

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
DEPARTAMENTO DE TELEMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Leandro Bezerra Marinho

**Uma nova abordagem de classificação com opção de rejeição para
localização e navegação de robôs móveis em mapas topológicos
utilizando imagens convencionais e omnidirecionais**

Fortaleza – CE

25 de maio de 2017

Leandro Bezerra Marinho

**Uma nova abordagem de classificação com opção de
rejeição para localização e navegação de robôs móveis
em mapas topológicos utilizando imagens convencionais
e omnidirecionais**

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Inteligência Artificial

Orientador: Prof. Dr. Pedro Pedrosa Rebouças Filho

Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Fortaleza – CE

25 de maio de 2017

Dados Internacionais de Catalogação na Publicação
Instituto Federal do Ceará - IFCE
Sistema de Bibliotecas - SIBI
Ficha catalográfica elaborada pelo SIBI/IFCE, com os dados fornecidos pelo(a) autor(a)

M337n Marinho, Leandro Bezerra.

Uma Nova abordagem de classificação com opção de rejeição para localização e navegação de robôs móveis em mapas topológicos utilizando imagens convencionais e omnidirecionais / Leandro Bezerra Marinho. - 2017. 135 f.

Dissertação (Mestrado Profissional) Instituto Federal do Ceará, Mestrado em Ciência da Computação, Campus Fortaleza, 2017.

Orientação: Prof. Dr. Pedro Pedrosa Rebouças Filho.

1. Robótica. 2. Mapa Topológico. 3. Localização de Robôs Móveis. I. Título.

CDD 004

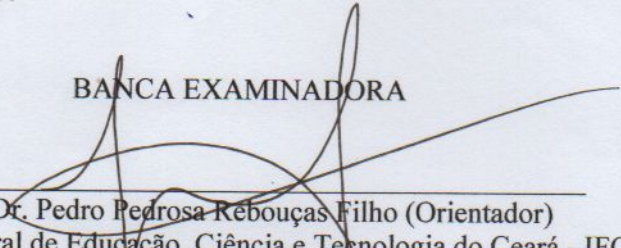
LEANDRO BEZERRA MARINHO

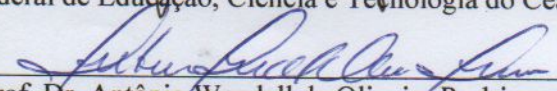
Uma nova abordagem de classificação com opção de rejeição para localização e navegação de robôs móveis em mapas topológicos utilizando imagens convencionais e omnidirecionais

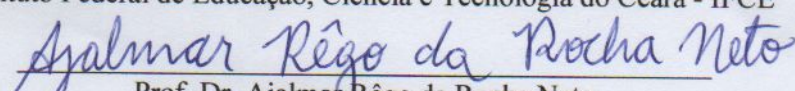
Dissertação submetida à Coordenação do Curso de Pós-graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação, área de concentração Ciência da Computação.

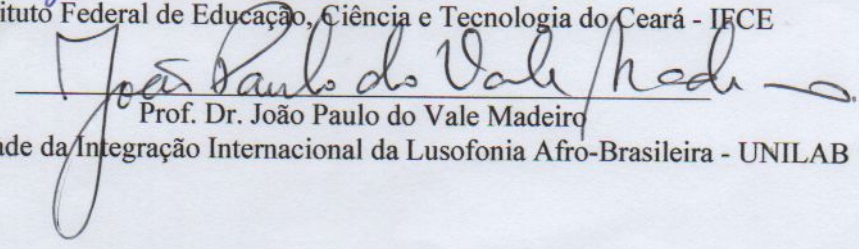
Aprovada em 01 / 02 /2017.

BANCA EXAMINADORA


Prof. Dr. Pedro Pedrosa Rebouças Filho (Orientador)
Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE


Prof. Dr. Antônio Wendell de Oliveira Rodrigues
Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE


Prof. Dr. Ajalmar Rêgo da Rocha Neto
Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE


Prof. Dr. João Paulo do Vale Madeiro
Universidade da Integração Internacional da Lusofonia Afro-Brasileira - UNILAB

Este trabalho é dedicado à minha namorada, meus pais e irmão pelo incentivo e apoio em todas as minhas escolhas.

Agradecimentos

A Deus, acima de tudo.

Ao meu orientador, Prof. Dr. Pedro Pedrosa, pelos ensinamentos, amizade, confiança, paciência e dedicação depositados ao longo deste trabalho. Seu conhecimento iluminador e entusiasmo contagiante foram fundamentais para o sucesso desta jornada.

À minha namorada, Renata Lira, pelo carinho, compreensão e incentivo para que eu pudesse concluir esta importante etapa da minha vida.

À minha família que sempre me incentivou a alcançar caminhos cada vez mais distantes.

A todos os meus novos amigos e companheiros de pesquisa do PPGCC, Laboratório de Processamento de Imagens e Simulação Computacional (LAPISCO) e PPGET. Em especial, ao Saulo Oliveira, Madson Dantas, Fausto Sampaio, Kauê Menezes, Lucas Silva, Róger Sarnento e Jardel das Chagas pelas conversas e discussões que foram de suma importância na conclusão desse trabalho.

Ao Wellington Souza e Jefferson Almeida pela ajuda na aquisição das imagens.

Aos professores Dr. Amauri Holanda e Dr. Victor Hugo por ter ajudado nas correções das publicações.

A todos os professores do mestrado que de alguma forma contribuíram para minha formação.

À Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUN-CAP) pelo suporte financeiro.

“É preciso força pra sonhar e perceber que a estrada vai além do que se vê.”
(Los Hermanos)

