veículo se encontra, juntamente com o cálculo da Aceleração (adicionado pelo autor e baseado nos dados da velocidade instantânea).

Figura 44: Exemplo do Relatório com a fusão dos dados,

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1			COO X	COO Y	FRA	VEL	FAI cen	ACELERAÇÃO		COO X	COO Y	FRA	VEL	FAI cen
2	ID	1	215.00	283.00	3.00	14.03	2.00	69.34	2	114.00	296.00	205.00	13.45	1.00
3	VEL med	43.73	217	282	4	22.35	2	22.42	27.85	90	294	206	19.31	1
4	FRA fin	141	217	280	5	25.04	2	29.67	423	96	293	207	25.41	1
5	FAI fin	2	217	279	6	28.6	2	26.75	1	101	292	208	24.49	1
6	PREDIÇÃO	n_m	215	276	7	31.81	2	3.33	null	100	291	209	26.54	1
7	CLASSIFICAÇÃO	n_m	214	275	8	32.21	2	2.67	n_m	99	289	210	28.94	1
8			214	274	9	32.53	2	9.17		99	288	211	29.86	1
9			214	272	10	33.63	2	3.00		99	287	212	29.58	1
10			214	271	11	33.99	2	4.67		95	287	213	27.24	1
11			214	269	12	34.55	2	6.25		95	286	214	28.31	1
12			214	267	13	35.3	2	-4.75		94	285	215	28.63	1
13			213	266	14	34.73	2	3.92		95	285	216	29.03	1
14			213	265	15	35.2	2	4.92		94	284	217	28.25	1
15			213	263	16	35.79	2	-3.00		95	282	218	29.07	1
16			213	262	17	35.43	2	4.08		93	283	219	27.09	1
17			213	260	18	35.92	2	3.00		94	283	220	27.01	1

201718041401_detalhado | Predição | 201718041401_BD

Fonte: O autor.

3.5. CONCLUSÕES

A associação realizada entre o módulo de Classificador de Motocicletas desenvolvido - CM nesta pesquisa (vide *Capítulo 2*) com o módulo de Identificação Automática do Comportamento do Tráfego – IACT proposto por MARCOMINI (2018) permitiu identificar os veículos presentes em vídeos de tráfego, coletar algumas de suas informações , tais como a velocidade e a aceleração praticadas durante seu trajeto e classificar o tipo destes veículos em motos ou não motos.

4. ANÁLISE EXPLORATÓRIA DO COMPORTAMENTO DOS MOTOCICLISTAS UTILIZANDO VÍDEOS DO TRÁFEGO

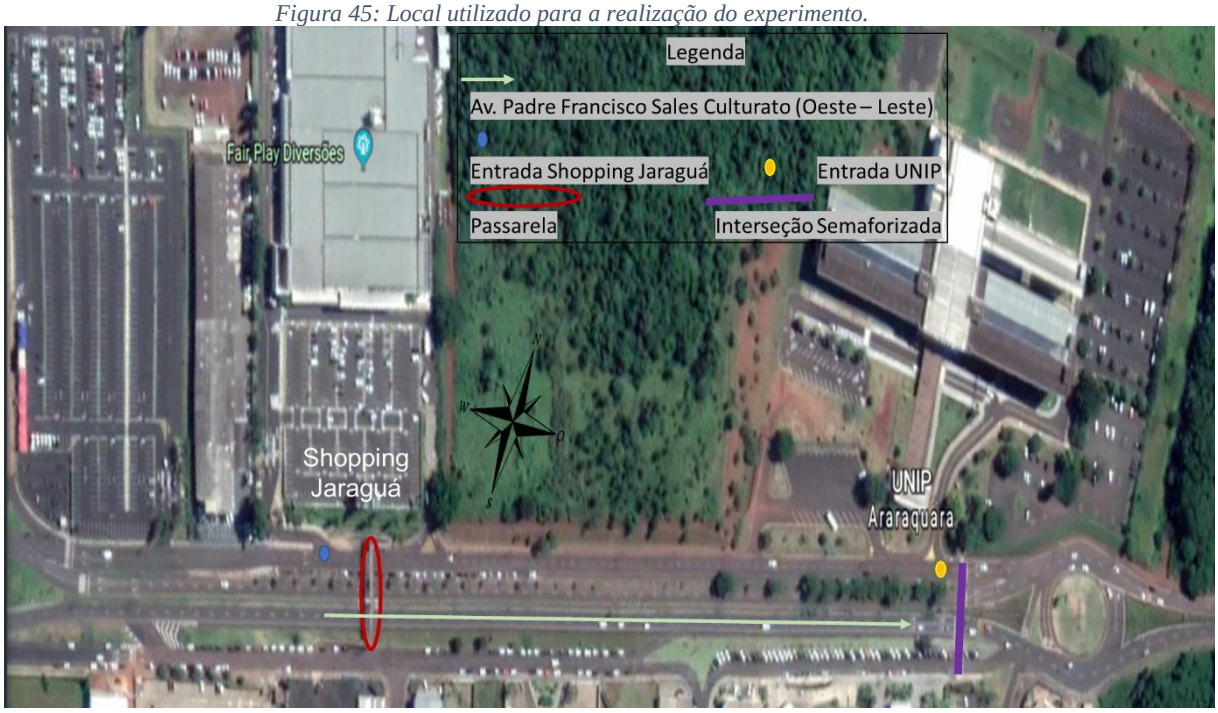
Este *Capítulo 4* tem como objetivo principal investigar o comportamento dos motociclistas, conjunto de reações apresentadas mediante a uma interação com o meio em que se encontra, a partir dos dados veiculares obtidos pela associação do módulo de Identificação Automática do Comportamento do Tráfego a Partir de Imagens de Vídeo – IACT, proposto por (MARCOMINI, 2018) com o módulo de Classificação de Motocicletas desenvolvido – CM, desenvolvido no *Capítulo 3* deste trabalho.

Os dados do comportamento das duas classes, motociclistas e não motociclistas foram analisados. Desta forma tanto as trajetórias de cada uma das motocicletas como as dos demais veículos presentes nos vídeos de monitoramento do tráfego foram registradas e armazenadas em um relatório final contendo além da contagem total de veículos em geral, a contagem total de veículos de cada uma das duas classes, além do resultado do cálculo da velocidade média de cada um dos veículos e também o resultado do cálculo da aceleração média de cada um dos veículos.

4.1. ESTUDO DE CASO

O local escolhido para a realização do experimento de análise exploratória do comportamento dos motociclistas utilizando vídeos do tráfego foi a Av. Padre Francisco Sales Culturato (sentido Oeste – Leste), localizada no município de Araraquara – SP sendo que os vídeos do tráfego utilizados registram imagens da corrente de tráfego iniciando próximas a entrada do shopping Jaraguá e finalizando próximo a entrada da UNIP, que se encontra nas proximidades de uma interseção semaforizada. A figura 45 ilustra o local do experimento. O trecho monitorado possui uma extensão de aproximadamente 400 metros, apresenta um fluxo de aproximadamente 601,1 veículos por hora, sendo que as motocicletas representam aproximadamente 19,6% desse fluxo. A coleta dos vídeos ocorreu das 10:50 as 14:10, totalizando 3 horas e 20 minutos de filmagem e coletando um total de 2.552 veículos sendo que destes 435 eram motocicletas (17%) conforme explicado no *Capítulo 2*,

Figura 45: Local utilizado para a realização do experimento.



Fonte: O autor sobre imagem do Google Earth.

4.2. MÉTODO PROPOSTO

O objetivo central desta pesquisa como um todo é testar a hipótese de que: *é possível identificar os comportamentos de motociclistas em uma corrente de tráfego utilizando imagens de vídeos e possibilitando assim a análise do comportamento desses condutores.* Mediante a esse questionamento foi elaborado um método dividido em quatro etapas: (i) Separação dos vídeos em conjuntos de treinamento e conjuntos de teste; (ii) Treinamento do Classificador de Motocicletas; (iii) Aplicação da Associação dos módulos IACT / CM ao conjunto teste e (iv) Análise dos dados fundidos obtidos.

4.2.1. SEPARAÇÃO DOS VÍDEOS EM CONJUNTOS DE TREINAMENTO E CONJUNTOS DE TESTE

Para padronizar o tempo de processamento do classificador optou-se por utilizar somente os vídeos a favor do fluxo do tráfego e excluir o vídeo de 7 minutos. Assim, neste tópico do trabalho utilizou-se somente os 15 vídeos, totalizando 180 minutos, para o processamento total do sistema.

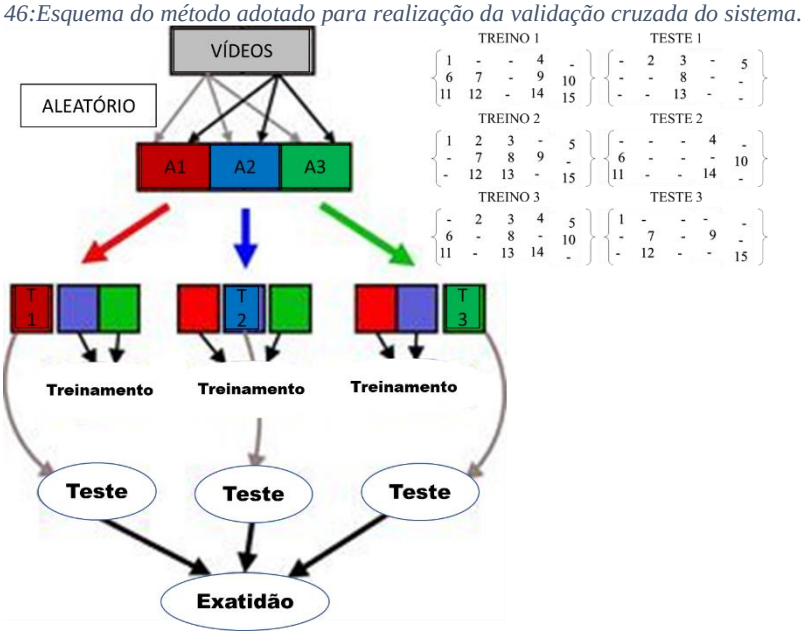
No momento em que se realizou o treinamento do modelo de predição do classificador de motocicletas desenvolvido no *capítulo 2* o conjunto de treinamento era constituído por imagens extraídas de todos os vídeos coletados a favor do fluxo do tráfego e no momento de validar as predições fornecidas pelo modelo de classificação o conjunto de teste também era constituído da mesma forma, entretanto esses conjuntos

eram mutuamente excludentes, ou seja o conjunto teste possuía elementos inéditos ao conjunto de treinamento.

No caso particular do experimento, o conjunto de dados não é constituído por imagens extraídas dos vídeos do tráfego e sim pelos próprios vídeos do tráfego. Sendo assim foi preciso separar novos conjuntos de treinamento e novos conjuntos de teste para que no momento de se avaliar o desempenho do classificar, todas as imagens que fossem apresentadas a ele, tivessem o título de inéditas de fato.

Como o conjunto de dados original possui 15 vídeos, utilizou-se o critério de validação cruzada, método *K-Fold*, com k = 3 (produção de 3 amostras) e método *Hold-Out*, no qual o conjunto treinamento representa 70% do conjunto de dados original (10 vídeos) e o conjunto teste representa 30% do conjunto de dados original (5 vídeos). Desta forma a Figura 46 ilustra o método utilizado para a subdivisão do conjunto de dados originais nos conjuntos de dados de treinamento e conjunto de dados de teste.

Figura 46:Esquema do método adotado para realização da validação cruzada do sistema.



Fonte: O autor sobre imagem do Wikipedia.

Somente após os três conjuntos de treino terem sido definidos (quais vídeos de fato foram utilizados para compor cada um dos três conjuntos) que pode se utilizar as imagens extraídas dos respectivos vídeos de tráfego para de fato montar o conjunto de imagens de treinamento para calibrar os classificadores de motocicletas.

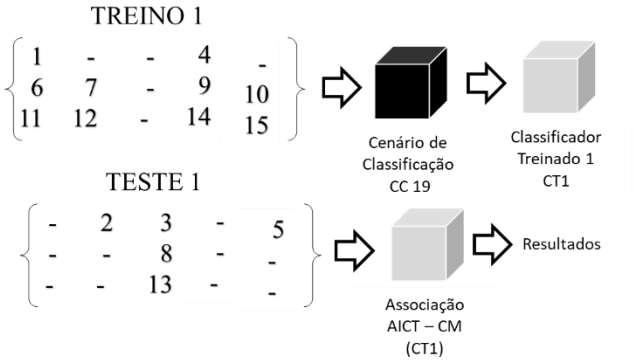
4.2.2. TREINAMENTO DO CLASSIFICADOR DE MOTOCICLETAS

O treinamento do classificador de motocicletas utilizado neste experimento ocorreu similarmente como o explicado no *capítulo 2*, porem utilizando as três novas amostras geradas e somente utilizando as estimativas dos parâmetros segundo o cenário de calibração 19 (parâmetros C = 150 e Tol = 0,01, mantendo os demais parâmetros em sua configuração padrão).

4.2.3. APLICAÇÃO DA ASSOCIAÇÃO DOS MÓDULOS IACT / CM AO CONJUNTO TESTE

Com a realização do item anterior foi obtido três classificadores de motocicletas treinados de acordo com o esquema mostrado na Figura 47. Desta forma o sistema proposto como um todo (associação dos módulos AICT com o módulo CM) foi aplicado aos 15 vídeos de tráfego disponíveis, sendo que para cada vídeo o sistema utilizava um dos três classificadores treinados.

Figura 47: Exemplo da aplicação da associação dos módulos do sistema proposto como um todo para a amostra 1.



Fonte: O autor.

4.3. DISCUSÃO E ANÁLISE DOS RESULTADOS OBTIDOS

Após a aplicação do sistema proposto a cada um dos 15 vídeos utilizados obteve-se 15 conjuntos de dados veiculares e de dados de classificação de veículos que foram devidamente agrupados em um único arquivo.

Em determinadas situações nas quais os veículos se encontravam muito próximos uns aos outros, o método agrupou *features* de ambos os veículos como se fossem um único objeto em movimento, resultando em centroide e uma predição errados, caso os veículos envolvidos pertencessem a classes distintas. Em outras situações, o método produziu mais de um centroide por veículo, isso acontecia com veículos maiores (caminhões ou ônibus) que eram rastreados mais de uma vez. Também foi observado situações onde o programa identificava veículos e os rastreava corretamente, porém nenhuma imagem deste havia sido coletada e nem classificada produzindo uma ausência de dados. Neste último caso, a marcação “null” foi registrada denominando uma nova classe a ser excluída, essa anotação pode ser observada na Figura 48, na célula “I6”, sendo que dos 2.566 veículos coletados ao todo, 185 foram classificados nesta nova classe, representando 7,2% de perda.

Figura 48: Exemplo do Relatório com a fusão dos dados. Em destaque a marcação "null" utilizada quando notada alguma discrepância entre os módulos para identificar e classificar algum veículo.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1			COO X	COO Y	FRA	VEL	FAI cen	ACELERAÇÃO		COO X	COO Y	FRA	VEL	FAI cen
2	ID	1	215.00	283.00	3.00	14.03	2.00	69.34	2	114.00	296.00	205.00	13.45	1.00
3	VEL med	43.73	217	282	4	22.35	2	22.42	27.85	90	294	206	19.31	1
4	FRA fin	141	217	280	5	25.04	2	29.67	423	96	293	207	25.41	1
5	FAI fin	2	217	279	6	28.6	2	26.75	1	101	292	208	24.49	1
6	PREDIÇÃO	n_m	215	276	7	31.81	2	3.33	null	100	291	209	26.54	1
7	CLASSIFICAÇÃO	n_m	214	275	8	32.21	2	2.67	n_m	99	289	210	28.94	1
8			214	274	9	32.53	2	9.17		99	288	211	29.86	1
9			214	272	10	33.63	2	3.00		99	287	212	29.58	1
10			214	271	11	33.99	2	4.67		95	287	213	27.24	1
11			214	269	12	34.55	2	6.25		95	286	214	28.31	1
12			214	267	13	35.3	2	-4.75		94	285	215	28.63	1
13			213	266	14	34.73	2	3.92		95	285	216	29.03	1
14			213	265	15	35.2	2	4.92		94	284	217	28.25	1
15			213	263	16	35.79	2	-3.00		95	282	218	29.07	1
16			213	262	17	35.43	2	4.08		93	283	219	27.09	1
17			213	260	18	35.92	2	3.00		94	283	220	27.01	1

Fonte: O autor.

Para realizar a análise dos dados veiculares agregados de velocidade e aceleração foi utilizado o programa de análise de dados SPSS da empresa IBM. Desta forma um novo arquivo foi produzido agregando os dados conforme mostrado na Figura 49.

Figura 49: Arquivo com os dados fundidos e agregados utilizado como entrada para o programa SPSS.