Resumo

Esta dissertação apresenta um estudo comparativo do desempenho de diversos descritores de imagem e classificadores na tarefa de reconhecimento de locais internos para localização e navegação de um robô móvel usando mapa topológico, avaliando dois tipos de câmera, convencional e omnidirecional. Na metodologia adotada são avaliadas combinações de extratores de atributos e classificadores promissores existentes na literatura para reconhecimento de cenas para definir a melhor configuração para cada tipo de câmera. Esse trabalho também propõe uma nova abordagem de classificação com opção de rejeição para localização de robôs móveis utilizando informação do mapa topológico para otimizar a navegação do robô, chamada de Suporte de Decisão para Localização (SDL). A melhor acurácia obtida com a câmera convencional foi 99,96% através da combinação da Máquina de Aprendizagem Mínima (*Minimal Learning Machine with Nearest Neighbors*, MLM-NN) com os Momentos Espaciais (ME), cujo tempo de classificação foi 128,8 μ s e de extração 0,165s, sendo este o melhor tempo de extração obtido. Quanto ao tempo de classificação, o melhor resultado obtido foi 47,3 μ s com classificador Bayesiano, cuja acurácia foi 99,94%. No segundo grupo dos experimentos com as imagens da câmera omnidirecional, a combinação com maior acurácia foi 96,67% através da MLM-NN com *Local Binary Patterns* (LBP), cujo tempo de classificação foi 1210 μ s e de extração 0,337s. Por outro lado, o menor tempo de extração foi de 0,178s com os ME e de classificação 36,4 μ s com classificador Bayesiano, cuja acurácia foi 87,5%. Quanto aos ganhos obtidos pelo SDL, verifica-se que a acurácia média do teste de navegação com a primeira câmera foi 83% sem o SDL e 91,9% com o SDL, e passou de 36,3% para 52,1% com SDL na câmera omnidirecional. A taxa de acerto da MLM-NN com ME e do classificador Bayesiano com ME dos testes de navegação utilizando a câmera de visão convencional sem o SDL foi 100%, e o SDL apenas confirmou o trajeto obtido. Entretanto, utilizando a câmera omnidirecional, os valores obtidos na navegação aumentaram 11,6% na taxa de acerto para MLM-NN com LBP, passando de 86% para 96%. De forma similar, a acurácia do classificador Bayesiano com ME passou de 82% para 93%. Conclui-se, portanto, que a combinação do classificador Bayesiano com os ME utilizando a câmera de visão convencional é a melhor escolha para a tarefa de navegação considerando a acurácia, custo computacional e de armazenamento, dentre as configurações avaliadas. Por fim, conclui-se que o SDL aumenta a eficácia e confiabilidade na navegação em robótica móvel, podendo ser utilizado para este fim em aplicações reais.

Palavras-chave: Reconhecimento de Locais Internos. Localização de Robôs Móveis. Mapa Topológico. Extrator de Características. Reconhecimento de Padrões.

Abstract

This dissertation presents a comprehensive performance comparison study of several image descriptors and classifiers applied to indoor place recognition to mobile robot localization and navigation using a topological map. We also assess the effectiveness of two types of cameras, conventional and omnidirectional, in the task of interest. In the adopted methodology, we assess combinations of promising feature extractors and classifiers existing in the literature to indoor place recognition in order to choose the best setting for each type of camera. In addition, we propose a novel approach of classification with reject option for localization of mobile robots using topological map information in order to optimize the robot navigation, called Decision Support for Localization (DSL). The best accuracy obtained with the first camera was 99.96% through the combination of the Minimal Learning Machine with Nearest Neighbors (MLM-NN) along with the Spatial Moments (SM), whose classification time was 128.8 μ s and extraction time was 0.165s, which is the best one. With regard to classification time, the best result was 47.3 μ s with Bayes classifier, which accuracy was 99.94%. In the second group of experiments with the omnidirectional camera images, the combination with the best accuracy obtained was 96.67% through MLM-NN along with *Local Binary Patterns* (LBP), which classification time was 1210 μ s and extraction time was 0.337s. On the other hand, the shortest extraction time was 0.178s with SM and of classification 36.4 μ s along with Bayes classifier, whose accuracy was 87.5%. With regard to gains obtained by DSL, the accuracy of the navigation test using the first camera was 83% with DSL and 91.9% with DSL, while with the omnidirectional camera it went from 36.3% to 52.1% with DSL. The accuracy of the MLM-NN with SM and of the Bayes classifier with SM in the navigation tests using the first camera without the DSL was 100%, and the DSL only confirmed the obtained route. However, the accuracy of the MLM-NN with LBP in the navigation test using the omnidirectional camera increased 11.6%, went from 86% to 96%. Likewise, the accuracy of the Bayes classifier along with SM went from 82% to 93%. We can conclude that the combination of Bayes classifier along with Spatial Moments using the camera with conventional field of view is a suitable choice for the navigation task considering the accuracy, computational and storage cost, among the evaluated configurations. Finally, we can see that the DSL increases the efficiency and reliability of navigation in mobile robotic, and can be used for this purpose in real applications.

Keywords: Indoor Place Recognition. Mobile Robot Localization. Topological Map. Feature Extraction. Pattern Recognition.

Lista de ilustrações

Figura 1 – Categorização das tecnologias de posicionamento em ambientes internos (ALARIFI et al., 2016).	29
Figura 2 – Imagem de uma cena incoerente ilustrando violações das cinco relações introduzidas por Biederman (1981): suporte : o notebook está flutuando; intersversão : o fundo aparece através da maleta; probabilidade : o hidrante está na sala; posição : a torneira está sobre o sofá; tamanho : a cadeira parece ser maior do que a mesa.	36
Figura 3 – Paradigmas de mapeamento e localização de robôs móveis.	41
Figura 4 – Estrutura e fluxo de dados de um sistema típico de reconhecimento de padrões.	42
Figura 5 – Para uma vizinhança 3×3 na imagem, o padrão binário local é determinado pela comparação do <i>pixel</i> central com seus vizinhos na ordem fixada atribuída pela seta. O padrão binário gera um número que é computado no histograma LBP.	53
Figura 6 – Diferença entre os padrões não uniforme e uniforme. Da esquerda para a direita; a primeira e segunda matriz mostram um exemplo de um padrão não uniforme. A terceira e quarta matriz exibem características do LBP uniforme. O padrão uniforme possui, apenas, duas regiões conectadas.	55
Figura 7 – Exemplo do LBP com padrões uniformes aplicado ao problema de localização de robôs móveis: (a) Imagem de entrada. (b) Resultado da aplicação do LBP com contraste melhorado para propósitos de visualização e (c) histograma da imagem após utilizar o LBP.	55
Figura 8 – Direções de adjacência da GLCM.	56
Figura 9 – GLCM com distância $D = 1$ e direção $\theta = 0$ de uma imagem I (GONZALEZ; WOODS, 2010).	56
Figura 10 – Criação de uma imagem resultante da diferença de Gaussianas (DoG) (LOWE, 2004).	61
Figura 11 – Descritor das características do SIFT (LOWE, 2004).	61
Figura 12 – Com imagens integrais, apenas quatro operações são necessárias para calcular a soma de intensidades em uma área retangular na imagem (BAY et al., 2008).	63
Figura 13 – Da esquerda para direita: a derivadas Gaussianas de segunda ordem discretizadas L_{yy} e L_{xy} e aproximações do SURF para L_{yy} e L_{xy} (EVANS, 2009).	63
Figura 14 – Problema de classificação binária.	68