	A	B	C	D	E	F	G	H	I	J
1	ID_Veículo	ID	VEL med [km/h]	ACELERAÇÃO [m/s²]	FRA fin	FAI fin	PREDIÇÃO	CLASSIFICAÇÃO	a < -2.5	2.5 > a
2	20171804_1102	1	45.14	1.07	304	2	1	1	0.00	0.00
3	20171804_1102	2	50.85	3.52	499	2	0	0	0.00	1.00
4	20171804_1102	3	41.12	0.00	504	1	1	1	0.00	0.00
5	20171804_1102	4	44.60	0.13	506	2	2	2	0.00	0.00
6	20171804_1102	5	34.47	-0.32	582	2	0	0	0.00	0.00
7	20171804_1102	6	48.65	0.71	909	2	0	0	0.00	0.00
8	20171804_1102	7	50.82	1.03	970	1	2	2	0.00	0.00
9	20171804_1102	8	47.43	-8.70	980	1	0	0	1.00	0.00
10	20171804_1102	9	41.90	-0.23	999	1	2	2	0.00	0.00
11	20171804_1102	10	40.52	-0.38	1003	1	2	2	0.00	0.00
12	20171804_1102	11	79.79	0.27	1052	1	2	2	0.00	0.00
13	20171804_1102	12	70.07	1.65	1119	2	2	2	0.00	0.00
14	20171804_1102	13	58.74	1.74	1154	2	2	2	0.00	0.00
15	20171804_1102	14	72.28	0.88	1195	2	2	2	0.00	0.00
16	20171804_1102	15	60.03	0.04	1222	2	2	2	0.00	0.00
17	20171804_1102	16	62.27	-0.06	1235	1	2	2	0.00	0.00

Fonte: O autor.

No programa SPSS, as variáveis nominais Predição e Classificação foram convertidas em numéricas, logo os valores “null”, “M” de Moto e “NM” de Não-Moto tornaram-se 0, 1 e 2, respectivamente. É possível observar na Figura 49 que nas duas últimas colunas, “I” e “J”, foi realizado a verificação dos valores obtidos da aceleração, colunas “D” para saber se estes se encontram dentro dos valores esperados para a situação real observada, $-2{,}5 < a < 2{,}5$ (trecho de declive leve). Caso o valor da aceleração estiver fora do intervalo esperado, a célula respectiva a este valor de aceleração nas colunas “I” ou “J” possui o valor igual a 1,00 e o preenchimento vermelho.

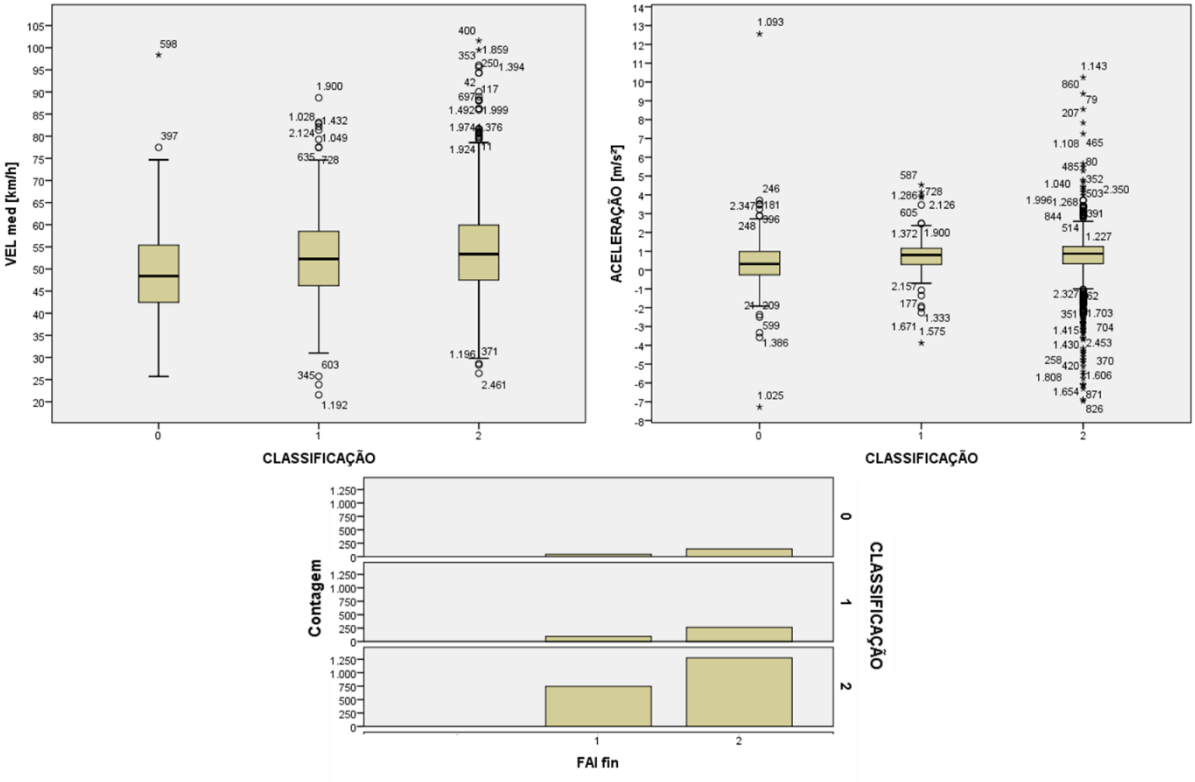
A análise exploratória dos dados mostrada na Figura 50 revelou a seguinte distribuição dos dados coletados e permitiu as seguintes observações:

i. A maioria dos veículos (95%) realmente respeita o limite de velocidade no trecho observado (V = 70 km/h), provavelmente por existir sinalização vertical informando o limite de velocidade localizada próxima ao local.

ii. Grande parte dos veículos (75%) trafega próximo a velocidade de 50 km/h, limite de velocidade este do final do trecho observado, próximo a interseção semaforizada devido a existência de uma rotatória a qual existe a instalação de equipamento de aferição da velocidade do tráfego.

Os dados de Velocidade média e de Aceleração das Classes Motocicletas e Não-Motocicletas mostraram-se muito similares, levantando a hipótese de que não há diferença na velocidade por tipo de veículo.

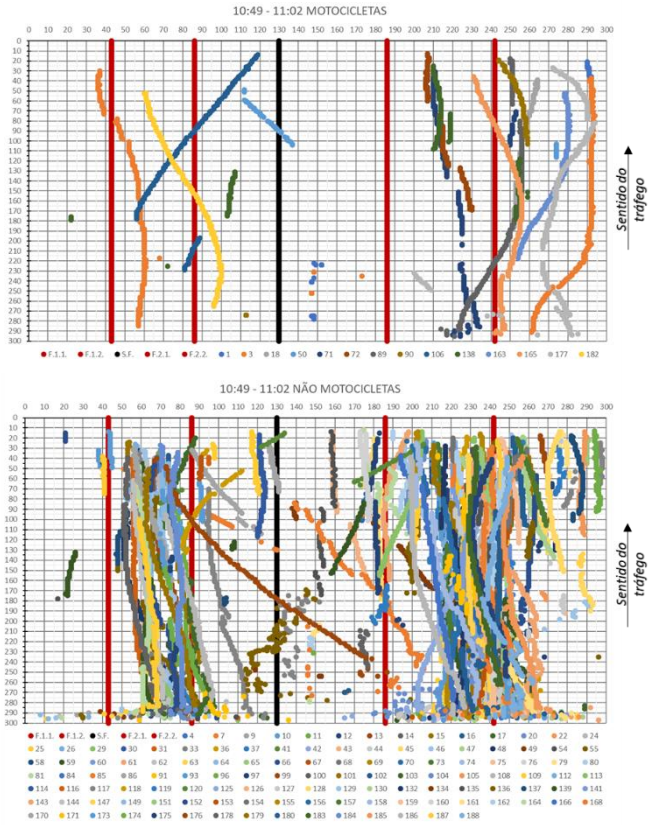
Figura 50: Distribuição dos dados por classes.



Classificação: 0 – "Null"; 1 – "Moto"; 2 – "Não Moto".
Faixa final: 1 – direita; 2 – esquerda.
Fonte: O autor.

Uma análise dos dados desagregados retrata o comportamento do tráfego a cada quinze minutos (duração de cada vídeo) utilizando as Coordenadas (X; Y) para ilustrar as trajetórias dos veículos em cada vídeo analisado, como exemplificado na Figura 51. Os gráficos que exemplificam as trajetórias dos veículos nos vídeos coletados distinguem os veículos por tipo. A linha vertical preta "C" representa a separação das faixas e as linhas verticais vermelhas representam a divisão de ambas as faixas em três partes.

Figura 51: Exemplo das trajetórias do primeiro vídeo coletado (11:02 até às11:14) representando 12 minutos de observação.



Fonte: O autor.

Todas as imagens geradas das trajetórias dos veículos separadas por vídeos e distinguidas por tipo de veículo podem ser encontradas nos Apêndice A.

A análise dos dados das trajetórias dos veículos nos permitiu levantar algumas hipóteses:

1. Devido as dimensões menores das motocicletas estas criam "faixas virtuais" dentro de uma faixa de tráfego real. E pela posição dos percursos deste tipo de veículo em particular essas "faixas de tráfego fantasmas" possuem aproximadamente 1/3 da largura da faixa de tráfego real.
2. Existe uma preferência por parte das motocicletas em trafegarem na faixa da direita, nos dados representado pelo número 2, mais precisamente na "faixa de tráfego fantasma" mais próxima ao acostamento. Provavelmente por ser a região com menor interação entre veículos.
3. Já os demais veículos também preferem trafegar pela faixa da direita, porém quando se analisa em qual porção das faixas estes veículos preferem transitar, é observado que em sua maioria eles transitam pela porção central das faixas. Provavelmente, pelo fato de suas dimensões serem maiores que as das motocicletas, transitar o mais centralizado na faixa se torne mais seguro para este tipo de veículo.

É importante comentar uma particularidade ao se analisar esse tipo de imagens. Como os gráficos retratam um cenário de 12 minutos, as linhas que estão se cruzando no gráfico não significam obrigatoriamente que os veículos as quais elas representam também se cruzaram. É preciso retornar aos

dados desagregados e confrontar a posição (X; Y) e os *Frames* de ambos os veículos para confirmar se realmente ocorreu uma interação entre esses veículos e compreender essa possível interação. Uma alternativa também é observar as imagens salvas, tanto as imagens da janela em Perspectiva como as imagens do *frame* original quando os veículos cruzam a linha de chegada.

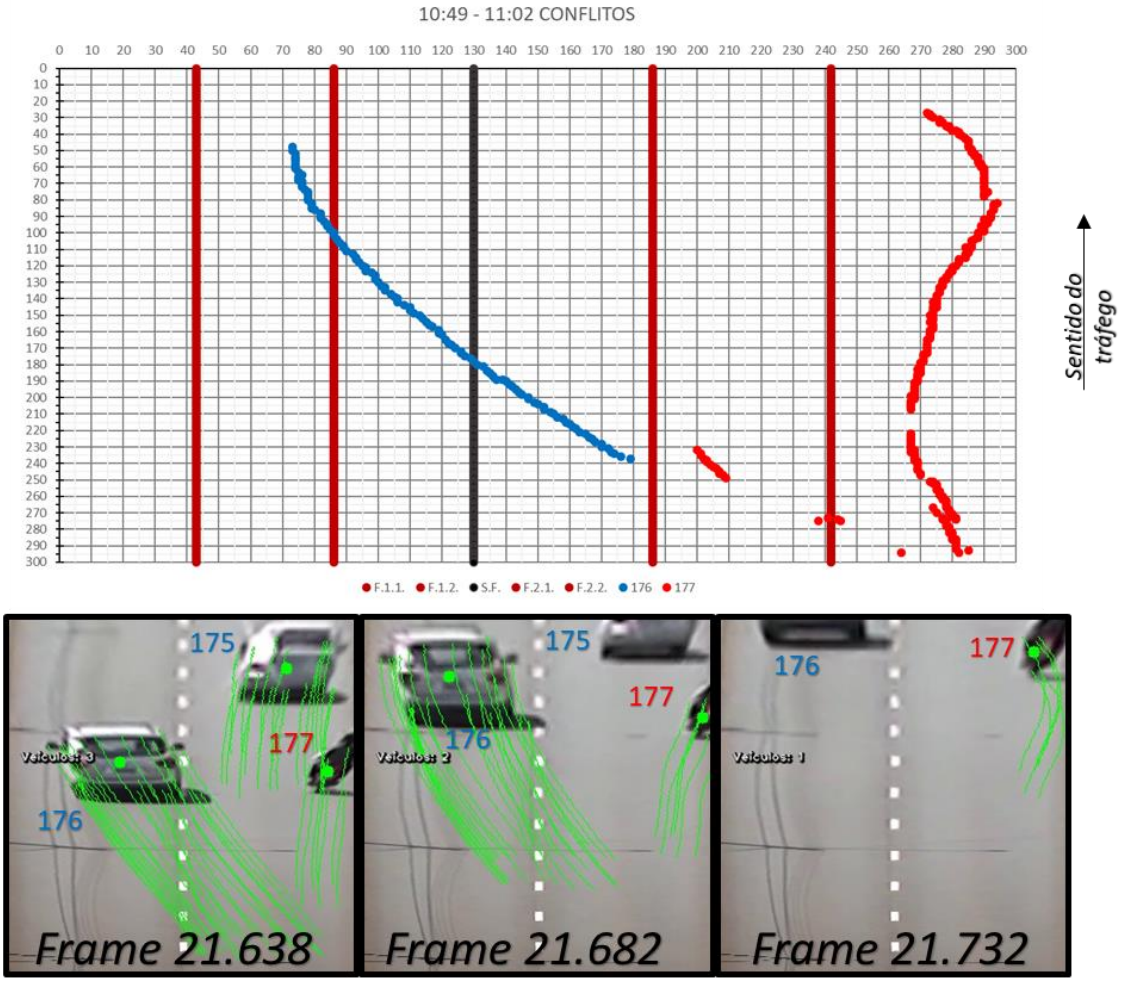
A análise das imagens das interações dos veículos nos permitiu levantar algumas hipóteses:

1. O movimento de ultrapassagem de um veículo de passeio a uma motocicleta, cada um localizado em faixas distintas porem consecutivas, ocasiona uma alteração no percurso retilíneo da motocicleta.
2. O movimento de ultrapassagem de uma motocicleta a um veículo de passeio, cada um localizado em faixas distintas porem consecutivas, não ocasiona uma alteração no percurso retilíneo do veículo de passeio.

Entretanto é bem visível constatar que alguns veículos realizaram manobras de mudanças de faixas e manobras consideráveis fora do usual. Como o objetivo do trabalho é o comportamento somente da Motocicletas, as análises de manobras consideráveis fora do usual se concentrou somente neste tipo de veículo.

Na Figura 52 é possível observar a interação entre o veículo 177 e o veículo 176, um automóvel e uma motocicleta respectivamente. Neste exemplo, nota-se no gráfico a trajetória da motocicleta realizando uma espécie de “S” na pista, comportamento esse bem diferente do esperado. Ao se analisar mais afundo sua trajetória pode se verificar que a mudança brusca realizada pelo veículo ao longo de sua trajetória se deu pela aproximação de outro veículo (automóvel 177) que realizou uma manobra de mudança de faixa seguida de uma ultrapassagem, como demostrado pelas imagens salvas da Figura 52.

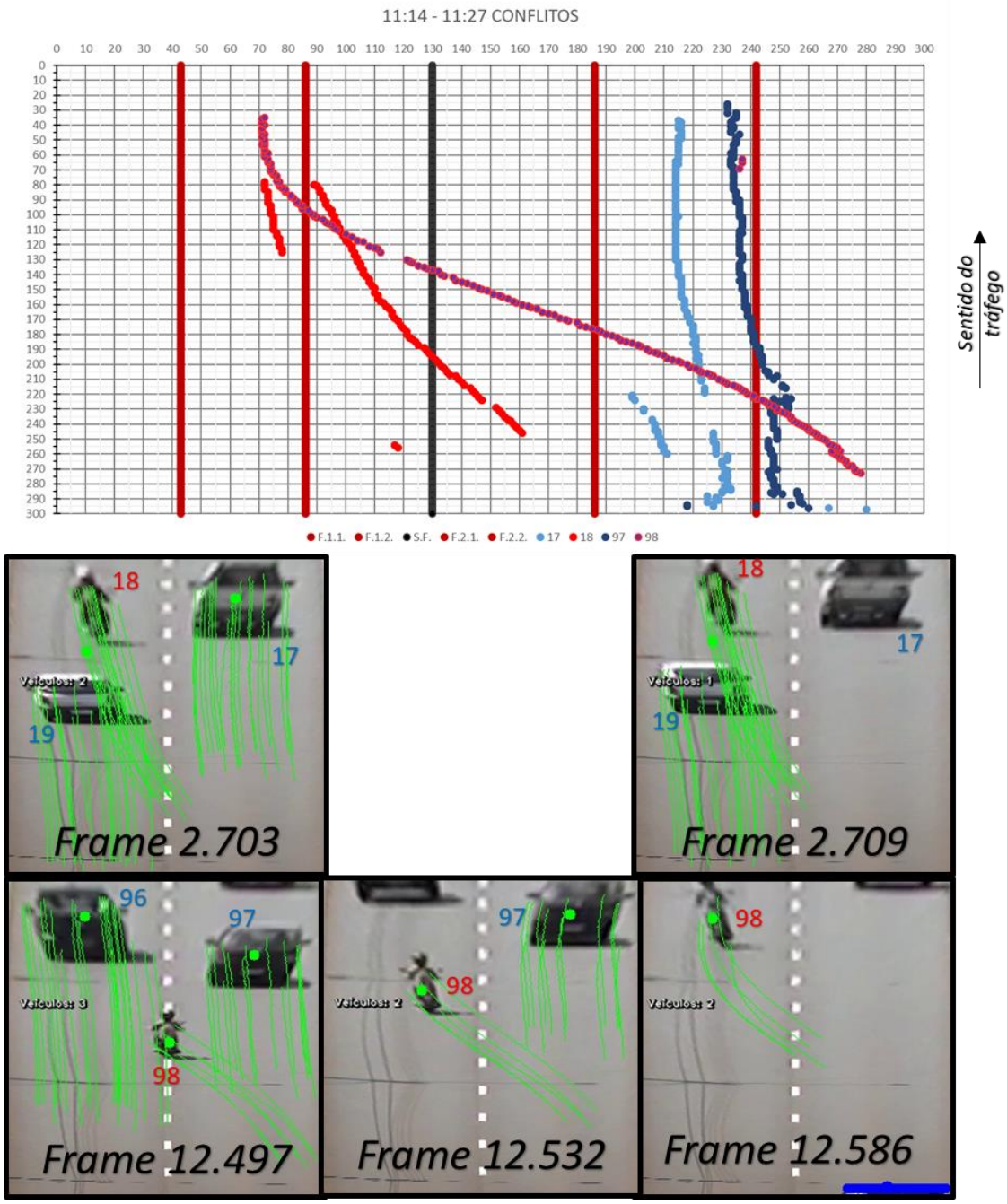
Figura 52: Exemplo das interações entre os veículos do primeiro vídeo coletado (10:49 até às11:02) representando 12 minutos de observação.



Interação entre o veículo Não-Motocicleta 176, em azul e o veículo Motocicleta 177, em vermelho. O gráfico distingue os veículos por tipo. A linha vertical preta "S.F." representa a separação das faixas e as linhas verticais vermelhas representam a divisão de ambas as faixas em três partes.
Fonte: O autor.

Outras duas situações interessantes a se comentar ocorreram no terceiro vídeo, como mostrado na Figura 53. Em ambas as situações é possível ver uma mudança de faixa realizada pelos veículos do tipo Motocicletas.

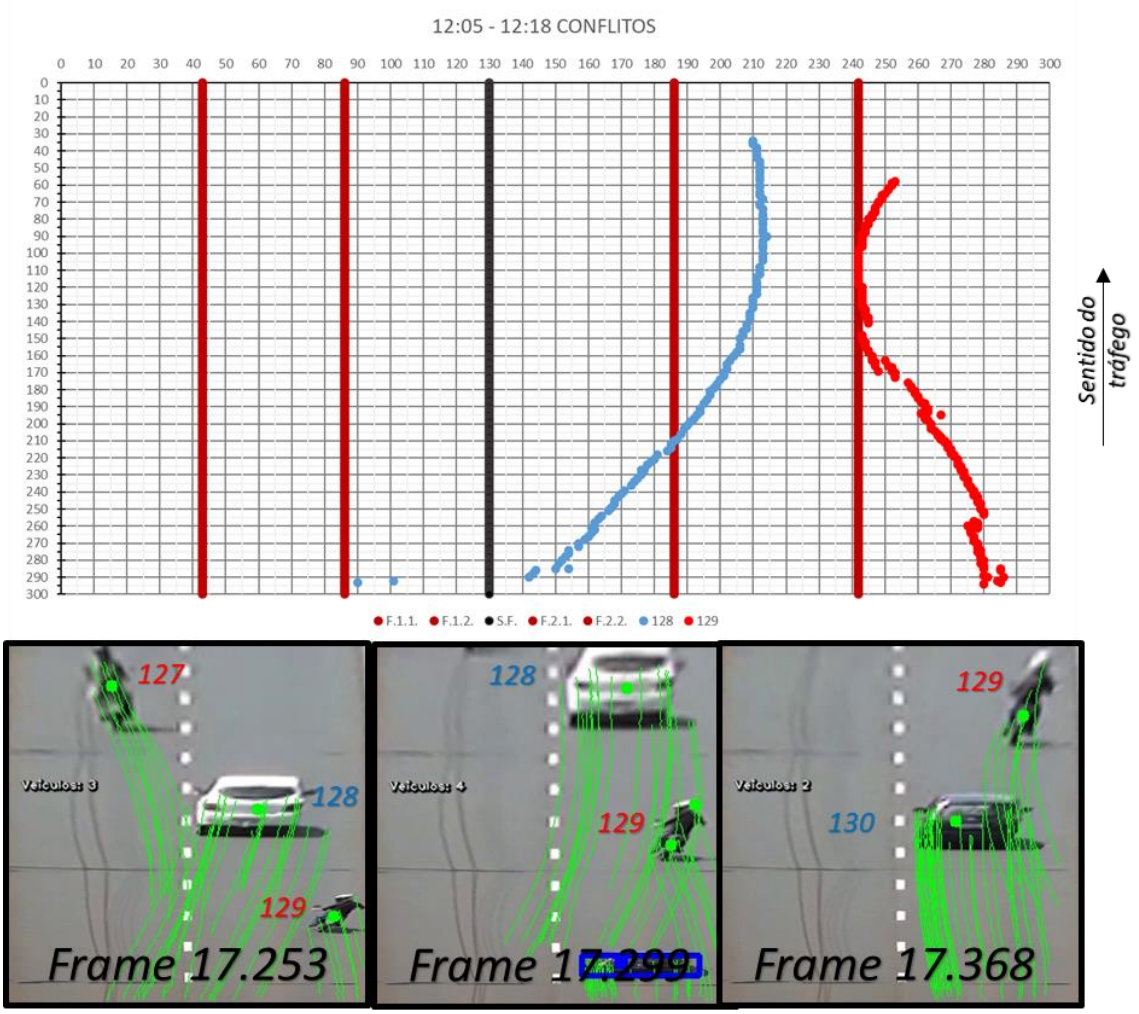
Figura 53: Exemplo das interações entre os veículos do terceiro vídeo coletado (11:14 até às11:27) representando 12 minutos de observação.



Interação entre o veículo 17e 18 e dos veículos 97 e 98. Os veículos Não-Motocicletas estão representados em azul e os veículos Motocicletas estão representados em vermelho. O gráfico distingue os veículos por tipo. A linha vertical preta "S.F." representa a separação das faixas e as linhas verticais vermelhas representam a divisão de ambas as faixas em três partes.
Fonte: O autor.

Durante a coleta dos dados foi possível observar que alguns condutores, principalmente motociclistas, ao notarem a presença das câmeras de coleta tiveram suas trajetórias comprometidas, como pode ser observado na Figura 54. Neste exemplo, é possível observar que o automóvel 128 realiza uma manobra de mudança de faixa sucedida por uma ultrapassagem a motocicleta 129. Interessante é observar o deslocamento lateral realizado pelo motociclista devido ao fato de estar olhando para trás (em direção as câmeras utilizadas na coleta).

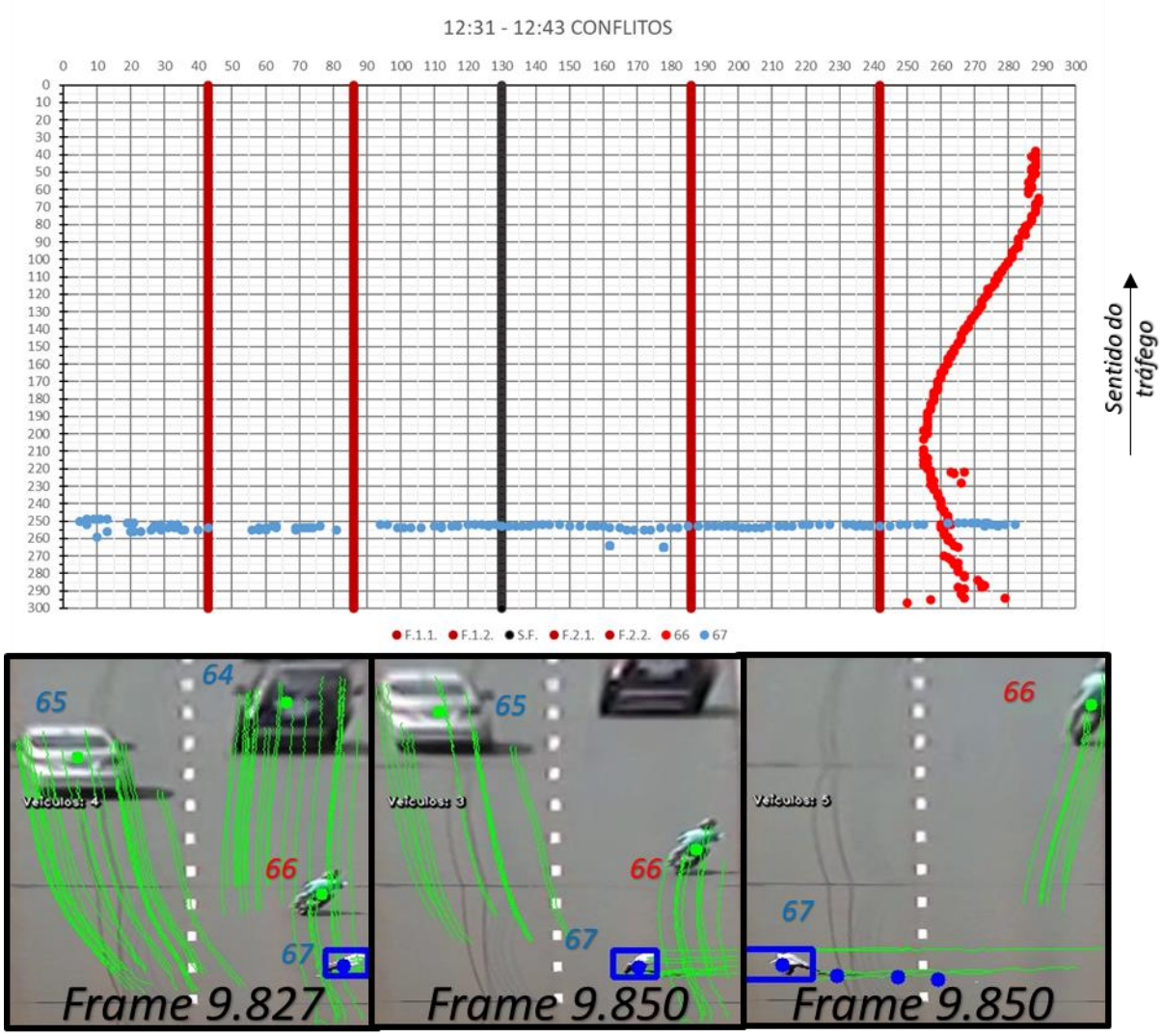
Figura 54: Exemplo das interações entre os veículos do sétimo vídeo coletado (12:05 até às12:18) representando 12 minutos de observação.



Interação entre o veículo 128 e o veículos 129. Os veículos Não-Motocicletas estão representados em azul e os veículos Motocicletas estão representados em vermelho. O gráfico distingue os veículos por tipo. A linha vertical preta "S.F." representa a separação das faixas e as linhas verticais vermelhas representam a divisão de ambas as faixas em três partes. Fonte: O autor.

Como a coleta ocorreu em um trecho urbano foi possível coletar e observar o transitar de outros objetos que não são classificados como veículos, como animais e pedestres cruzando a via de tráfego. Na Figura 55 é possível observar o deslocamento de um pedestre que ao atravessar a rua é classificado como não motocicleta e seus dados são atribuídos ao veículo de ID 67. Nesta mesma imagem é possível observar que a presença deste transeunte provoca um deslocamento lateral no veículo de ID 66 classificado corretamente como Motocicleta.

Figura 55: Exemplo das interações entre os veículos do nono vídeo coletado (12:31 até às12:43) representando 12 minutos de observação.



Interação entre o veículo 66 e o veículos 67. Os veículos Não-Motocicletas estão representados em azul e os veículos Motocicletas estão representados em vermelho. O gráfico distingue os veículos por tipo. A linha vertical preta "S.F." representa a separação das faixas e as linhas verticais vermelhas representam a divisão de ambas as faixas em três partes. Fonte: O autor.

Para poder avaliar o módulo de Identificação Automática do Comportamento do Tráfego – IACT, a Tabela 12 foi gerada. Verifica-se que tirando três casos em particular a diferença entre a contagem realizada manualmente, pelo autor, e a contagem realizada automaticamente, pelo sistema proposto por MARCOMINI (2018) é inferior a 10 %. Após uma análise mais detalhada, observou-se que nestes três casos em particular, a participação de veículos grandes (ônibus e caminhões) é superior a 7%, o que pode ter ocasionado o erro na contagem automática, umas vez que em veículos de grandes dimensões o módulo proposto tende a produzir mais de um centroide por veículo.

Tabela 12:Avaliação do módulo IACT, contagem de veículos manual e a automática.

ID	VÍDEO(O-L)	TOTAL VEÍCULOS		RELAÇÃO (AUTOM. - MANUAL)	
		MANUAL	AUTOMÁTICA	dif.	%
1	2017/18/04/11:02	134	189	55	41.045
2	2017/18/04/11:14	173	214	41	23.699
3	2017/18/04/11:27	169	182	13	7.6923
4	2017/18/04/11:40	156	166	10	6.4103
5	2017/18/04/11:52	152	170	18	11.842
6	2017/18/04/12:05	174	182	8	4.5977
7	2017/18/04/12:18	162	172	10	6.1728
8	2017/18/04/12:31	178	172	-6	-3.3708
9	2017/18/04/12:43	164	158	-6	-3.6585
10	2017/18/04/12:56	191	172	-19	-9.9476
11	2017/18/04/13:09	165	149	-16	-9.697
12	2017/18/04/13:22	145	140	-5	-3.4483
13	2017/18/04/13:35	179	173	-6	-3.352
14	2017/18/04/13:48	162	149	-13	-8.0247
15	2017/18/04/14:01	192	184	-8	-4.1667

Fonte: O autor.

Para poder avaliar o módulo de Classificação de Motocicletas desenvolvido - CM, a Tabela 13 foi gerada. Observa-se que no experimento o classificador utilizado teve um desempenho muito próximo do classificador desenvolvido no *capítulo 2* deste trabalho.

Tabela 13: Avaliação do módulo CM. Comparação entre a métrica de exatidão do classificador desenvolvido no capítulo 2 com classificador utilizado no experimento.

ID	VÍDEO(O-L)	EXATIDÃO		
		TESTE 1	TESTE 2	TESTE 3
1	11:02	*	*	0.93
2	11:14	0.96	*	*
3	11:27	0.94	*	*
4	11:40	*	0.86	*
5	11:52	0.95	*	*
6	12:05	*	0.94	*
7	12:18	*	*	0.94
8	12:31	0.93	*	*
9	12:43	*	*	0.94
10	12:56	*	0.92	*
11	13:09	*	0.97	*
12	13:22	*	*	0.95
13	13:35	0.97	*	*
14	13:48	*	0.95	*
15	14:01	*	*	0.97
MÉDIA TESTE		0.95	0.93	0.94
MÉDIA AVALIAÇÃO		0.94		
DESEMPENHO CLASSIFICADOR *CAPÍTULO 2*		EXATIDÃO		
		0.92		

Os valores em destaque representam ovídeo que apresentou melhor desempenho por grupo de teste.
Fonte: O autor.

Para poder avaliar o sistema proposto como um todo (associação dos módulos IACT – CM), a Tabela 14 foi gerada. Percebe-se que o tempo de processamento da associação dos módulos demora em média sete vezes mais que a duração real do vídeo utilizado.

Tabela 14: Avaliação do sistema proposto como um todo (associação dos módulos IACT – CM). Tempo de processamento.

ID	VÍDEO(O-L)	DURAÇÃO [hh:mm:ss]		DURAÇÃO [segundos]		PROCES. / VÍDEO
		VÍDEO	PROCESSAMENTO	VÍDEO	PROCESSAMENTO	
1	11:02	00:12:42.26	01:29:19.54	762.26	5,359.54	7.0
2	11:14	00:12:41.19	01:28:04.50	761.19	5,284.50	6.9
3	11:27	00:12:41.69	01:33:11.88	761.69	5,591.88	7.3
4	11:40	00:12:41.69	01:32:32.91	761.69	5,552.91	7.3
5	11:52	00:12:41.19	01:27:16.66	761.19	5,236.66	6.9
6	12:05	00:12:48.70	01:28:24.60	768.7	5,304.60	6.9
7	12:18	00:12:45.70	01:18:55.05	765.7	4,735.05	6.2
8	12:31	00:12:44.20	01:26:02.34	764.2	5,162.34	6.8
9	12:43	00:11:44.20	01:16:18.59	764.2	4,578.59	6.0
10	12:56	00:12:51.20	01:27:35.63	771.2	5,255.63	6.8
11	13:09	00:13:04.72	01:28:29.55	784.72	5,309.55	6.8
12	13:22	00:13:03.22	01:14:39.44	783.22	4,479.44	5.7
13	13:35	00:12:53.71	01:34:04.27	773.71	5,644.27	7.3
14	13:48	00:12:41.19	01:23:54.11	761.19	5,034.11	6.6
15	14:01	00:12:41.19	01:17:03.36	761.19	4,623.36	6.1
			MÉDIA	767	5,143.50	6.7

Fonte: O autor.