Figura 15 – Exemplo de um problema que não pode ser resolvido de forma satisfatória por um classificador linear (SOUZA JÚNIOR et al., 2015).	71
Figura 16 – Mapeamento de dados para um espaço de características de mais alta dimensão através do truque do <i>kernel</i> (ROCHA NETO, 2011).	71
Figura 17 – Modelo esquemático de uma rede MLP com uma camada oculta.	73
Figura 18 – Gráficos das funções logística (a) e tangente hiperbólica (b).	75
Figura 19 – Etapas do fluxo adotado da abordagem proposta neste trabalho para navegação de robôs móveis.	84
Figura 20 – Robô utilizado na navegação com as respectivas câmeras adotadas.	84
Figura 21 – Mapa topológico do ambiente de navegação através de duas perspectivas.	85
Figura 22 – Imagens capturadas no local referente à classe 7 (nó 2) nos ambientes (a e c) sintético e (b e d) real com a câmera convencional e omnidirecional, respectivamente.	87
Figura 23 – Ilustração da rota 7 utilizada no teste de navegação. Nessa rota, o robô começa na classe 8 nó 1 e segue em frente duas vezes. Ele vira à esquerda e novamente segue em frente duas vezes. Finalmente, ele alcança a classe 12 no nó 6.	90
Figura 24 – Situação hipotética na qual o classificador prediz que há uma maior probabilidade da posição atual ser um lugar muito distante do real. O suporte de decisão descarta a primeira sugestão do classificador pelo critério de vizinhança no mapa e escolhe a próxima.	91
Figura 25 – Exemplo de validação cruzada com <i>k-fold</i> . Os conjuntos de treinamento são representados em cor azul e o de validação em vermelho.	94
Figura 26 – Taxa de acerto média do teste de navegação, com e sem o Suporte de Decisão para Localização, obtidos pela combinação entre os classificadores e extratores utilizando a GoPro	103
Figura 27 – Taxa de acerto média do teste de navegação, com e sem o Suporte de Decisão para Localização, obtidos pela combinação entre os classificadores e extratores utilizando a câmera omnidirecional	112

Lista de tabelas

Tabela 1	– Comparação entre tecnologias de localização interna. Intensidade do sinal recebido (<i>received signal strength</i> , RSS). Tempo de retorno (<i>time of arrival</i> , ToA). Diferença de tempo de chegada (<i>time difference of arrival</i> , TDOA). Ângulo de chegada (<i>angle of arrival</i> , AoA).	30
Tabela 2	– Soluções e abordagens de localização topológica baseado em visão computacional.	31
Tabela 3	– <i>Kernels</i> comumente utilizados.	72
Tabela 4	– Quantidade de imagens por classe e resolução dos quatro conjuntos de imagens obtidas.	86
Tabela 5	– Número de atributos gerados por cada técnica de extração de características sobre os quatro conjuntos de imagens.	88
Tabela 6	– Comandos para mover o robô e os pontos iniciais e finais usados no teste de navegação: seguir em frente (SF), virar à esquerda (VE), virar à direita (VD) e girar 180 graus (G180).	89
Tabela 7	– Matriz de confusão para um problema de classificação multiclasse. . . .	92
Tabela 8	– Especificidade (Es), sensibilidade (Se), valor preditivo positivo (VPP), F1-score, média harmônica (MH) and Acurácia (Ac) obtidos pelos extratores e classificadores no ambiente sintético usando as imagens com campo de visão convencional	98
Tabela 9	– Especificidade (Es), sensibilidade (Se), valor preditivo positivo (VPP), F1-score, média harmônica (MH) and Acurácia (Ac) obtidos pelos extratores e classificadores no ambiente real usando as imagens provenientes da câmera GoPro®	100
Tabela 10	– Acurácia (Ac), tempo de treino, tempo de teste e de extração com todos extratores combinados com as melhores configurações dos classificadores utilizando as imagens da GoPro®.	101
Tabela 11	– Taxa de acerto por rota das melhores configurações dos classificadores combinados com os extratores sem o Suporte de Decisão para Localização utilizando a GoPro®.	104
Tabela 12	– Taxa de acerto por rota das melhores configurações dos classificadores combinados com os extratores com o Suporte de Decisão para Localização utilizando a GoPro®.	105
Tabela 13	– Espaço de armazenamento dos modelos embarcados para cada classificador com a GoPro®.	106

Tabela 14 – Especificidade (Es), sensibilidade (Se), valor preditivo positivo (VPP), F1-score, média harmônica (MH) and Acurácia (Ac) obtidos pelos extratores e classificadores no ambiente sintético usando as imagens da câmera omnidirecional	108
Tabela 15 – Especificidade (Es), sensibilidade (Se), valor preditivo positivo (VPP), F1-score, média harmônica (MH) and Acurácia (Ac) obtidos pelos extratores e classificadores no ambiente real usando as imagens da câmera omnidirecional	109
Tabela 16 – Acurácia (Ac), tempo de treino, tempo de teste e de extração com todos extratores combinados com as melhores configurações dos classificadores utilizando as imagens da câmera omnidirecional.	111
Tabela 17 – Taxa de acerto por rota das melhores configurações dos classificadores combinados com os extratores sem o Suporte de Decisão para Localização utilizando a câmera omnidirecional	113
Tabela 18 – Taxa de acerto por rota das melhores configurações dos classificadores combinados com os extratores com o Suporte de Decisão para Localização utilizando a câmera omnidirecional	114
Tabela 19 – Tamanho em kilobyte dos modelos embarcados para cada classificador com a câmera omnidirecional.	115
Tabela 20 – Valores dos hiperparâmetros dos classificadores utilizado no teste de navegação com o robô usando a câmera GoPro . <i>CO</i> significa o número de neurônios na camada oculta e <i>K</i> é a porcentagem dos dados de treinamento usando como pontos de referência.	133
Tabela 21 – Valores dos hiperparâmetros dos classificadores utilizado no teste de navegação com o robô usando a câmera omnidirecional . <i>CO</i> significa o número de neurônios na camada oculta e <i>K</i> é a porcentagem dos dados de treinamento usando como pontos de referência.	133

Lista de Algoritmos

1	Algoritmo de treinamento da MLM e MLM-NN	80
2	Algoritmo de Teste da MLM-NN	80

Lista de abreviaturas e siglas

Ac	Acurácia
DoG	<i>Difference of Gaussians</i> (Diferença de Gaussianas)
Es	Especificidade
GLCM	<i>Gray-Level Co-occurrence Matrix</i> (Matriz de Co-ocorrência)
GPS	<i>Global Positioning System</i> (Sistema de Posicionamento Global)
LBP	<i>Local Binary Pattern</i> (Padrão Binário Local)
LMS	<i>Least Mean Square</i> (Erro Médio Quadrático)
LM	Levenberg-Marquardt (LM)
LRF	<i>Laser Range Finder</i> (Telêmetro à laser)
MAP	Máximo <i>a posteriori</i>
MH	Média Harmônica
MLM-NN	MLM- <i>Nearest Neighbors</i> (Máquina de Aprendizagem Mínima com Vizinhos Mais Próximos)
MLM	<i>Minimal Learning Machine</i> (Máquina de Aprendizagem Mínima)
MC	Momentos Centrais
ME	Momentos Espaciais
MI	Momentos Invariantes
MLP	<i>Multi Layer Perceptron</i> (Perceptron de Múltiplas Camadas)
PPA	<i>parahippocampal place area</i> (Área Para-hipocampal de Lugares)
RBF	<i>Radial Basis Function</i> (Função de Base Radial)
RFID	<i>Radio-Frequency Identification</i> (Identificação por Rádio Frequência)
RNA	Redes Neurais Artificiais
SDL	Suporte de Decisão para Localização

SIFT	<i>Scale-Invariant Feature Transformation</i> (Transformada de Características Invariante à Escala)
SURF	<i>Speeded Up Robust Features</i> (Características Robustas Aceleradas)
SVM	<i>Support Vector Machines</i> (Máquinas de Vetores de Suporte)
Se	Sensibilidade
VPP	Valor Preditivo Positivo

Lista de símbolos

$\mathbb{R}$	Conjunto dos números reais
$f(x, y)$	Função bidimensional de intensidade de luz (imagem)
$\mathbf{I}(x, y)$	Matriz bidimensional de intensidade de luz (imagem)
$\max$	Máximo
$\min$	Mínimo
M_{pq}	Momento geral de ordem p, q de uma imagem
m_{pq}	Momentos Espaciais
μ_{pq}	Momentos Centrais
η_{pq}	Momentos Centrais Normalizados
ϕ_i	Momentos de Hu
$\mathbf{G}_D^\theta$	Matriz de co-ocorrência com distância D e direção θ
L	Número de níveis de cinza
$\mathbf{x}$	Vetor de características
C	Classe de uma amostra
$P(X, Y)$	Probabilidade conjunta
$P(Y X)$	Probabilidade <i>a posteriori</i>
$sign(\cdot)$	Função sinal
$K(\cdot, \cdot)$	Função de <i>kernel</i>
$\mathbf{H}$	Matriz Hessiana
$\mathbf{J}$	Matriz Jacobiana

Sumário

1	INTRODUÇÃO	27
1.1	Trabalhos Relacionados	28
1.2	Objetivos	32
1.2.1	Objetivo Geral	32
1.2.2	Objetivos Específicos	32
1.3	Produção científica	32
1.4	Organização da Dissertação	33
2	RECONHECIMENTO DE LOCAIS INTERNOS POR VISÃO COMPUTACIONAL	35
2.1	Locais e Cenas	35
2.2	Descrição do Problema	37
2.3	Percepção Humana da Cena	37
2.3.1	Percepção e Reconhecimento da Cena	38
2.3.2	Representação da Cena	39
2.4	Localização e Mapeamento	40
2.5	Estrutura de um Sistema Típico de Reconhecimento de Padrões	41
2.5.1	Sensoriamento	42
2.5.2	Segmentação e Pré-Processamento	43
2.5.3	Extração de Características	43
2.5.4	Classificação	44
2.5.5	Pós-processamento	44
2.5.6	Treinamento	45
2.5.7	Otimização	46
2.6	Conclusão	46
3	EXTRATORES DE CARACTERÍSTICAS	47
3.1	Descrição de Características	47
3.2	Invariância	47
3.3	Descritores baseados em Momentos	48
3.3.1	Definição Geral dos Momentos	49
3.3.2	Momentos Espaciais	49
3.3.3	Momentos Centrais	50
3.3.4	Momentos Centrais Normalizados	51
3.3.5	Momentos Invariantes de Hu	51
3.4	Descritores baseados em texturas	52

3.4.1	Padrões Binários Locais	52
3.4.1.1	Padrões Uniformes	54
3.4.2	Matriz de Coocorrência e os Descritores de Haralick	54
3.5	Descritores de pontos-chave	59
3.5.1	SIFT	60
3.5.2	SURF	62
3.6	Conclusão	63
4	RECONHECIMENTO DE PADRÕES	65
4.1	Classificador Bayesiano	66
4.2	Máquinas de Vetores Suporte	67
4.2.1	Problema Primal	68
4.2.2	Formulação Dual	70
4.2.3	O Truque do <i>Kernel</i>	70
4.3	Máquinas de Vetores de Suporte por Mínimos Quadrados	72
4.4	Perceptron de Múltiplas Camadas	73
4.4.1	Algoritmos de Treinamento	75
4.4.1.1	Algoritmo de Retropropagação do Erro	75
4.4.1.2	Levenberg-Marquardt	75
4.5	Máquina de Aprendizagem Mínima	76
4.5.1	Formulação	77
4.5.1.1	Regressão entre Distâncias	77
4.5.1.2	Estimativa da saída	78
4.5.2	Máquina de Aprendizagem Mínima para Classificação	79
4.6	Máquina de Aprendizagem Mínima com Vizinhos Mais Próximos . .	80
4.7	Conclusão	80
5	MATERIAIS E MÉTODOS	83
5.1	Robô e Mapa Topológico	83
5.2	Aquisição de Imagens	85
5.3	Etapa de Extração	86
5.4	Etapa de Classificação	88
5.5	Rotas de Navegação Adotadas	89
5.6	Suporte de Decisão para Localização (Proposta)	89
5.7	Avaliação de Desempenho	92
5.7.1	Métricas de Avaliação	92
5.7.2	Validação Cruzada	94
5.8	Conclusão	95
6	RESULTADOS E DISCUSSÕES	97

6.1	Resultados com a câmera GoPro®	97
6.1.1	Análise do tempo de processamento (GoPro®)	99
6.1.2	Teste de Navegação (GoPro®)	102
6.1.3	Custo de armazenamento (GoPro®)	106
6.2	Resultados com a câmera omnidirecional	107
6.2.1	Análise do tempo de processamento (Omnidirecional)	110
6.2.2	Teste de Navegação (Omnidirecional)	110
6.2.3	Custo de armazenamento (Omnidirecional)	115
6.3	Análise geral dos resultados	115
7	CONCLUSÃO	117
7.1	Proposta de Trabalhos Futuros	119
	 Referências	 121
	 APÊNDICES	 131
	APÊNDICE A – HIPERPARÂMETROS ADOTADOS NO TESTE DE NAVEGAÇÃO	 133

1 Introdução

A capacidade de adquirir, representar e combinar a informação perceptiva das memórias de locais armazenados em um mapa cognitivo interno é uma característica fundamental do ser humano e de inúmeros animais. Embora essa tarefa seja realizada de diversas maneiras, ela continua a ser um processo complexo que envolve inúmeros mecanismos perfeitamente cooperantes. Diante desses desafios, essa dissertação propõe uma abordagem de navegação e localização de robôs móveis em locais internos usando visão computacional em dois tipos de câmera. Além disso, um método de classificação com opção de rejeição para tarefa de interesse é apresentado. Enfatiza-se a robustez e eficiência à taxa de acerto, tempo de processamento e espaço de armazenamento, uma vez que essas são as propriedades fundamentais dos sistemas de reconhecimento com o objetivo de ser usado em aplicações do mundo real.