5. CONSIDERAÇÕES FINAIS

Esta pesquisa em particular se fundamenta na hipótese de que é possível identificar motocicletas dentro de uma corrente de tráfego utilizando técnicas de Visão computacional. Desta forma, a meta desta pesquisa de mestrado é, através da e utilização de um sistema de Identificação Automática do Comportamento do Tráfego a Partir de Imagens de Vídeo e sua associação com um Classificador SVM de Imagens de Motocicletas, identificar motociclistas de maneira automática em imagens de vídeos e realizar uma análise exploratória dos dados obtidos em busca de comportamentos de condução de motociclistas brasileiros.

Para isto foram propostos os seguintes objetivos principais:

1) Criar um Banco de Imagens de Veículos para utilizar em um Classificador de Imagens;
2) Desenvolver um Classificador de motocicletas a partir de imagens de veículos;
3) Identificar motocicletas em vídeos de tráfego utilizando um Classificador de Imagens;
4) Coletar informações de vídeos de tráfego como velocidade, aceleração e tipo de veículos (moto ou não moto) presentes no trânsito e
5) Analisar o comportamento dos motociclistas a partir dos dados veiculares obtidos.

2.1. CONCLUSÕES

A pesquisa em questão permitiu a criação de um banco de imagens de veículos contendo imagens extraídas a partir de vídeos de monitoramento produzindo exemplos nítidos, bem preciso e com visão posterior de motocicletas em meio ao fluxo urbano de uma cidade brasileira. O banco de dados concebido é constituído de (i) 721 imagens de motos (classe “M”), (ii) 13.393 imagens de exemplos de qualquer outro tipo de veículo ou parte de veículo (classe “NM”) e (iii) 50.481 imagens de exemplos da paisagem de fundo (classe “F”).

Como grande contribuição, a pesquisa possibilitou o desenvolvimento e calibração de um classificador de motocicletas, utilizando um critério linear de separação para treinar o classificador (algoritmo *LinearSVC*) em conjunto com vetores de características criados a partir de histogramas em escala de cinza de imagens de veículos produzidos pela função LBP (*local_binary_pattern*). A calibração do classificador escolhido para dar continuidade a pesquisa foi o Cenário de Classificação CC 19, obtido pela alteração dos parâmetros C = 150 e Tol = 0,01, mantendo os demais parâmetros em sua configuração padrão. Este modelo foi o que apresentou uma das melhores métricas de AUC (0,96) e valores de TVP (0,92), TFP (0,09), TVN (0,91), Precisão (0,91) e Exatidão (0,92) considerados como excelente segundo a literatura (FAWCETT, 2006; PRATI; BATISTA; MONARD, 2008; MINH; SANO; MATSUMOTO, 2012;

SILVA et al., 2013b; ARRÓSPIDE; SALGADO, 2014; LUO et al., 2018; SILVA; AIRES; VERAS, 2018).

Desta forma foi possível desenvolver um classificador de veículos do tipo motocicleta com um desempenho considerado bom e que em conjunto com um sistema de Identificação Automática do Comportamento do Tráfego a Partir de Imagens de Vídeo foi capaz de identificar as motocicletas em vídeos de monitoramento do tráfego e também extrair as informações veiculares do fluxo presente nos vídeos, como velocidade, aceleração.

A análise dos dados obtidos pela utilização do sistema de Identificação Automática do Comportamento do Tráfego a Partir de Imagens de Vídeo e sua associação com o Classificador de Motocicletas Desenvolvido permitiu verificar que para o trecho e período observado a maioria dos veículos respeitam o limite de velocidade de trânsito ($V = 70\ km/h$), sendo que 75% destes devem conhecer o trecho em questão e devido a esse fato transitam próximos a $V = 50\ km/h$, sendo que este é limite de velocidade para o cruzamento semaforizado que se localiza após o trecho observado.

Analisando as classes separadamente foi possível observar uma preferência por parte das motocicletas em trafegarem na faixa da direita, nos dados representado pelo número 2, mais precisamente no seu 1/3 de faixa mais próxima ao acostamento. Entretanto os demais veículos, provavelmente devido as suas dimensões serem maiores que as das motocicletas, preferem transitar pela porção central das faixas.

O conjunto de dados obtidos permitiu observar algumas situações muito interessantes como as manobras de mudanças de faixa e manobras de ultrapassagem. Entretanto como esses comportamentos foram observados somente nas imagens ilustradas, não é possível inferir que foi encontrado um tipo de padrão de comportamento para os veículos do tipo motocicletas, uma vez que esses comportamentos observados podem ter ocorrido por uma casualidade.

É muito importante salientar que os módulos utilizados nessa pesquisa foram calibrados para a situação encontrada nos vídeos coletados, o que implica na necessidade de uma nova calibração para situações distintas, tanto no quesito fluxo da corrente de tráfego, que demanda um olhar mais apurado nos parâmetros de agrupamento das *features*, como coleta de imagens diferentes, seja por mudança no sentido do fluxo de tráfego ou simplesmente por um posicionamento diferente, altura ou inclinação, que irá alterar as dimensões da imagem padrão e todos os procedimentos que são baseados nesta imagem.

Como os principais objetivos proposta para essa pesquisa foram alcançados, é plausível afirmar que a meta da pesquisa também foi alcançada

5.2. SUGESTÕES DE TRABALHOS FUTUROS

Como a pesquisa realizada tinha como objetivo focar nos comportamentos das motocicletas o classificador desenvolvido tinha como premissa identificar somente as motocicletas dos demais veículos presentes na corrente de tráfego. Sendo assim sugere-se a elaboração de um classificador que seja capaz de identificar diversos tipos de veículos dentro da corrente de tráfego.

O conjunto de dados obtidos se mostrou capaz de elucidar vários aspectos interessantes sobre o comportamento de condução de veículos em geral. Desta forma sugere-se o desenvolvimento de uma pesquisa que analise esses comportamentos habituais, porém interessantes como as manobras de mudanças de faixa e manobras de ultrapassagem, comportamentos esses que não foram abordados nesta pesquisa e que necessitaria de uma coleta maior.

Quanto ao uso dos programas, uma sugestão é melhorar a associação entre os módulos utilizados para que a fusão dos dados (classificação com os dados de posição) seja realizada automaticamente e seja disponibilizada em um relatório único como saída do sistema proposto.

Outra sugestão é reproduzir a pesquisa realizada, porém utilizando as imagens de vídeo obtidas no sentido contrário a corrente de tráfego. Nesta sugestão em particular, seria interessante comparar a qualidade da classificação dos dois classificadores para os mesmos veículos, o desenvolvido nessa pesquisa com imagens a favor da corrente de tráfego e do novo classificador, o desenvolvido com imagens contrarias a corrente de tráfego.

APÊNDICE A

Essa seção apresenta as imagens geradas das trajetórias dos veículos separadas por vídeos e distinguidas por tipo de veículo.

Figura A. 1. Trajetórias do primeiro vídeo coletado (10:49 até às11:02) representando 12 minutos de observação.

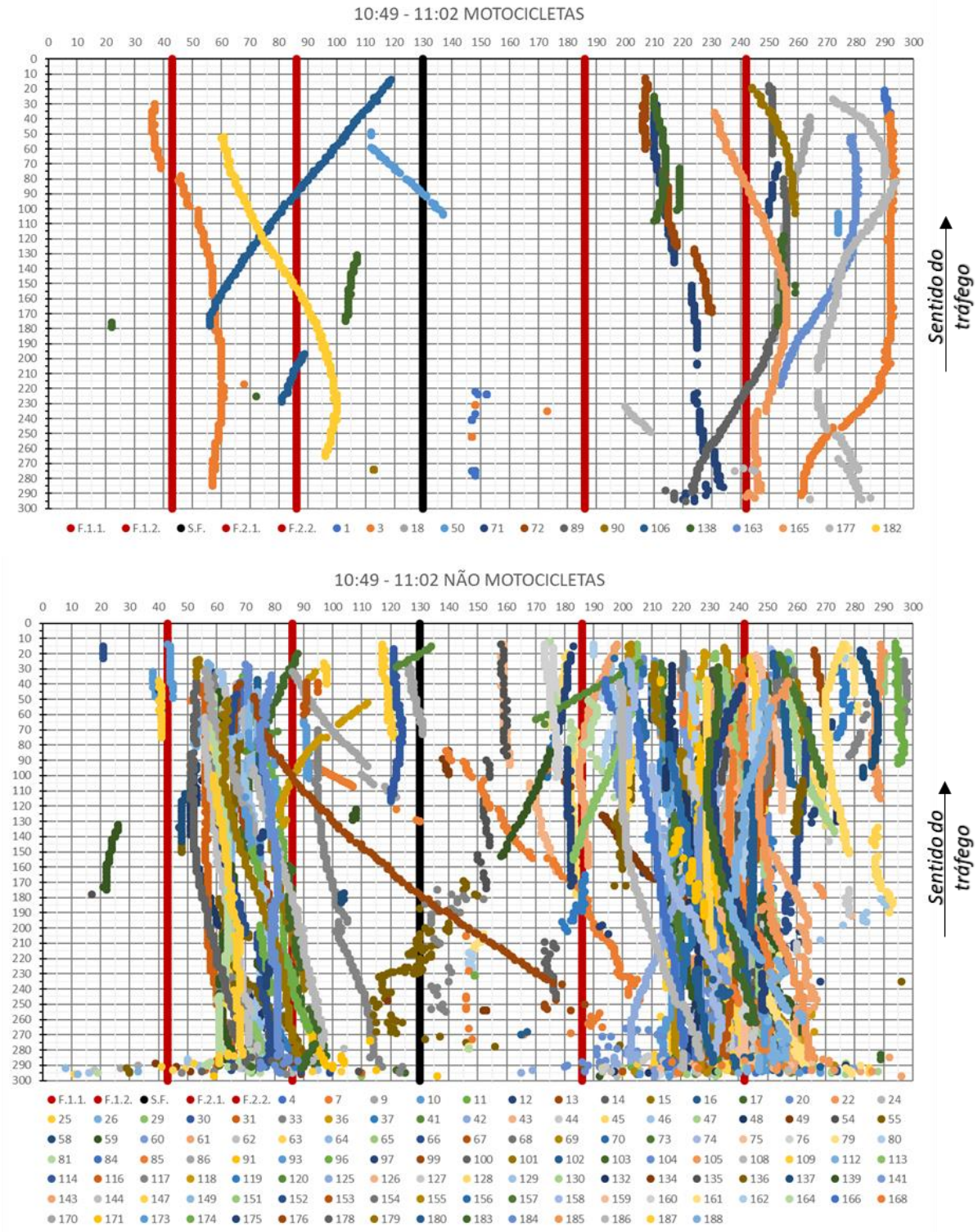


Figura A. 2. Trajetórias do segundo vídeo coletado (11:02 até às11:14) representando 12 minutos de observação.

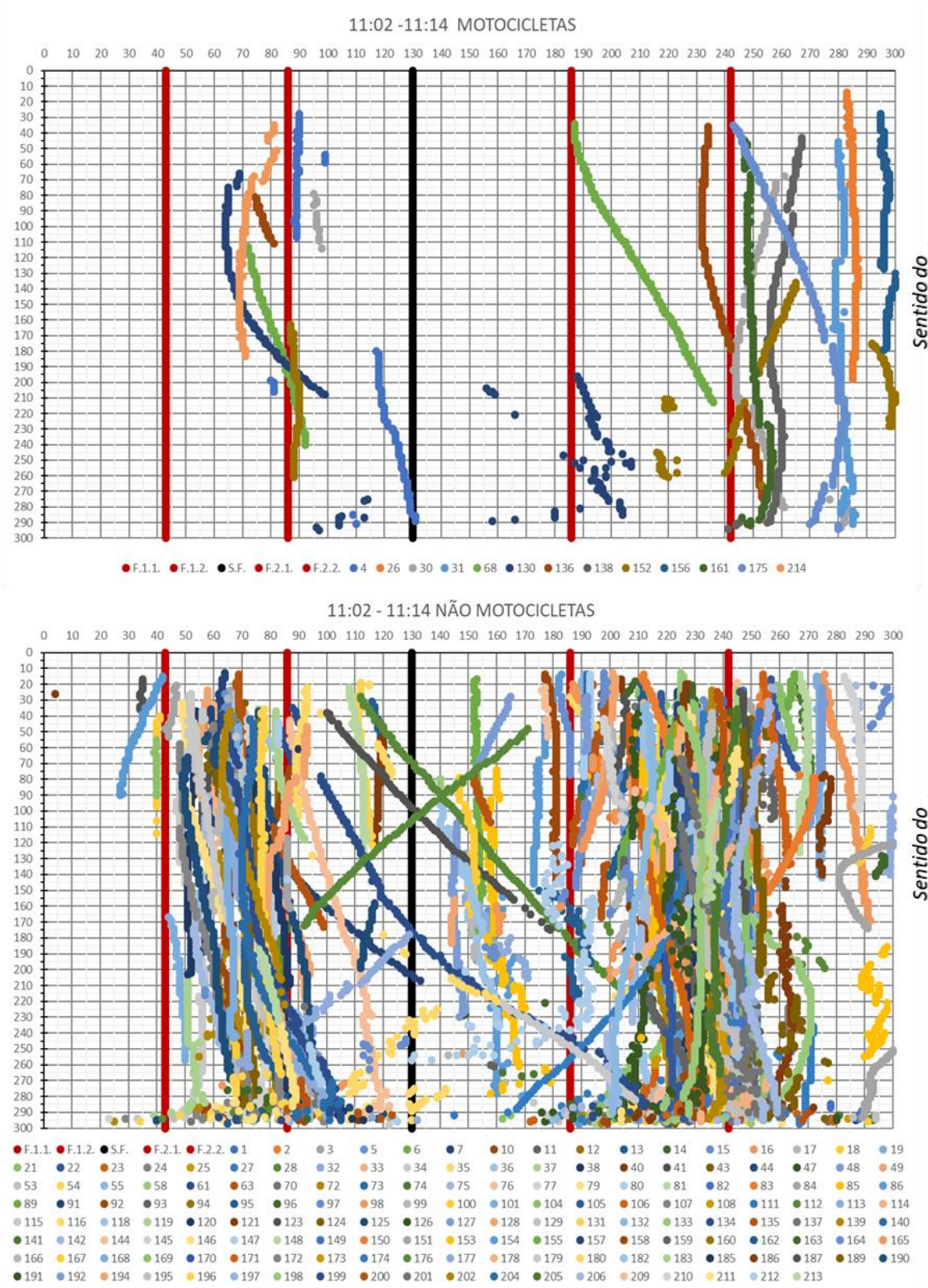


Figura A. 3. Trajetórias do terceiro vídeo coletado (11:14 até às11:27) representando 12 minutos de observação.