Em ambientes externos, sistemas de navegação como o Sistema de Posicionamento Global (*Global Positioning System*, GPS) fornecem posições precisas e podem ser aplicados no reconhecimento de locais. No entanto, tais sistemas não funcionam bem em ambientes internos. Diante disso, diferentes métodos de posicionamento foram propostos. Eles costumam ser baseados em ultrassom, Bluetooth, Wi-Fi, reconhecimento de imagem e navegação inercial (SONG; JIANG; HUANG, 2011; MA et al., 2014; XU et al., 2015). Yayan, Yucel e Yazıcı (2015) apresentam um sistema de posicionamento de locais internos que utiliza sinais de ultrassom para calcular a posição de um robô móvel. Esses sensores, no entanto, são prejudicados por pequenas fontes de ruído e pelo formato dos objetos. Ko e Kuc (2015) propõem um método para a localização de robôs móveis em ambientes internos parcialmente desconhecidos, que combina sensores ultrassônicos com um telêmetro¹ à laser (*Laser Range Finder*, LRF). Entretanto, o LRF possui um custo financeiro elevado. Em Song et al. (2016), um sistema de posicionamento com Identificação por Rádio Frequência (*Radio-Frequency Identification*, RFID) é apresentado. Esses sistemas, no entanto, exigem uma infraestrutura complexa.

Os estudos atuais de navegação de robôs móveis baseados em visão computacional estão cada vez mais crescentes (WANG et al., 2012). Isso ocorre devido ao fato da percepção humana ser primeiramente visual. Além disso, compreender e utilizar esse tipo de informação é essencial para permitir a interação homem-robô. As propriedades visuais de um local podem ser usadas para determinar a sua categoria funcional (por exemplo, cozinha, escritório). Ademais, pontos de referência visuais são comumente usados por seres humanos para planejar caminhos e descrever lugares. Essas características motivaram pesquisadores a empregar a informação visual na tarefa de localização de robôs móveis.

¹ Telêmetro é um dispositivo de precisão destinado à medição de distâncias em tempo real.

Além da localização topológica de robôs móveis, há outras aplicações, tais como recuperação de imagens baseada em conteúdo, que usam o reconhecimento de locais com visão computacional. Atualmente, tecnologias da Internet como a *World Wide Web* fornecem acesso a enormes volumes de dados, dos quais grande parte são imagens. Consequentemente, a capacidade de obter imagens com base em uma descrição do seu conteúdo torna-se indispensável.

Os métodos de aprendizado de máquina têm sido amplamente aplicados juntamente com técnicas de visão computacional devido a sua capacidade eficaz de aprender padrões complexos e tomar decisões com base em dados (YANG et al., 2011). Charalampous, Kostavelis e Gasteratos (2015) apresentam um método de planejamento para caminhos locais com SVM. Caron, Filliat e Gepperth (2014) utilizam uma rede neural MLP juntamente com uma câmera RGB-D para o reconhecimento de objetos. Em Kasat e Thepade (2016), uma abordagem de classificação de imagem foi proposta usando o classificador Bayesiano.

Com objetivo de enfatizar a robustez e eficiência à taxa de acerto, tempo de processamento e espaço de armazenamento, o sistema de reconhecimento de posicionamento de robôs móveis com visão computacional foi avaliado com diversos extratores e classificadores em conjunto, verificando a eficiência de dois tipos de câmera, convencional e omnidirecional, nos ambientes simulado e real. Por fim, uma abordagem de classificação com opção de rejeição para navegação de robôs móveis é proposta.

1.1 Trabalhos Relacionados

A localização é uma habilidade básica necessária para navegação. Com objetivo de realizar tal tarefa, o robô precisa estimar sua posição no ambiente (CASTELLANOS; TARDOS, 2012). Assim, o sistema de localização deve adquirir a maior quantidade possível de informação do ambiente. Existem inúmeros tipos de sensores utilizados na robótica móvel para localização, tais como as tecnologias baseadas em ultrassom, em laser ou baseadas em visão computacional (DEAK; CURRAN; CONDELL, 2012). Alarifi et al. (2016) classifica as tecnologias de posicionamento em ambientes internos de acordo com a infraestrutura do sistema que as utiliza, ver Figura 1. A Tabela 1 compara as vantagens e desvantagens dessas tecnologias e apresenta alguns estudos relacionados. Dentre as vantagens do uso de um método de localização baseado em visão computacional, pode-se destacar a facilidade de escalabilidade sem aumento significativo do custo. Além disso, esses métodos não demandam mudanças no ambiente (MAUTZ; TILCH, 2011).

Garcia-Fidalgo e Ortiz (2015) apresenta um estudo sobre várias abordagens de localização e mapeamento topológico utilizando visão computacional. Os autores classificam essas abordagens em descritores globais, descritores locais e bolsa de palavras (*bag-of-words*, BoW). A Tabela 2 mostra as vantagens e desvantagens de cada abordagem e

algumas aplicações. De acordo com a Tabela 2, observam-se diversas contribuições da visão computacional na localização e navegação de robôs. Esta dissertação utiliza essa abordagem, no entanto, com duas contribuições, uma do ponto de vista científico e outra do ponto de vista tecnológico.

Os métodos de apoio à decisão evoluíram com intuito de tomar decisões semelhantes às dos seres humanos. No entanto, esses métodos ainda apresentam erros. Assim, paradigmas, que fornecem uma alternativa mais viável, foram desenvolvidos quando a confiança na tomada de decisão é baixa (SOUSA et al., 2015). Diante desses desafios, esta dissertação propõe um método de localização por meio de classificação com uma opção de rejeição para navegação de robôs móveis usando imagens denominado Suporte de Decisão para Localização (SDL). O SDL utiliza a informação de distância do mapa topológico para fornecer, automaticamente, uma opção adequada durante a localização quando a confiança na decisão é baixa. Os resultados mostram que o método proposto aumentou a precisão de todos os classificadores avaliados neste trabalho, ou seja, a classificação é otimizada.

A contribuição deste trabalho do ponto de vista tecnológico é a utilização de técnicas de extração de características globais e locais baseadas em momentos (Momentos Espaciais, Momentos Centrais e Momentos de Hu), textura (LBP e GLCM) e descritores de ponto-chave (SIFT e SURF), juntamente com técnicas consolidadas de aprendizagem de máquina para tarefas de localização de robôs móveis com imagens de campo de visão convencional e omnidirecional. Como pode ser visto na Tabela 2, os demais trabalhos realizam uma comparação simples com alguns métodos. Em contrapartida, diversas abordagens para otimizar a solução são avaliadas nesta dissertação.

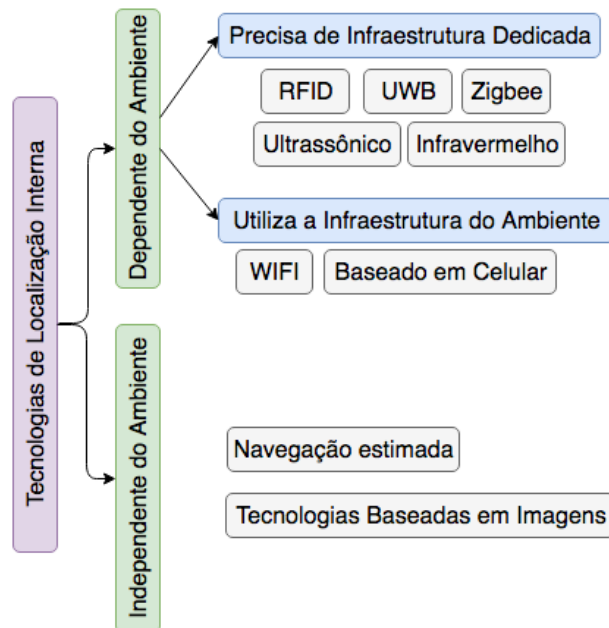


Figura 1 – Categorização das tecnologias de posicionamento em ambientes internos (ALARIFI et al., 2016).

Tabela 1 – Comparação entre tecnologias de localização interna. Intensidade do sinal recebido (*received signal strength*, RSS). Tempo de retorno (*time of arrival*, ToA). Diferença de tempo de chegada (*time difference of arrival*, TDOA). Ângulo de chegada (*angle of arrival*, AoA).

Tecnologia	Método Comum	Vantagem	Desvantagem	Sistema
Baseado em Imagem	Reconhecimento de Padrões	Não necessita de mudanças no ambiente (implantação de tag RFID ou roteadores, p. ex.) (DEAK; CURRAN; CONDELL, 2012); facilmente escalável sem significativo aumento do custo (MAUTZ; TILCH, 2011).	Depende do campo de visão; alguns métodos de processamento de imagens requerem altos custos computacionais (HARTMANN, 2011).	Rubio et al. (2016), Son, Kim e Sohn (2015), Debski et al. (2015), Li (2014).
Bluetooth	Proximidade, RSS	Presente em dispositivos comuns como smartphones (DEAK; CURRAN; CONDELL, 2012); Não precisa de campo de visão (SONG; JIANG; HUANG, 2011).	O usuário precisa fazer uma varredura (DEAK; CURRAN; CONDELL, 2012); o limite de cobertura é limitado (MAINETTI; PATRONO; SERGI, 2014).	Kwicień et al. (2015), Li (2014).
Zigbee	RSS, Medidas de Mudança de Fase	Requer pouca energia; baixo custo (SONG; JIANG; HUANG, 2011).	Opera em banda não licenciada tornando-se vulnerável a interferência causada por outros sinais (ALARIFI et al., 2016).	Fang et al. (2012)
WLAN	RSS	O hardware é econômico e fácil de instalar (CHOWDHURY et al., 2016).	Afetado por múltiplos caminhos e efeito de fadiga do sinal (CHOWDHURY et al., 2016).	Nowicki e Skrzypczyński (2015), Campos, Lovisolo e Campos (2014), Lim et al. (2010)
UWB	ToA, TDOA	Seu sinal possui boa penetração de obstáculos; possui resolvabilidade direta de múltiplos caminhos (ALARIFI et al., 2016).	Elevado custo financeiro (SONG; JIANG; HUANG, 2011).	Zandian e Witkowski (2016), Monica e Ferrari (2015), Reyna (2013).
RFID	Proximidade, RSS	Adequado para ambientes densos (DEAK; CURRAN; CONDELL, 2012).	Alto custo dos leitores, infraestrutura complexa (DEAK; CURRAN; CONDELL, 2012).	Calderoni et al. (2015), Miah e Gueaieb (2011).
Ultrassom	ToA, TDOA	Baixo custo; não sofre de interferências eletromagnéticas (DEAK; CURRAN; CONDELL, 2012).	Afetado por sons de alta frequência; escalabilidade é complexa (DEAK; CURRAN; CONDELL, 2012).	Yayan, Yücel e Yazıcı (2015), Kim e Choi (2008).
Infravermelho	Proximidade, Mudança de Fase Diferencial, AoA	não sofre de interferências eletromagnéticas (CHOWDHURY et al., 2016).	Custo de hardware e manutenção alta (CHOWDHURY et al., 2016).	Sioutis e Tan (2014), Schindler, Metzger e Starnier (2006).
Baseado em Celular	RSS	Pode trabalhar na mesma frequência de outros dispositivos (DEAK; CURRAN; CONDELL, 2012).	Baixa precisão (DEAK; CURRAN; CONDELL, 2012).	Varshavsky et al. (2007).
Navegação Estimada	Rastreamento	Não é necessário hardware extra (ALARIFI et al., 2016).	Computa uma posição aproximada (ALARIFI et al., 2016).	Kothari et al. (2012), Jimenez et al. (2009).

Tabela 2 – Soluções e abordagens de localização topológica baseado em visão computacional.