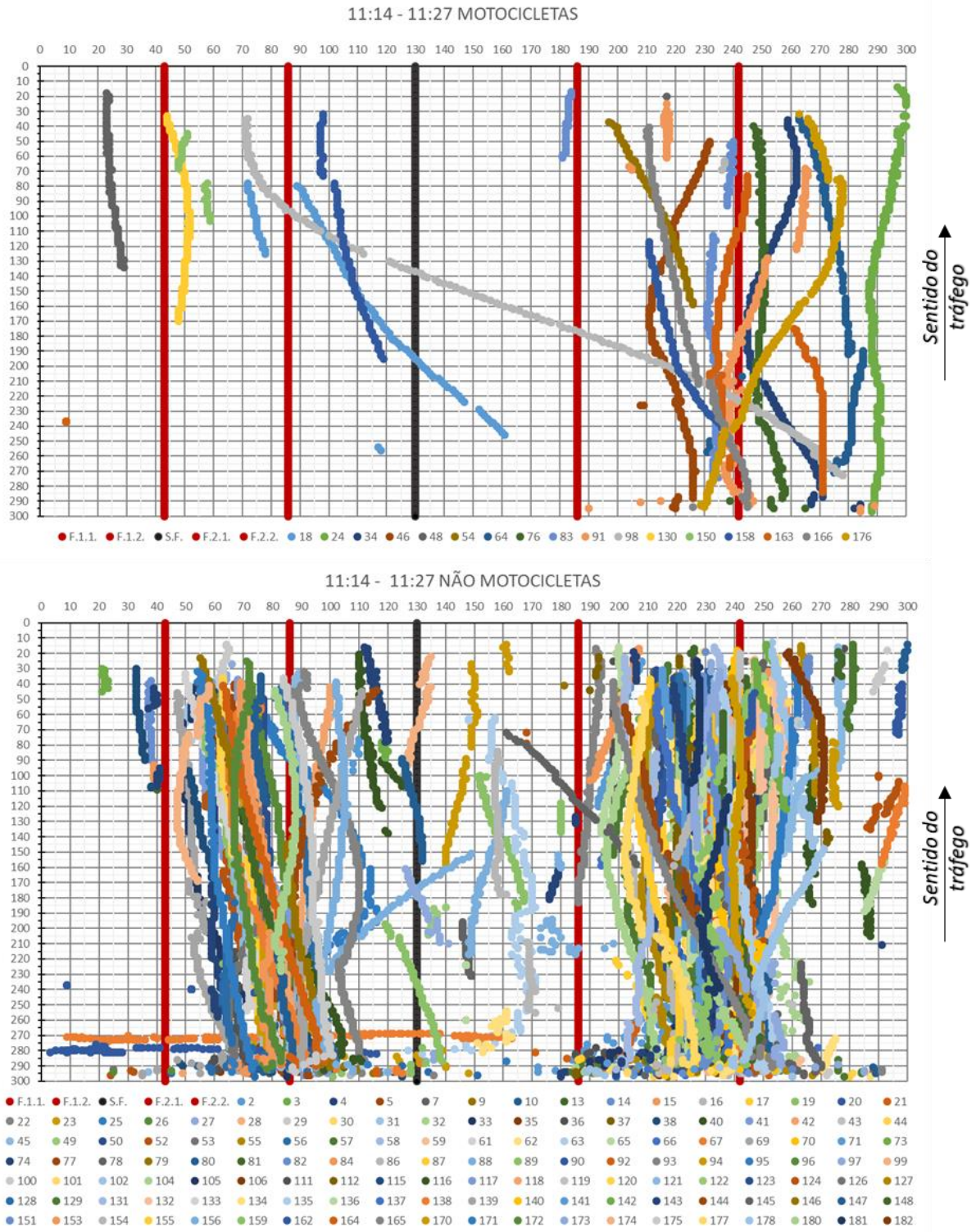
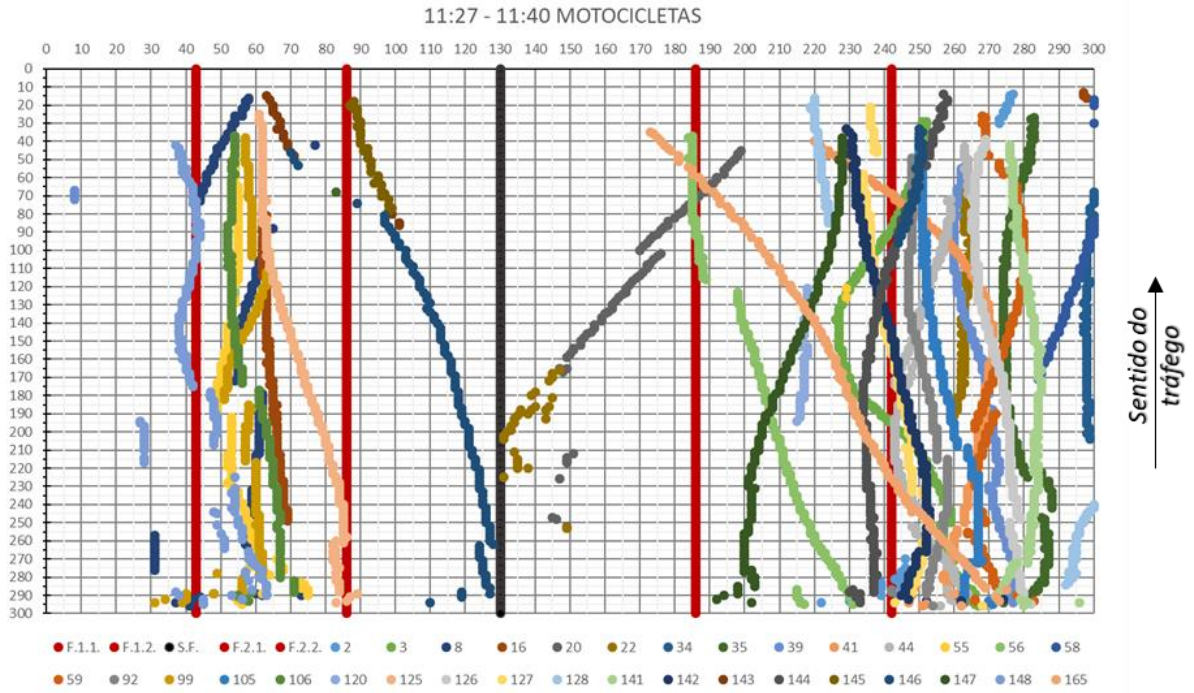


Figura A. 4. Trajetórias do quarto vídeo coletado (11:27 até às11:40) representando 12 minutos de observação.



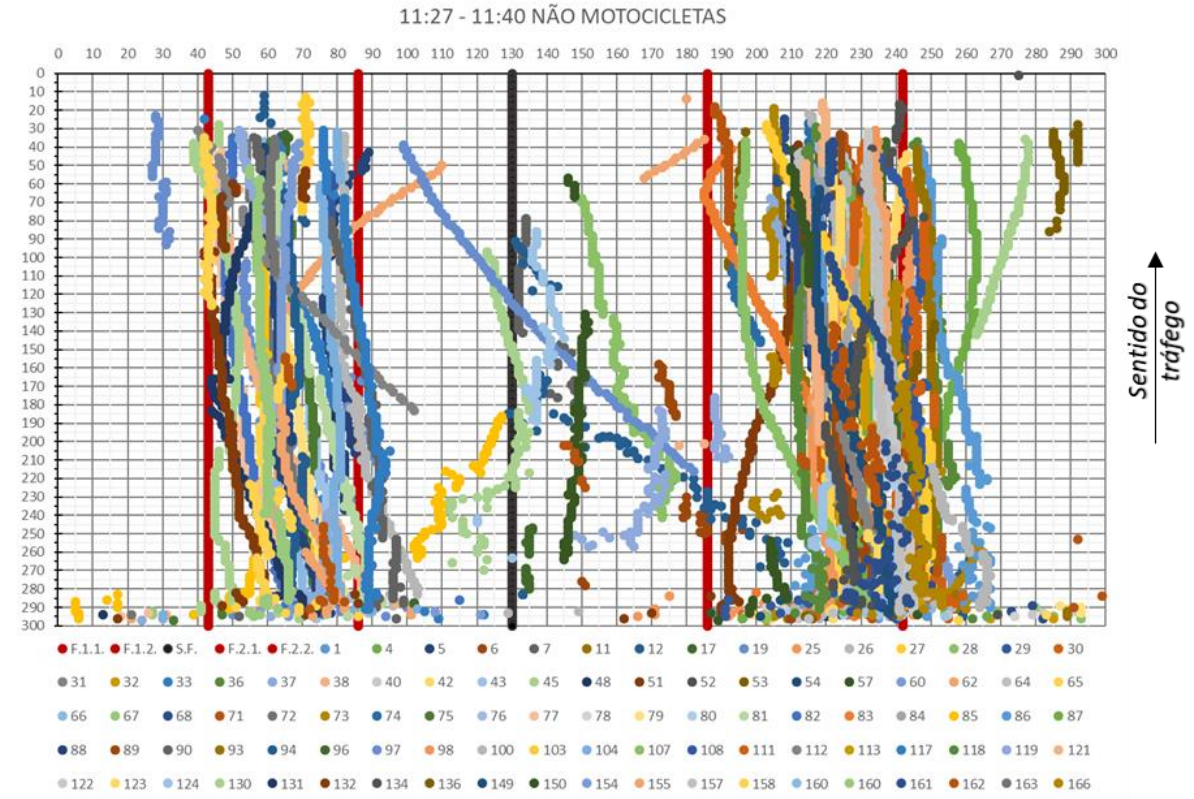


Figura A. 5. Trajetórias do quinto vídeo coletado (11:40 até às11:52) representando 12 minutos de observação.

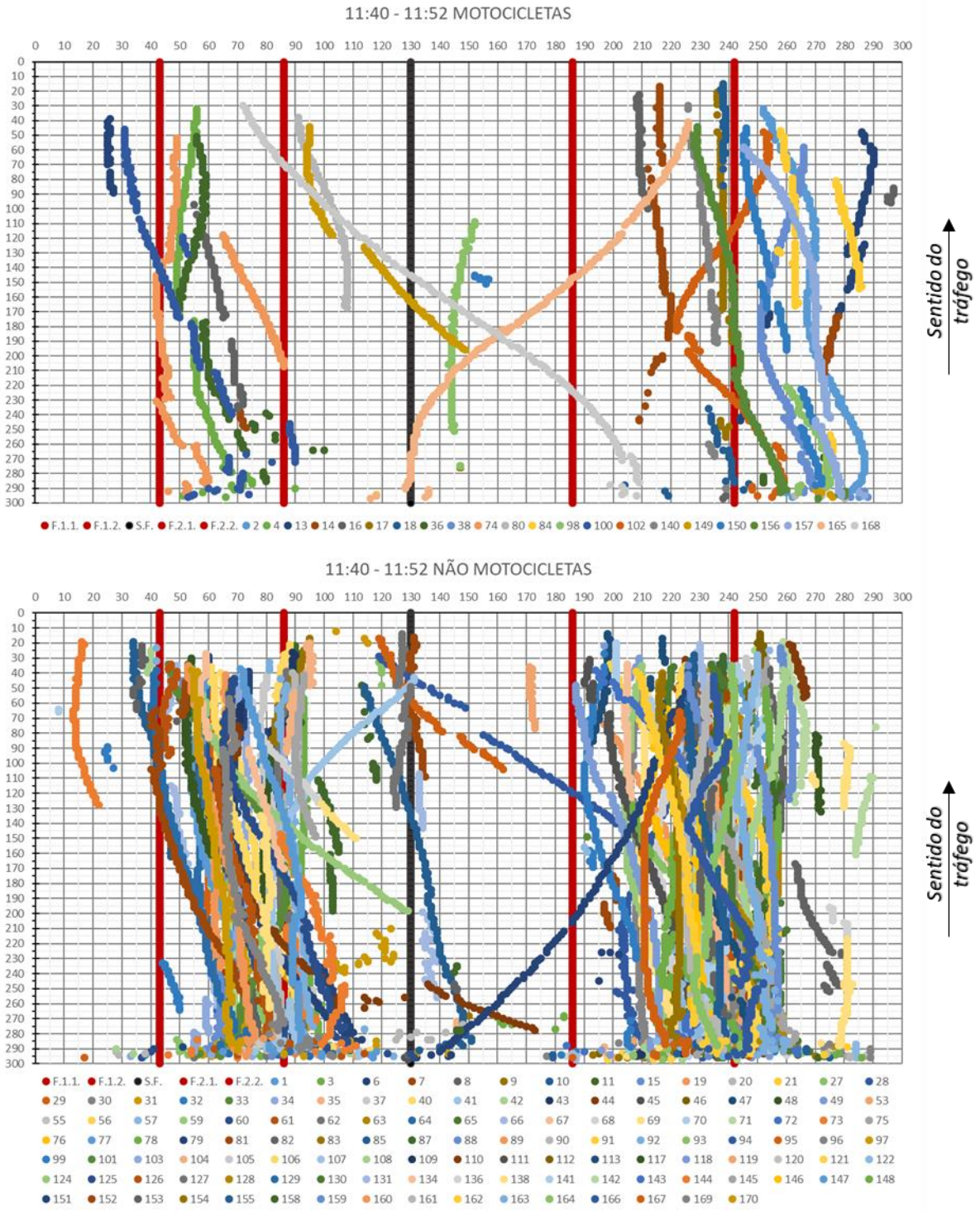
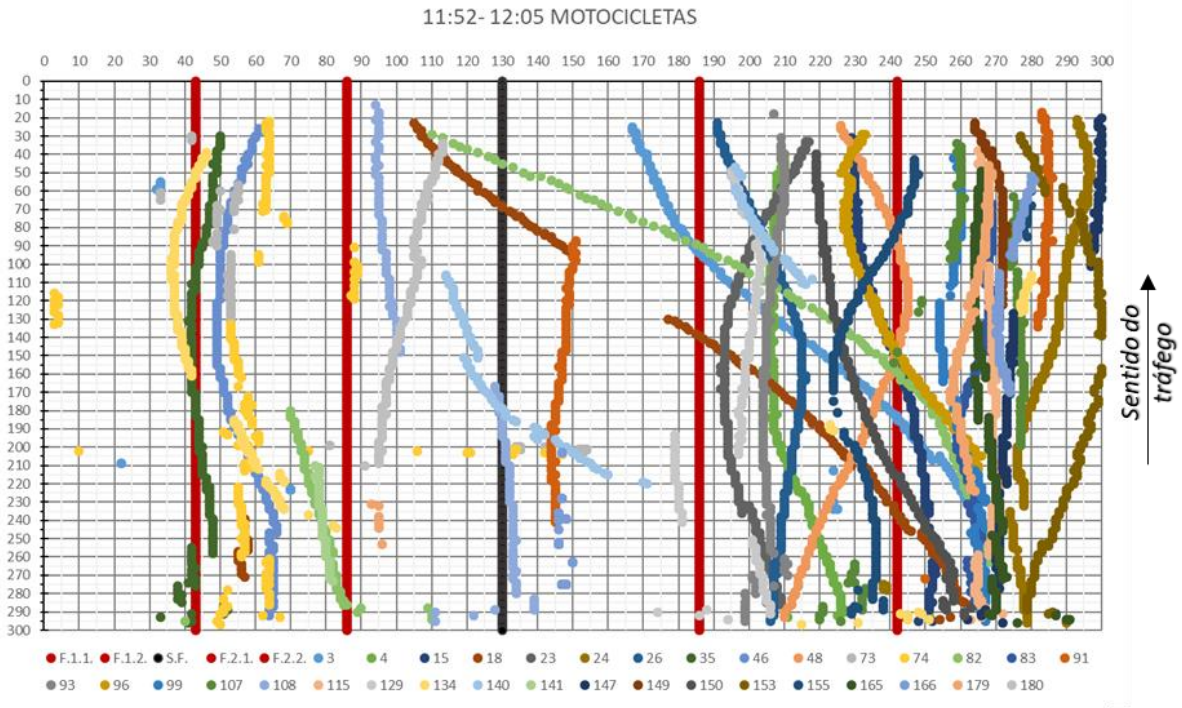


Figura A. 6. Trajetórias do sexto vídeo coletado (11:52 até às12:05) representando 12 minutos de observação.



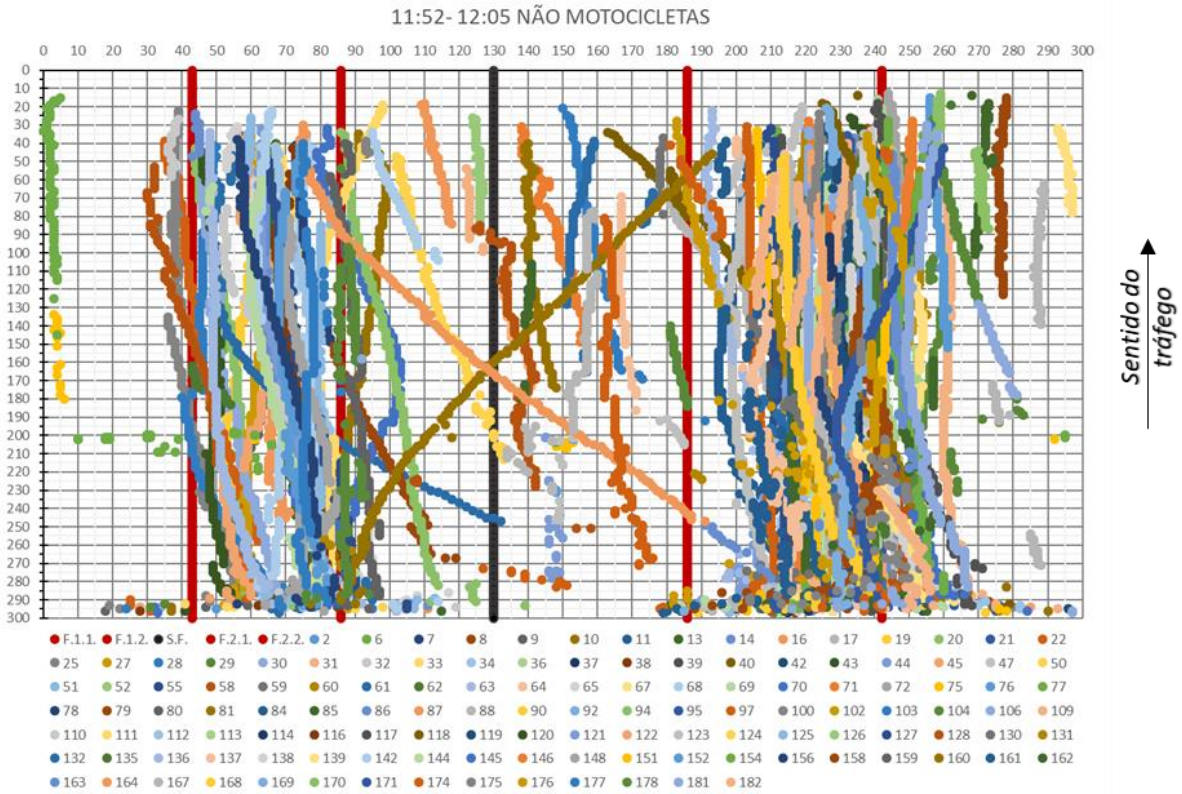


Figura A. 7. Trajetórias do sétimo vídeo coletado (12:05 até às12:18) representando 12 minutos de observação.