	Vantagem	Desvantagem	Descritor	Referência
Descritor Global	Baixo custo computacional e de armazenamento (BONIN-FONT; ORTIZ; OLIVER, 2008).	Pouco robusto aos efeitos de oclusão e iluminação, reduzindo o poder de discriminação (BONIN-FONT; ORTIZ; OLIVER, 2008).	PCA Omni-gist DP-FACT	Bessa et al. (2015) Rituerto, Murillo e Guerrero (2014) Liu e Siegart (2012)
Descritor Local	Alto poder de discriminação; são geralmente mais robustos às mudanças de escala, oclusão, iluminação e rotação (GARCIA-FIDALGO; ORTIZ, 2015)	O custo computacional e as necessidades de armazenamento são maiores do que para descritores globais (FUENTES; RUIZ; RENDÓN, 2015).	SIFT Wavelets 3D-PIRF (SURF)	Johns e Yang (2011) Saedan, Lim e Ang (2007) Morioka, Yi e Hasegawa (2011)
BoW	Adequado com elevado número de imagens para processar (FUENTES; RUIZ; RENDÓN, 2015).	Há presença de palavras com ‘ruído’ devido o modo de construção do vocabulário e a perda das relações espaciais entre as palavras; precisa de uma moderada quantidade de armazenamento (GARCIA-FIDALGO; ORTIZ, 2015).	FAST/BRIEF SURF SIFT	Galvez-López e Tardos (2012) Cummins e Newman (2011) Ranganathan e Dellaert (2011)

1.2 Objetivos

O objetivo geral desta dissertação, bem como seus objetivos específicos, são detalhados nas subseções a seguir.

1.2.1 Objetivo Geral

O objetivo geral desta dissertação é criar um sistema de navegação para robôs móveis utilizando uma abordagem proposta de classificação com opção de rejeição para localização denominada Suporte de Decisão para Localização (SDL).

1.2.2 Objetivos Específicos

Em seu sentido mais específico, essa dissertação tem os seguintes objetivos:

- Criar um ambiente simulado computacionalmente para cada tipo de câmera, convencional e omnidirecional;
- Criar um banco de imagem do ambiente virtual para cada tipo de câmera, convencional e omnidirecional;
- Criar um banco de imagem do ambiente real para cada tipo de câmera, convencional e omnidirecional;
- Analisar combinações de extratores de características e classificadores consolidados com imagens provenientes de dois tipos de câmera, convencional e omnidirecional;
- Identificar a viabilidade do uso das técnicas na aplicação em tempo real;
- Propor uma abordagem para auxiliar a tomada de decisão durante a localização do robô móvel;
- Desenvolver um sistema autônomo de navegação e localização com mapas topológicos para robôs móveis.

1.3 Produção científica

Ao longo do desenvolvimento desta dissertação os seguintes artigos científicos foram publicados:

- **MARINHO, L.B.; ALMEIDA, J. S.; MENDES S., J. WELLINGTON; de ALBUQUERQUE, V.H.C.; REBOUÇAS FILHO, P. P. A novel mobile robot navigation approach based on topological maps using classification with reject option in omnidirectional images.** *Expert Systems with Applications*, 2016.

- **MARINHO, L.B.**; REBOUÇAS FILHO, P. P.; DE SOUZA JUNIOR, Amauri Holanda. **A new approach to Human Activity Recognition using Machine Learning techniques**. In: Proceedings of the sixteenth International Conference on Intelligent Systems Design and Applications (ISDA'16), 2016.
- SILVA, E.M.; **MARINHO, L.B.**; REBOUÇAS FILHO, P. P.; LEITE, J.P.; LEITE, J.P.; FIALHO, W.M.L.; de ALBUQUERQUE, V.H.C.; TAVARES, J.M.R.S. **Classification of Induced Magnetic Field Signals for the Microstructural Characterization of Sigma Phase in Duplex Stainless Steels**. Metals, v. 6, p. 164, 2016.
- CUNHA, M. A. Q.; **MARINHO, L.B.**; REBOUÇAS FILHO, P. P.; **Mechanical Traction Machine Retrofitting via Digital Data Acquisition and Processing**. IEEE Latin America Transactions, 2016.
- RODRIGUES, M. B.; **MARINHO, L.B.**; NÓBREGA, M. R. V.; MENDES S., J. WELLINGTON; REBOUÇAS FILHO, P. P.; **Lung Segmentation in chest Computerized Tomography images using Border Following algorithm**. In: Proceedings of the sixteenth International Conference on Intelligent Systems Design and Applications (ISDA'16), 2016.

Além disto, outros trabalhos estão em processo de avaliação, os quais foram submetidos conforme segue:

- **MARINHO, L. B.**; ALMEIDA, J. S.; MENDES S., J. WELLINGTON; SOUZA JUNIOR, A. H.; REBOUÇAS FILHO, P. P. **A novel mobile robot navigation approach based on classification with rejects option using topological position via artificial vision**. Submetido para *Robotics and Computer-Integrated Manufacturing*.
- **MARINHO, L.B.**; REBOUÇAS FILHO, P. P.; DE SOUZA JUNIOR, Amauri Holanda. **A practical approach to feature selection applied to Human Activity Recognition**. Submetido para *Expert Systems*.
- REBOUÇAS FILHO, P. P.; REBOUÇAS, E. S.; **MARINHO, L. B.**; SARMENTO, R. M.; TAVARES, J.M.R.S.; de ALBUQUERQUE, V.H.C. **Analysis of Human Tissue Densities: A new approach to extract features from medical images**. Submetido para *Pattern Recognition Letters*.

1.4 Organização da Dissertação

O restante desta dissertação está organizada segundo os capítulos abaixo.

O **Capítulo 2** define o problema de reconhecimento de locais, discute vários aspectos de percepção humana da cena e, finalmente, apresenta uma arquitetura de um sistema de reconhecimento visual.

O **Capítulo 3** apresenta os métodos usados para extrair os atributos das imagens, a partir do qual serão classificadas pelos métodos apresentados no **Capítulo 4**.

O **Capítulo 5** explica as etapas empregadas na abordagem adotada, como foi realizado o teste de navegação, o método proposto para otimizar o processo de classificação e as métricas de avaliação.

O **Capítulo 6** apresenta os resultados e as discussões dos experimentos isolados, com as imagens sintéticas e reais utilizando os dois tipos de câmera, assim como do teste de navegação juntamente com a abordagem proposta para otimizar o processo de classificação.

No **Capítulo 7** estão as conclusões do estudo realizado nesta dissertação e as propostas para trabalhos futuros.

Os valores dos hiperparâmetros para os classificadores adotados no teste de campo são apresentados no **Apêndice A**.