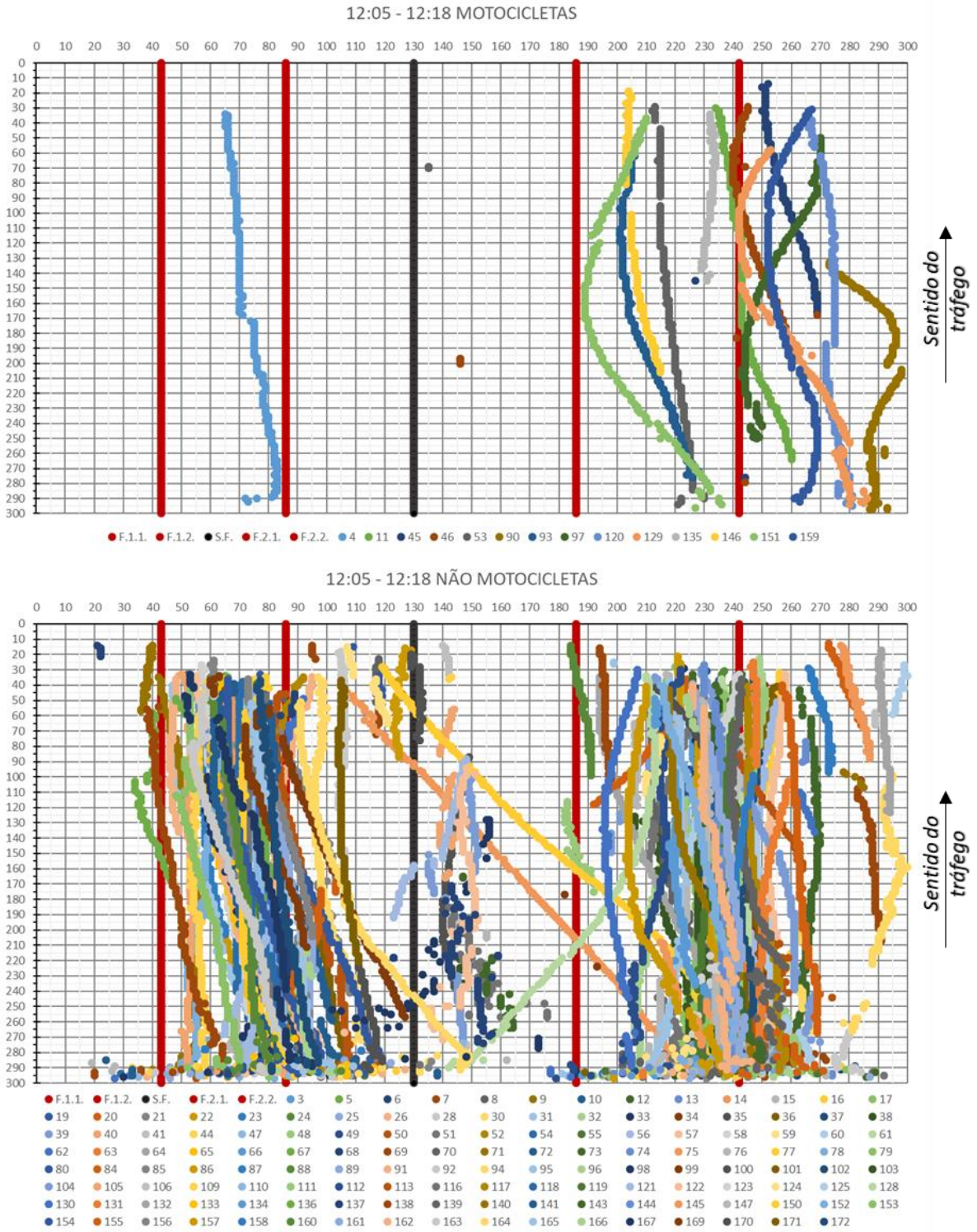


Figura A. 8. Trajetórias do oitavo vídeo coletado (12:18 até às12:31) representando 12 minutos de observação.

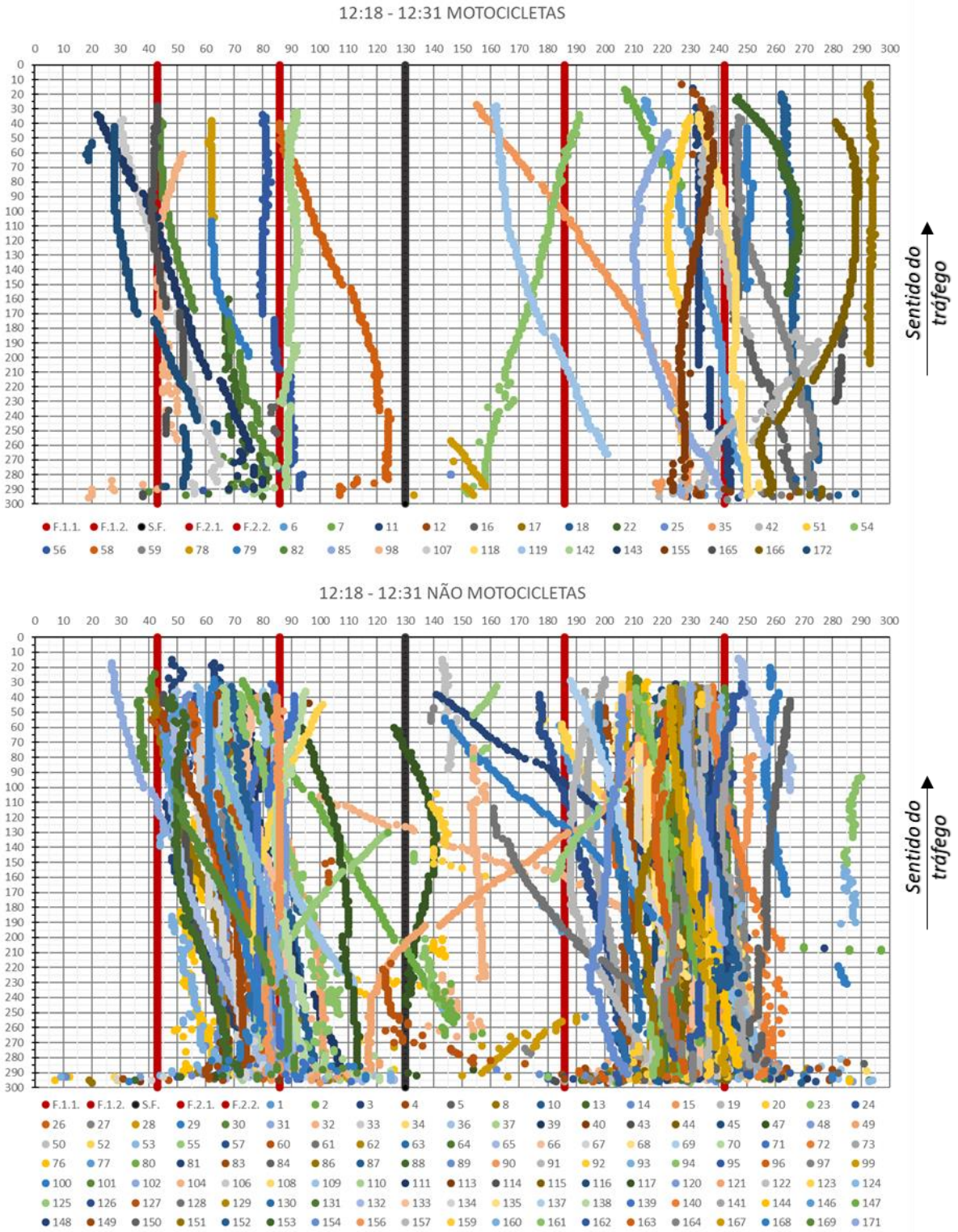


Figura A. 9. Trajetórias do nono vídeo coletado (12:31 até às12:43) representando 12 minutos de observação.

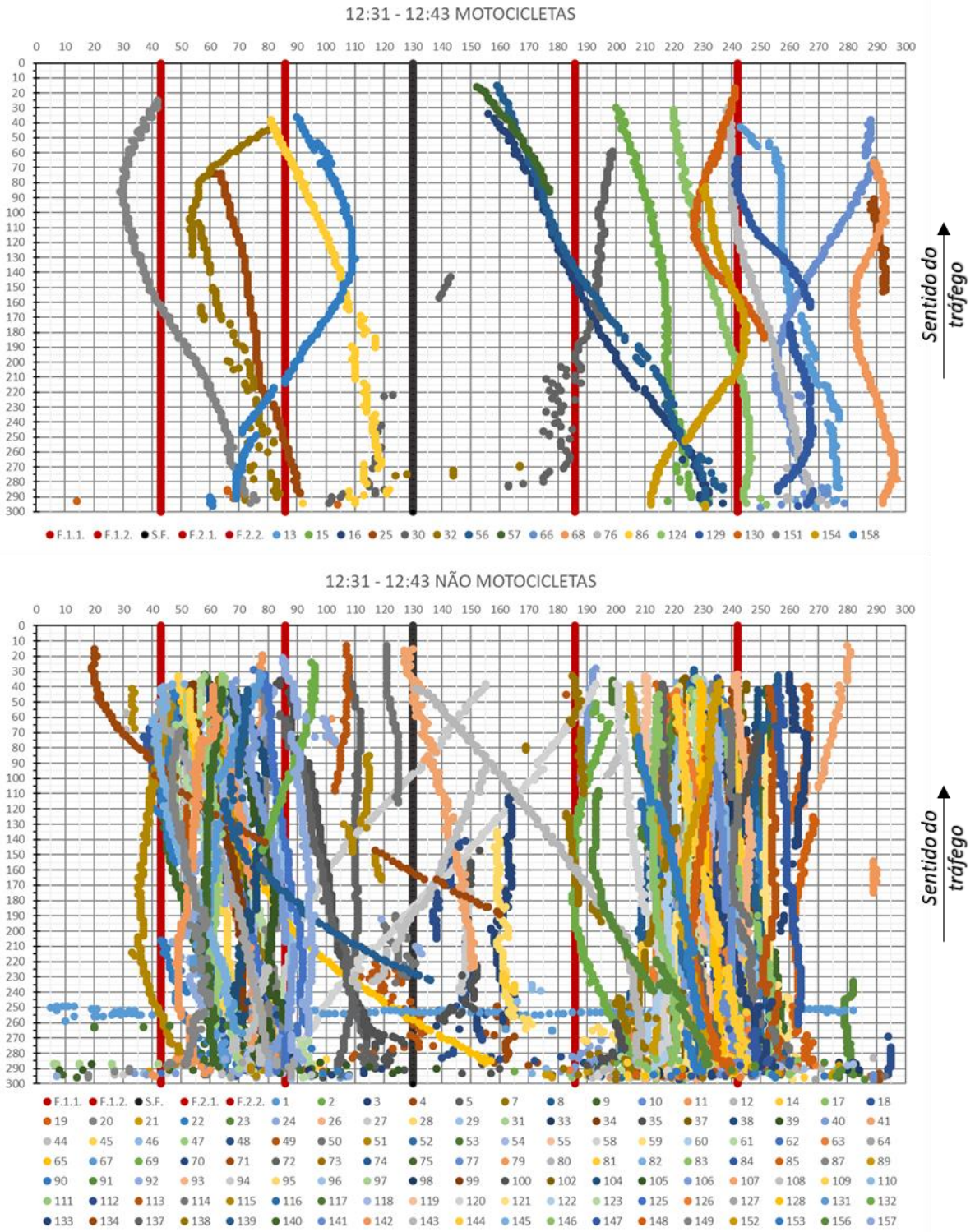


Figura A. 10. Trajetórias do décimo vídeo coletado (12:43 até às12:56) representando 12 minutos de observação.

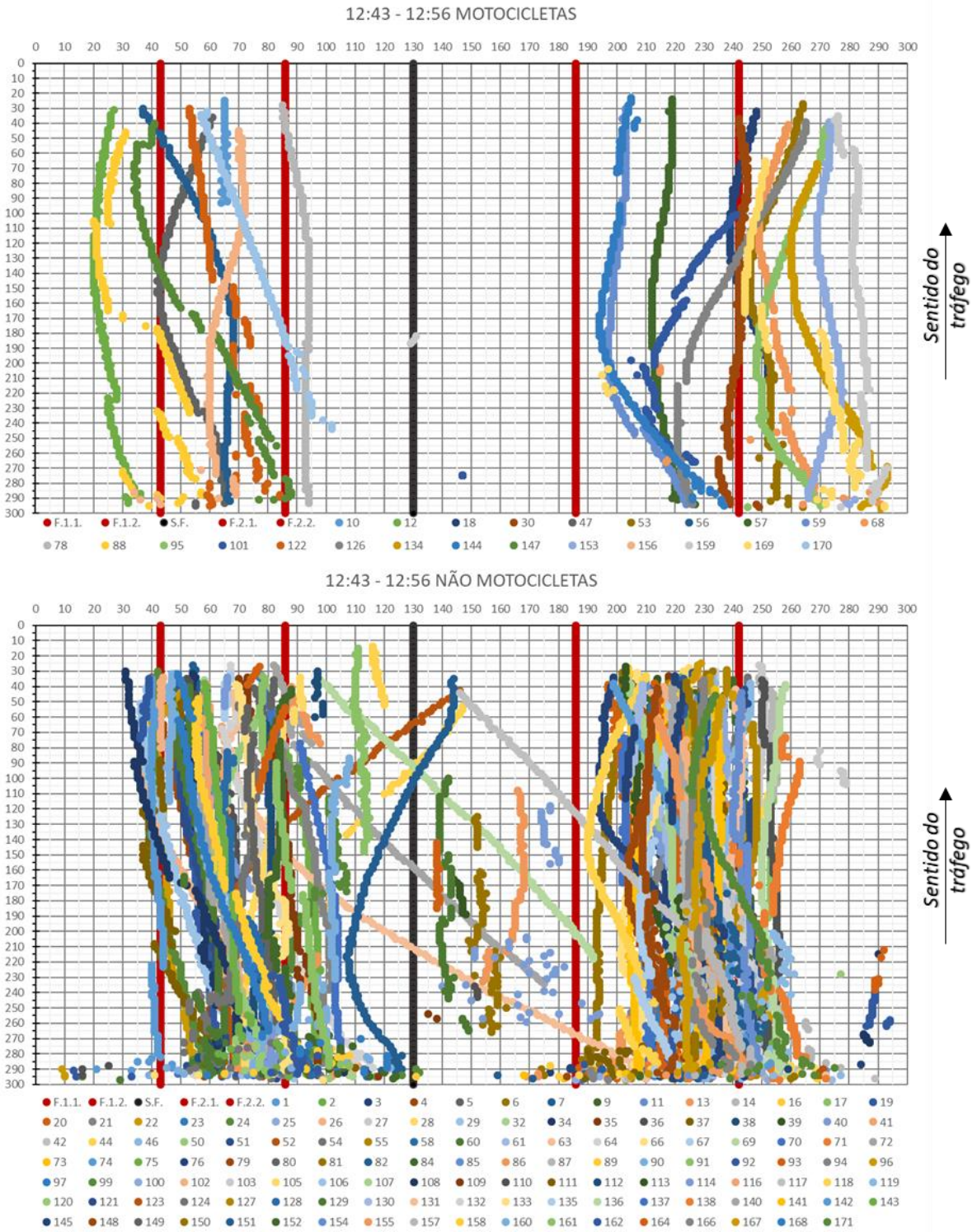
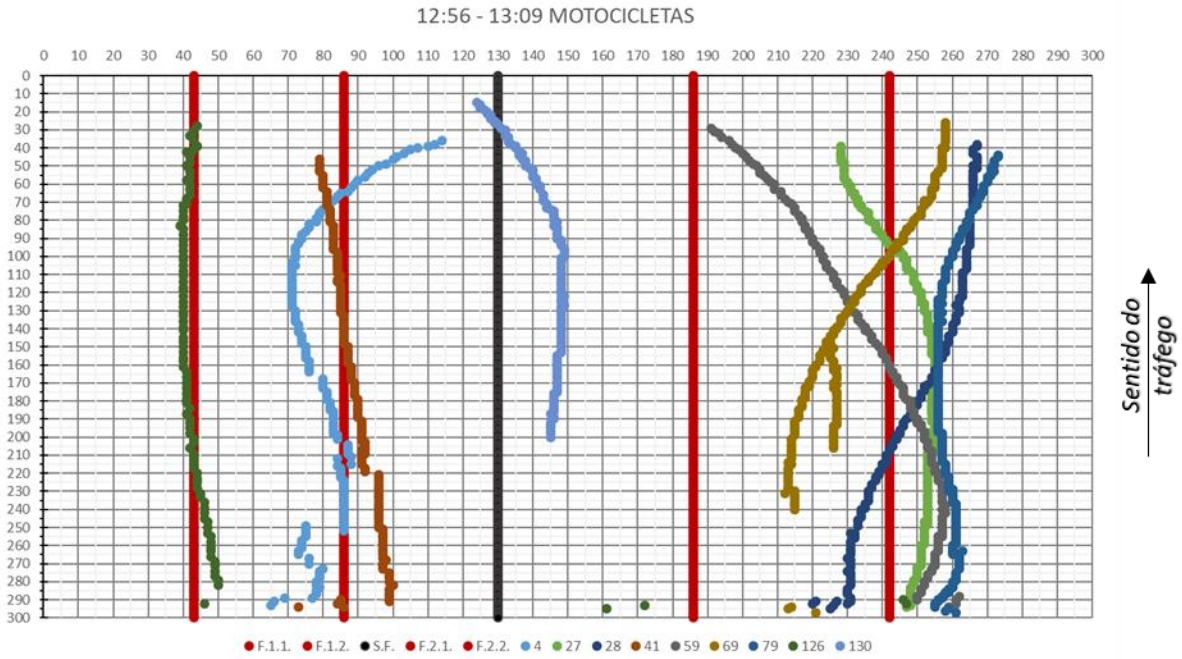


Figura A. 11. Trajetórias do décimo primeiro vídeo coletado (12:56 até às13:09) representando 12 minutos de observação.



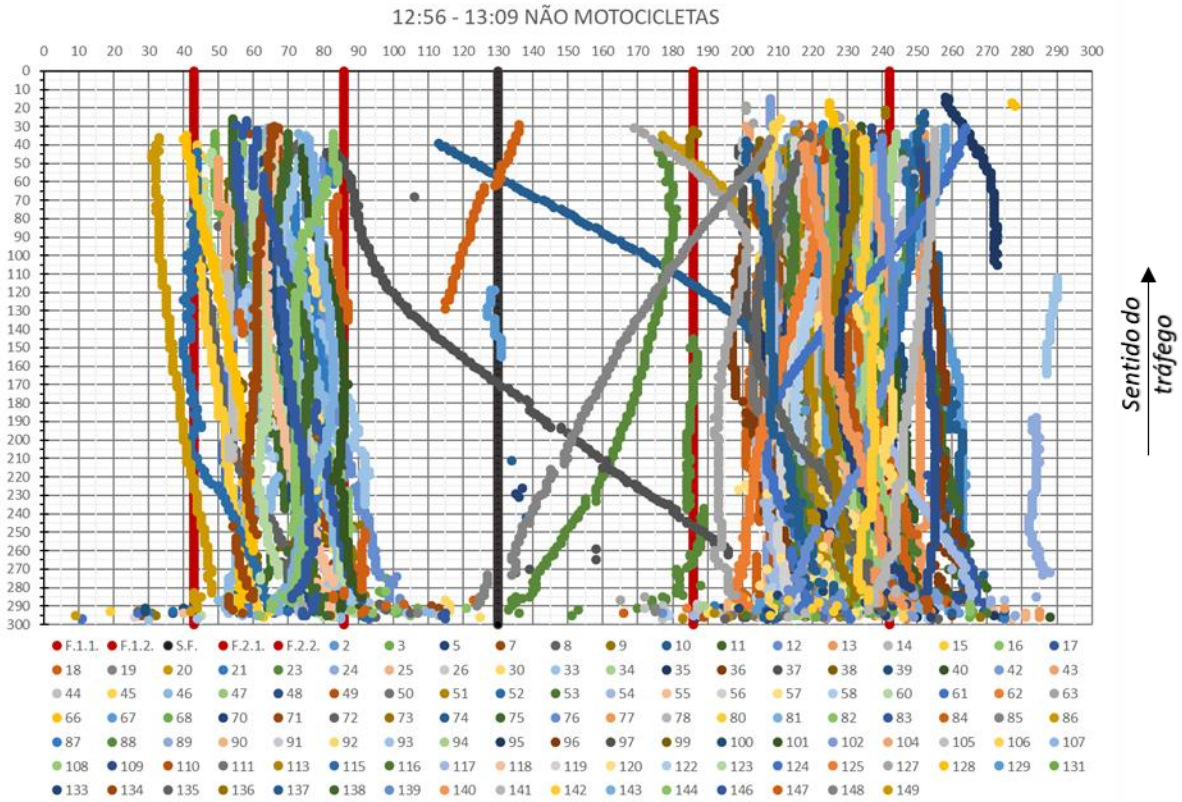


Figura A. 12. Trajetórias do decimo segundo vídeo coletado (13:09 até às13:22) representando 12 minutos de observação.

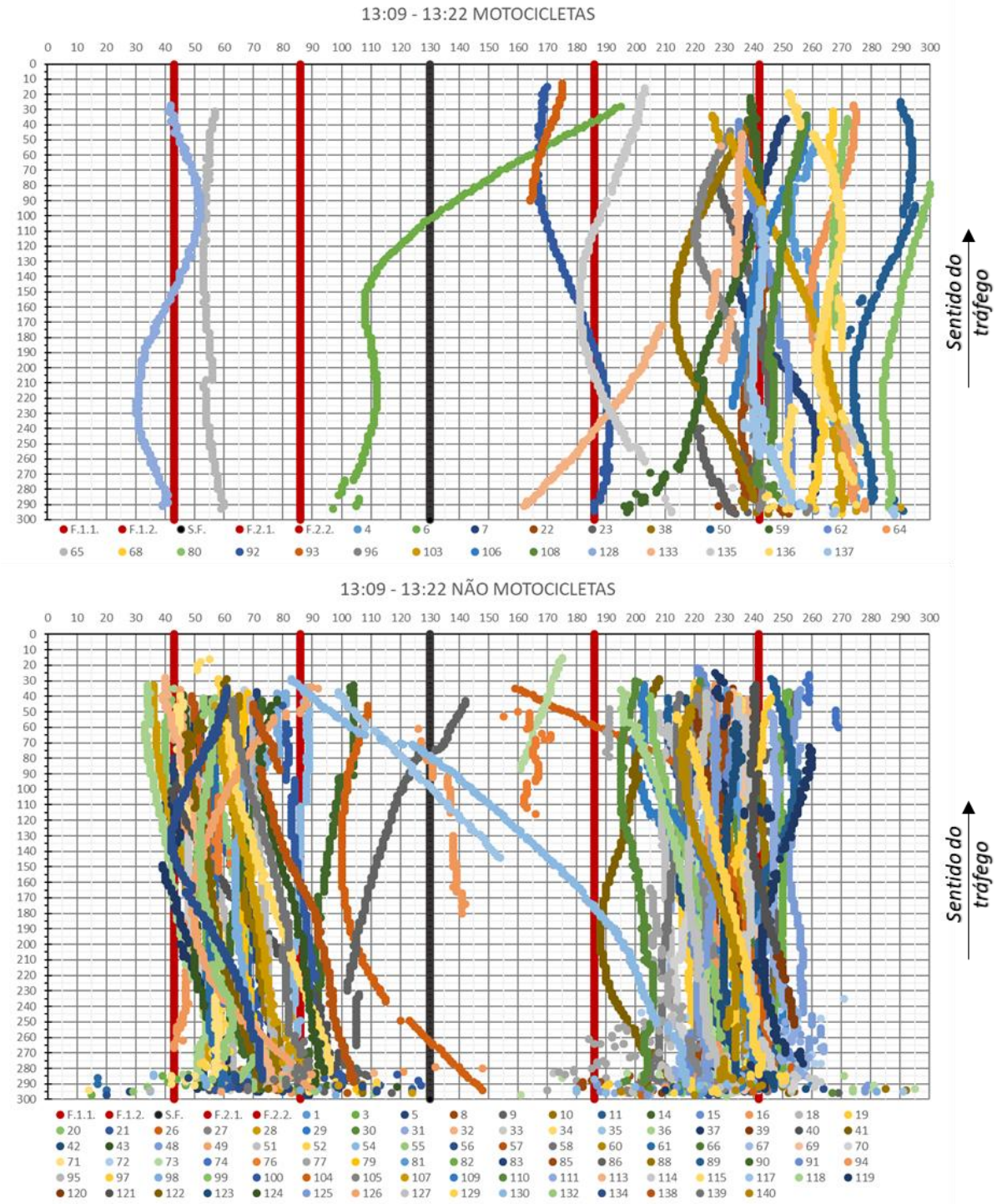


Figura A. 13. Trajetórias do decimo terceiro vídeo coletado (13:22 até às13:35) representando 12 minutos de observação.

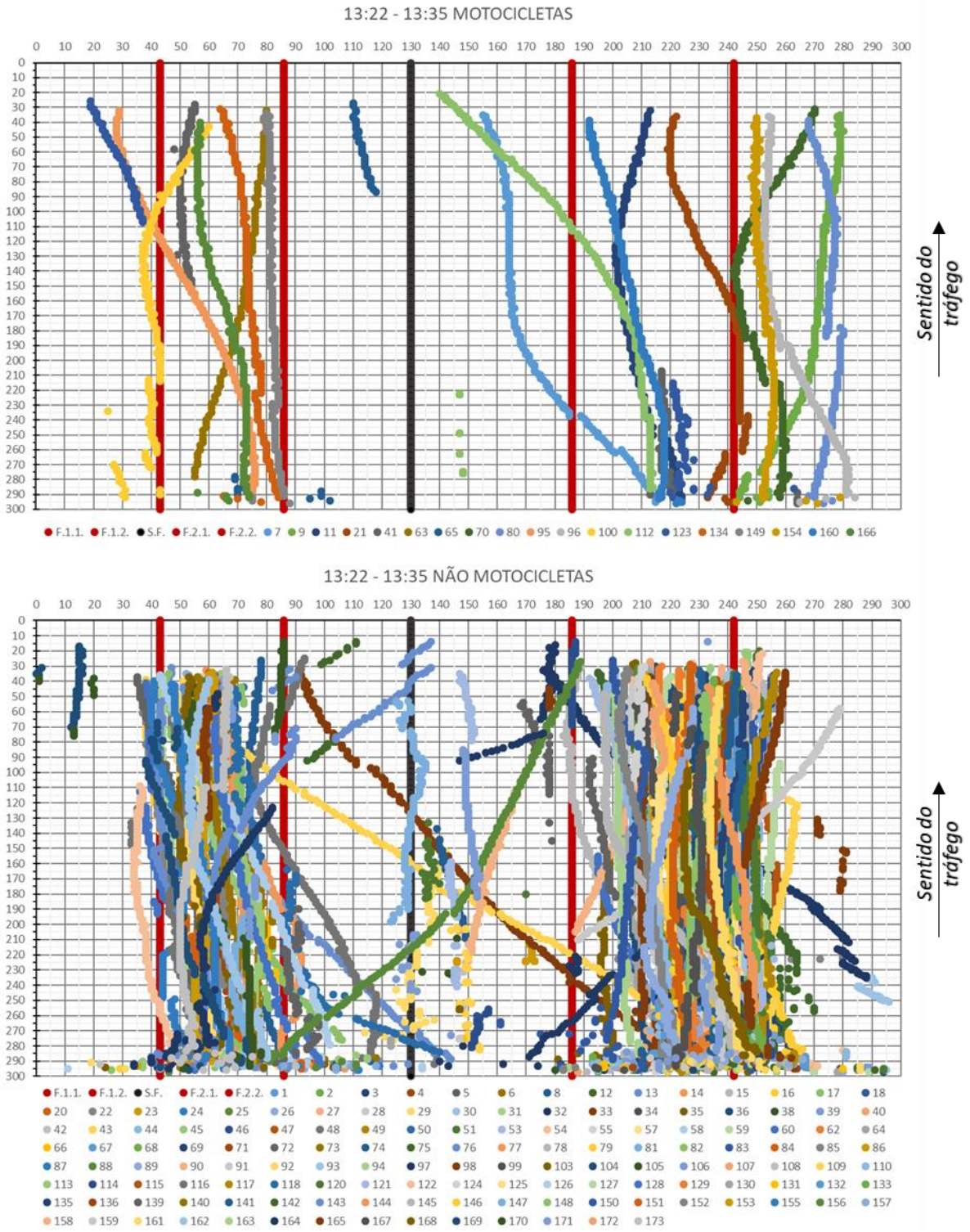


Figura A. 14. Trajetórias do decimo quarto vídeo coletado (13:35 até às13:48) representando 12 minutos de observação.

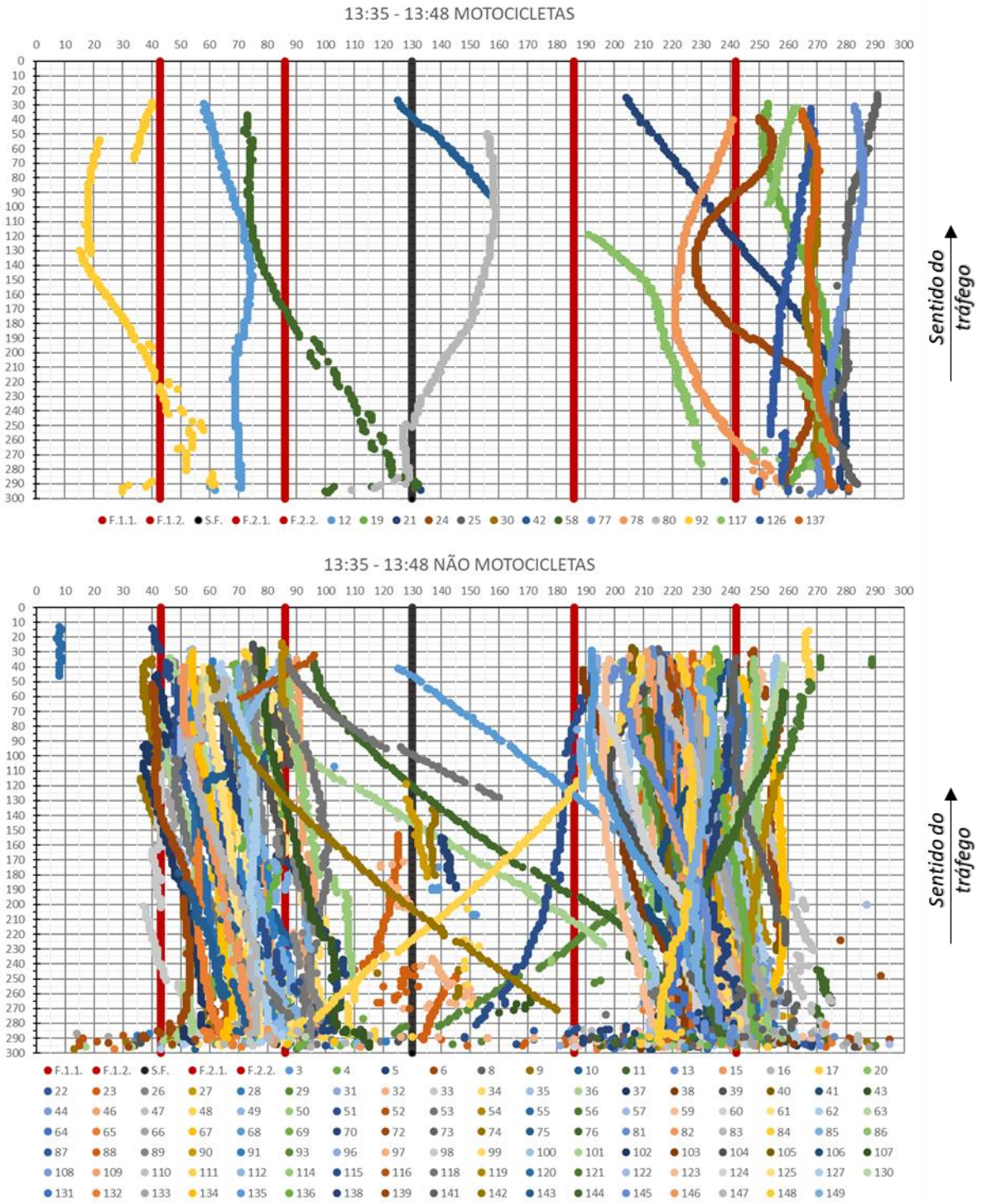
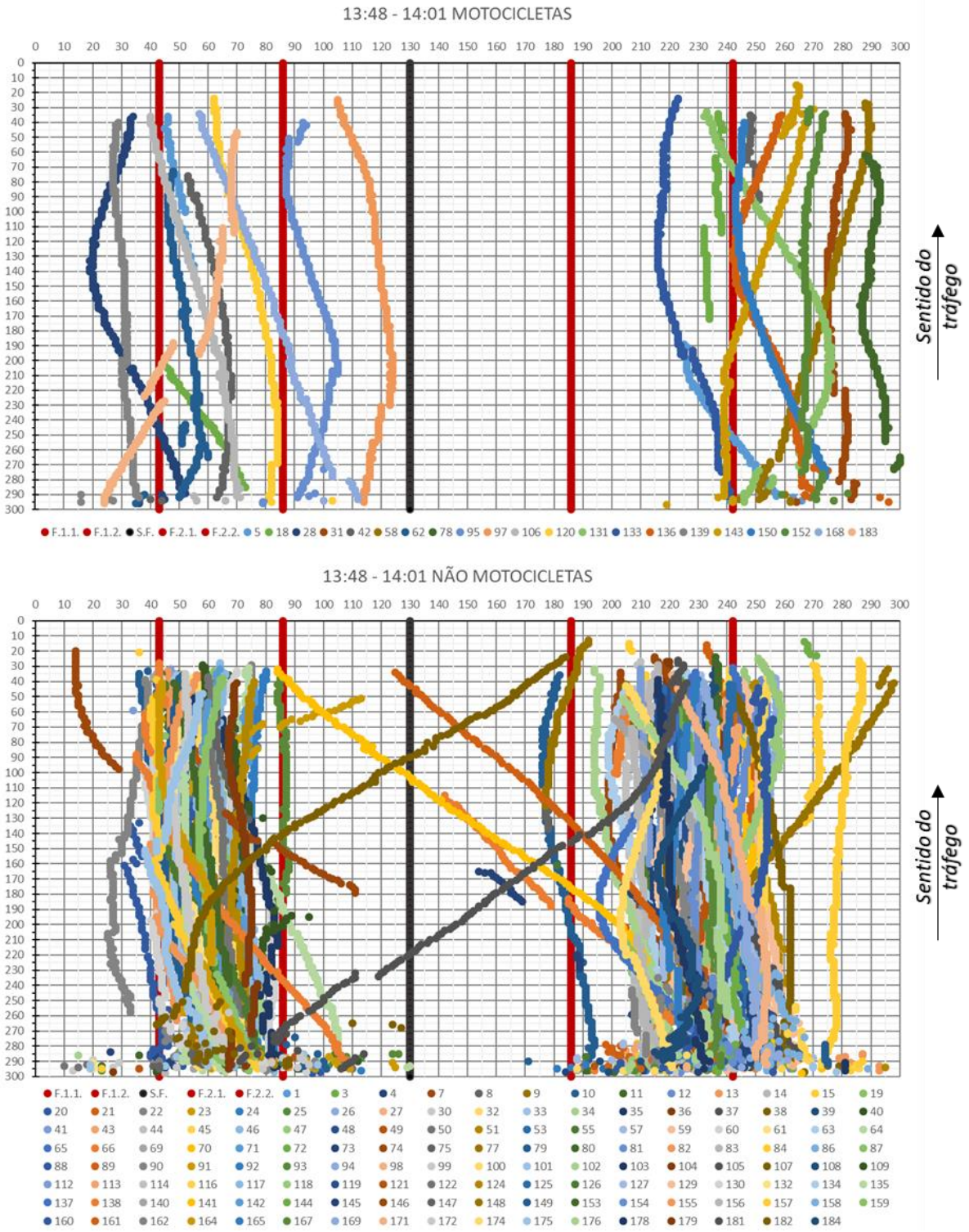


Figura A. 15. Trajetórias do decimo quinto vídeo coletado (13:48 até às14:01) representando 12 minutos de observação.



REFERÊNCIAS BIBLIOGRÁFICAS

APYB. **Python para quem está começando**. Disponível em: <https://python.org.br/>. Acesso em: 17 nov. 2018.

ARRÓSPIDE, J.; SALGADO, L. A Study of Feature Combination for Vehicle Detection Based on Image Processing. v. 2014, 2014.

BABU, F. A. M.; VORTISCH, P.; MATHEW, T. V. Modelling of motorcycle movements in mixed traffic conditions. In: 94th Annual Meeting of the Transportation Research Board, Washington, D.C. **Anais**... Washington, D.C.: 2015.

BARCELLOS, P.; BOUVIÉ, C.; ESCOUTO, F. L.; SCHARCANSKI, J. A novel video based system for detecting and counting vehicles at user-defined virtual loops. **Expert Systems with Applications**, v. 42, n. 4, p. 1845–1856, mar. 2015. Disponível em: <http://dx.doi.org/10.1016/j.eswa.2014.09.045>.

BAY, H.; TUYTELAARS, T.; VAN GOOL, L. SURF: Speeded Up Robust Features. In: [s.l: s.n.]p. 404–417.

BEYELER, M. **OpenCV with Python Blueprints**. [s.l.] Packt Publishing - Open source community distilled, 2015.

BRAGA, A. de P.; CARVALHO, A. C. P. de L. F. de; LUDERMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. 2. ed. Rio de Janeiro: Ltc Editora, 2007.

BUENO, L. M. **Análise de descritores locais de imagens no contexto de detecção de semi-réplicas**. 2011. Universidade Estatual de Campinas, 2011. Disponível em: <http://www.dca.fee.unicamp.br/~dovalle/recod/works/lucasBueno2011mscDissertation.pdf>.

CAO, L.; JIANG, Q.; CHENG, M.; WANG, C. Neurocomputing Robust vehicle detection by combining deep features with exemplar classi fi cation. **Neurocomputing**, v. 215, p. 225–231, 2016. Disponível em: <http://dx.doi.org/10.1016/j.neucom.2016.03.094>.

CORINNA, C.; VAPNIK, V. Support-Vector Networks. **Machine Learning**, v. 297, p. 273–297, 1995.

CUNHA, A. L. B. N. DA. **SISTEMA AUTOMÁTICO PARA OBTENÇÃO DE PARÂMETROS DO TRÁFEGO VEICULAR A PARTIR DE IMAGENS DE VÍDEO USANDO OPENCV**. 2013. Universidade de São Paulo, 2013.

DALAL, N.; TRIGGS, B. Histograms of Oriented Gradients for Human Detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), **Anais**...IEEE, 2005. Disponível em: <http://ieeexplore.ieee.org/document/1467360/>.

DENATRAN, D. N. D. T. **Frota de Veículos - 2018**. Disponível em: <http://www.denatran.gov.br/estatistica/261-frota-2016>. Acesso em: 19 mar. 2019.

DEVMEDIA. **Guia Completo de Python**. Disponível em:

<https://www.devmedia.com.br/guia/python/37024>. Acesso em: 20 maio. 2018.

DUDA, R. O.; HART, P. E. Use of the Hough transformation to detect lines and curves in pictures. **Communications of the ACM**, v. 15, n. 1, p. 11–15, 1 jan. 1972. Disponível em: <http://portal.acm.org/citation.cfm?doid=361237.361242>.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification**. Second Edi ed. [s.l.] Wiley, 2001.

FAWCETT, T. An introduction to ROC analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861–874, jun. 2006. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S016786550500303X>.

FERRAZ, A. C. P.; TORRES, I. G. E. **Transporte público urbano**. 2ª ed. ed. São Carlos: RIMA, 2004.

FERRAZ, C.; RAIA JR, A.; BEZERRA, B.; BASTOS, T.; RODRIGUES, K. **Segurança viária**. São Carlos - SP: Suprema Gráfica e Editora Ltda., 2012.

FORSSÉN, P.-E.; GRANLUND, G. Robust Multi-scale Extraction of Blob Features. In: BIGUN, J.; GUSTAVSSON, T. (Ed.). **Image Analysis. 13th Scandinavian Conference, SCIA Proceedings**. [s.l.] Springer-Verlag Berlin Heidelberg, 2003. p. 11–18.

FU, H.; MA, H.; LIU, Y.; LU, D. A vehicle classification system based on hierarchical multi-SVMs in crowded traffic scenes. **Neurocomputing**, v. 211, p. 182–190, out. 2016. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0925231216305732>.

GONZALEZ, R. C.; WOODS, R. E. **PROCESSAMENTO DIGITAL DE IMAGENS**. 3ª ed. São Paulo: Pearson trentice Hall, 2010.

GUOYING ZHAO; AHONEN, T.; MATAS, J.; PIETIKAINEN, M. Rotation-Invariant Image and Video Description With Local Binary Pattern Features. **IEEE Transactions on Image Processing**, v. 21, n. 4, p. 1465–1477, abr. 2012. Disponível em: <http://ieeexplore.ieee.org/document/6078431/>.

HE, Y.; DU, Y.; SUN, L.; WANG, Y. Improved waveform-feature-based vehicle classification using a single-point magnetic sensor. **Journal of Advanced Transportation**, v. 49, n. 5, p. 663–682, ago. 2015. Disponível em: <http://doi.wiley.com/10.1002/atr.1299>.

HEIKKILA, M.; PIETIKAINEN, M. A texture-based method for modeling the background and detecting moving objects. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 28, n. 4, p. 657–662, abr. 2006. Disponível em: <http://ieeexplore.ieee.org/document/1597122/>.

HOWSE, J. **OpenCV Computer Vision with Python**. [s.l.] Packt Publishing - Open source community distilled, 2013.

INFOSIGA SP. **Relatórios INFOSIGA SP**. Disponível em: <http://www.infosiga.sp.gov.br/>. Acesso em: 20 mar. 2017.

KIM, D.; DAHYOT, R. Face Components Detection Using SURF Descriptors and SVMs. In: 2008 International Machine Vision and Image Processing Conference, **Anais**...IEEE, set.

2008. Disponível em: <http://ieeexplore.ieee.org/document/4624384/>.

KIM, S.; YU, Z.; MAN, R.; LEE, M. Deep learning of support vector machines with class probability output networks. **Neural Networks**, v. 64, p. 19–28, 2015. Disponível em: <http://dx.doi.org/10.1016/j.neunet.2014.09.007>.

KLEIN, L. A. **Sensor Technologies and Data Requirementes for ITS**. USA: Artech House ITS Library, 2001.

KLEIN, L. A.; MILLS, M. K.; GIBSON, D. R. P. **Traffic Detector Handbook** (Federal Highway Administration, Ed.). [s.l: s.n.]. Disponível em: <https://www.fhwa.dot.gov/publications/research/operations/its/06108/06108.pdf>.

KOTSIANTIS, S. B.; ZAHARAKIS, I. D.; PINTELAS, P. E. Machine learning: a review of classification and combining techniques. **Artificial Intelligence Review**, v. 26, n. 3, p. 159–190, 10 nov. 2006. Disponível em: <http://link.springer.com/10.1007/s10462-007-9052-3>.

KOV, M.; YAI, T. Traffic characteristics of motorcycle-dominated urban street considering the effect of light vehicle. **Journal of the Eastern Asia Society for Transportation Studies**, v. 7, 2009. Disponível em: <http://www.easts.info/publications/journal_proceedings/journal2010/100370.pdf>.

LANGLEY, P. **Elements of Machine Learning**. San Francisco, CA: Morgan Kaufmann Inc., 1996.

LORENA, A. C.; CARVALHO, A. C. P. L. F. De. Uma Introdução às Support Vector Machines. **Revista de Informática Teórica e Aplicada**, v. 14, p. 25, 2007.

LOWE, D. G. **Object recognition from local scale-invariant featuresProc. of the International Conference on Computer Vision**. [s.l: s.n.]. Disponível em: <https://www.cs.ubc.ca/~lowe/papers/iccv99.pdf>.

LUO, Z.; B.-CHARRON, F.; LEMAIRE, C.; KONRAD, J.; LI, S.; MISHRA, A.; ACHKAR, A.; EICHEL, J.; JODOIN, P. M. MIO-TCD: A new benchmark dataset for vehicle classification and localization. **IEEE Transactions on Image Processing**, v. 27, n. 10, p. 5129–5141, 2018.

MARCOMINI, L. A. **Identificação automática do comportamento do tráfego a partir de imagens de vídeo**. 2018. Universidade de São Paulo, 2018. Disponível em: <http://www.teses.usp.br/teses/disponiveis/18/18144/tde-01102018-102649/pt-br.php>.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115–133, dez. 1943. Disponível em: <http://link.springer.com/10.1007/BF02478259>.

MINH, C. C.; SANO, K.; MATSUMOTO, S. Maneuvers of motorcycles in queues at signalized intersections. **Journal of Advanced Transportation**, v. 46, n. 1, p. 39–53, jan. 2012. Disponível em: <http://onlinelibrary.wiley.com/doi/10.1002/atr.144/full>.

MITCHELL, T. M. **Machinie Learning**. USA: WCB/McGraw-Hill, 1997.

MORGUL, E. F.; YANG, H.; KURKCU, A.; OZBAY, K.; BARTIN, B.; KAMGA, C.;

SALLOUM, R. Virtual Sensors: Web-Based Real-Time Data Collection Methodology for Transportation Operation Performance Analysis. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2442, n. 1, p. 106–116, jan. 2014. Disponível em: <http://journals.sagepub.com/doi/10.3141/2442-12>.

MUNIGETY, C.; VICRAMAN, V.; MATHEW, T. Semiautomated Tool for Extraction of Microlevel Traffic Data from Videographic Survey. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2443, p. 88–95, dez. 2014. Disponível em: <http://trrjournalonline.trb.org/doi/10.3141/2443-10>.

NAKAGAWA, K.; IWASAKI, A.; OISHI, Y.; HORISAKI, R.; TSUKAMOTO, A.; NAKAMURA, A.; HIROSAWA, K.; LIAO, H.; USHIDA, T.; GODA, K.; KANNARI, F.; SAKUMA, I. Sequentially timed all-optical mapping photography (STAMP). **Nature Photonics**, v. 8, n. 9, p. 695–700, 10 ago. 2014. Disponível em: <http://www.nature.com/doifinder/10.1038/nphoton.2014.163>.

NEUMEYER, D. **The Oxford Handbook of Film Music Studies**. [s.l.] Oxford University Press (OUP), 2013.

NGUYEN, H. D. Saturation Flow Rate Analysis at Signalized Intersections for Mixed Traffic Conditions in Motorcycle Dependent Cities. **Transportation Research Procedia**, v. 15, p. 694–708, 2016. Disponível em: <http://dx.doi.org/10.1016/j.trpro.2016.06.058>.

NGUYEN, L. X.; HANAOKA, S.; KAWASAKI, T. Traffic conflict assessment for non-lane-based movements of motorcycles under congested conditions. **IATSS Research**, v. 37, n. 2, p. 137–147, mar. 2014. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0386111213000265>.

NIXON, M. S.; AGUADO, A. S. **Feature Extraction and Image Processing**. First edit ed. Londres: Elsevier, 2002.

PIETIKÄINEN, M.; HADID, A.; ZHAO, G.; AHONEN, T. **Computer Vision Using Local Binary Patterns**. London: Springer London, 2011. v. 40

PONTIL, M.; VERRI, A. Support Vector Machines for 3D Object Recognition. **EEE Transactions on Pattern Analysis and Machine Intelligence**, v. 20, n. 6, p. 637–646, 1998. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.319.3554&rep=rep1&type=pdf>.

PRATI, R. C.; BATISTA, G. E. A. P. A.; MONARD, M. C. Evaluating Classifiers Using ROC Curves. **IEEE Latin America Transactions**, v. 6, n. 2, p. 215–222, jun. 2008. Disponível em: <http://ieeexplore.ieee.org/document/4609920/>.

RIBEIRO, E. R. **Análise exploratória de método utilizando Wavelet para detecção de padrões e anomalias em dados históricos do tráfego veicular**. 2017. Universidade de São Paulo - Escola de Engenharia de São Carlos, 2017.

SANTOS, E. M. **Teoria e Aplicação de Support Vector Machines à Aprendizagem e Reconhecimento de Objetos Baseado na Aparência**. 2002. Universidade Federal da Paraíba, 2002.

SCHMIDHUBER, J. Deep learning in neural networks : An overview. **Neural Networks**, v.

61, p. 85–117, 2015. Disponível em: <http://dx.doi.org/10.1016/j.neunet.2014.09.003>.

SCIKIT - IMAGE. Disponível em: <https://scikit-image.org/>. Acesso em: 20 maio. 2018.

SILVA, A. T. da. **Descritores de Imagem**. Disponível em: <http://www.joinville.udesc.br/portal/professores/andretavares/materiais/descritores.pdf>. Acesso em: 5 ago. 2017.

SILVA, A. T. da. **SVM (Support Vector Machine)**. Disponível em: <http://www.joinville.udesc.br/portal/professores/andretavares/materiais/RP_Aula14_SVM.pdf>. Acesso em: 7 ago. 2017.

SILVA, I. N. da; SPATTI, D. H.; FLAUZINO, R. A. **Redes Neurais Artificiais para Engenharia e Ciências Aplicadas**. São Paulo: Artliber, 2010.

SILVA, R.; AIRES, K.; SANTOS, T.; ABDALLA, K.; VERAS, R. SEGMENTAÇÃO, CLASSIFICAÇÃO E DETECÇÃO DEMOTOCICLISTAS SEM CAPACETE. In: Fortaleza. **Anais**... Fortaleza: XI Simpósio Brasileiro de Automação Inteligente (SBAI), 2013a.

SILVA, R.; AIRES, K.; VERAS, R.; SANTOS, T.; LIMA, K.; SOARES, A. Automatic Motorcycle Detection on Public Roads. **CLEI ELECTRONIC JOURNAL**, v. 16, n. 3, p. 1–9, 2013b. Disponível em: <http://www.scielo.edu.uy/scielo.php?pid=S0717-50002013000300004&script=sci_abstract&tlng=pt>.

SILVA, R. R. V. e; AIRES, K. R. T.; VERAS, R. D. M. S. Detection of helmets on motorcyclists. **Multimedia Tools and Applications**, v. 77, n. 5, p. 5659–5683, 18 mar. 2018. Disponível em: <http://link.springer.com/10.1007/s11042-017-4482-7>.

SOUZA, R. **Conheça a Linguagem Python e saiba como é possível usá-la!** Disponível em: <https://www.impacta.com.br/blog/2018/09/17/conheca-a-linguagem-python-e-saiba-como-e-possivel-usa-la/>. Acesso em: 30 set. 2018.

TAKALA, V.; AHONEN, T.; PIETIKÄINEN, M. Block-Based Methods for Image Retrieval Using Local Binary Patterns. In: **Scia 2005**. [s.l: s.n.]p. 882–891.

XIAO, J.; GAO, X.; KONG, Q.-J.; LIU, Y. More robust and better: a multiple kernel support vector machine ensemble approach for traffic incident detection. **Journal of Advanced Transportation**, v. 48, n. 7, p. 858–875, nov. 2014. Disponível em: <http://doi.wiley.com/10.1002/atr.1231>.

YU, S.-H.; HSIEH, J.-W.; CHEN, Y.-S.; HU, W.-F. An Automatic Traffic Surveillance System for Vehicle Tracking and Classification. In: BIGUN, J.; GUSTAVSSON, T. (Ed.). **IEEE Transactions on Intelligent Transportation Systems**. [s.l.] Springer-Verlag Berlin Heidelberg, 2003. 7p. 379–386.