2 Reconhecimento de Locais Internos por Visão Computacional

Nesse capítulo, um estudo sobre o problema de reconhecimento de locais é realizado. Na Seção 2.1, é discutida a definição dos objetos de interesse: local e cena. Além disso, serão analisados as propriedades de uma cena coerente e as limitações que devem ser satisfeitas para uma percepção visual ser chamada de cena. Tal conhecimento pode ser explorado na concepção de um sistema de reconhecimento de locais, bem como em experimentos psicofísicos¹ na percepção humana. Em seguida, na Seção 2.2, é formulado o problema do reconhecimento de locais internos por visão computacional e definido diversos termos-chaves. Na Seção 2.3, é discutida a percepção humana da cena e como os dados visuais são processados pelo cérebro humano. Essa seção constitui uma revisão das publicações sobre o tema. A Seção 2.4 apresenta os principais conceitos de localização na robótica móvel. Finalmente, a Seção 2.5, apresenta a estrutura de um sistema típico de reconhecimento de padrões.

2.1 Locais e Cenas

Essa seção inicia as considerações sobre reconhecimento de locais juntamente com as definições dos termos cena e local. De acordo com Henderson e Hollingworth (1999), cena é uma percepção visual semanticamente coerente (e muitas vezes nomeável) de um ambiente do mundo real compreendendo elementos de fundo (*background*) com vários objetos distintos dispostos de uma maneira espacialmente tolerada. O *background* consiste de estruturas e superfícies imóveis de grande escala, enquanto que objetos são entidades manipuláveis de menor escala. Diferenciar esses elementos, contudo, pode ser difícil. Por exemplo, a distinção entre o fundo e os objetos é arbitrária. Uma mesa pode ser considerado como um objeto quando vista de longe. No entanto, também pode constituir um fundo para objetos como caneta, teclado ou copo. Este problema é uma consequência das propriedades hierárquicas de uma cena. Portanto, os objetos devem ser organizados de uma maneira espacialmente tolerada para formar uma cena coerente.

Diante dessa problemática, é possível fazer as seguintes indagações: Será que a imagem apresentada na Figura 2 constitui uma cena coerente? É possível definir um conjunto de regras que uma cena deve seguir? Biederman (1981) define algumas restrições que devem ser satisfeitas para uma cena ser considerada coerente. Ele introduziu cinco

¹ A psicofísica é o ramo da psicologia que estuda a relação entre estímulos físicos e as respectivas sensações.

relações que podem ser utilizadas para descrever a diferença entre uma cena bem formada e um vetor de objetos estranhos. Elas podem ser facilmente explicadas usando uma ilustração gráfica de suas violações, ver Figura 2. As relações resultam das restrições físicas gerais (suporte e interversão), bem como da semântica, da função dos objetos e do ambiente circundante (probabilidade, posição e tamanho).



Figura 2 – Imagem de uma cena incoerente ilustrando violações das cinco relações introduzidas por Biederman (1981): **suporte**: o notebook está flutuando; **interversão**: o fundo aparece através da maleta; **probabilidade**: o hidrante está na sala; **posição**: a torneira está sobre o sofá; **tamanho**: a cadeira parece ser maior do que a mesa.

É possível formular o termo local de maneira semelhante. Os locais também possuem uma estrutura hierárquica. Um apartamento pode ser considerado como um local; no entanto, pode ser decomposto em compartimentos, tais como sala ou corredores que por sua vez podem compreender outras seções, por exemplo, quarto ou cozinha. Portanto, pode ser difícil determinar com precisão a fronteira entre os locais. Diferentes locais podem cumprir funções distintas, podem ter aparências diferentes ou podem apenas ser separados por paredes. Como resultado, um local pode ser considerado como um segmento geralmente nomeável de um ambiente do mundo real diferenciado devido a uma determinada funcionalidade, aparência ou fronteira artificial.

Embora esses termos não possam ser definidos com precisão, é possível tirar várias conclusões importantes. Em primeiro lugar, é esperado uma relação entre o tipo de um local e os objetos presentes nele. Observa-se, ainda, que os elementos que compõem o ambiente estão relacionados e devem satisfazer várias restrições. Os sistemas de reconhecimento devem ser capazes de lidar com a estrutura hierárquica de locais e cenas, bem como em

situações nas quais a fronteira entre os locais não é óbvia, ou apenas os locais diferem ligeiramente na aparência.

2.2 Descrição do Problema

O reconhecimento de locais internos (*indoor*) por visão computacional pode ser considerado um caso particular de reconhecimento de padrões restrito a representações visuais de um ambiente interno. Segundo Gonzalez e Woods (2010), padrão é uma descrição quantitativa ou estrutural de um objeto ou alguma outra entidade de interesse. De acordo com essa definição, a representação digital de uma imagem de um local, adquirida a partir de um ponto de vista particular, sob determinadas condições e em um determinado momento, também é um padrão, uma vez que constitui uma descrição de uma entidade nomeável, neste caso - um local.

Normalmente, os padrões podem ser agrupados em classes com características comuns. Essas propriedades, também conhecidas como atributos, permitem criar modelos a partir dos padrões em classes distintas. No caso do reconhecimento de locais, as classes correspondem aos locais, e os modelos são construídos através de características distintivas, por exemplo, os níveis de intensidade. Os principais conceitos de reconhecimento de padrões, juntamente com os métodos adotados nessa dissertação, são detalhados no Capítulo 4.

2.3 Percepção Humana da Cena

Os estudos atuais de visão computacional, muitas vezes, exploram o conhecimento de aspectos biológicos e psicológicos da percepção e cognição do ser humano. Na verdade, a pesquisa em áreas como a detecção e reconhecimento de padrões é realizada em paralelo aos experimentos psicofísicos em seres humanos. Isso pode ser explicado pelo fato dos melhores algoritmos disponíveis serem, ainda, superados pelos seres humanos em algumas tarefas. Além disso, a criação de máquinas capazes de interagir com pessoas e ambientes feitos pelo homem necessita da compreensão por trás dessa interação. Como resultado, várias soluções, já utilizadas em visão computacional e aprendizado automático, são motivadas por pesquisas psicológicas e biológicas.

Os seres humanos são também extremamente eficientes no reconhecimento de locais através da percepção visual. Por esta razão, é importante aprender como são processados dados visuais, o que se vê quando se olha para uma cena, como se representa o conhecimento espacial e, finalmente, como isso é feito de forma eficiente e robusta.

2.3.1 Percepção e Reconhecimento da Cena

Numerosas experiências relatam que os seres humanos são capazes de perceber e processar cenas rapidamente. A tarefa de reconhecer uma cena, determinando sua categoria semântica, extraindo as informações estruturais gerais, assim como o reconhecimento de alguns objetos básicos requer apenas uma fixação dos olhos e pode ser realizada em menos de 200ms. Esse tipo de informação é normalmente referenciada como a essência (*gist*) de uma cena. Potter (1976) mostra que pessoas são capazes de detectar uma imagem descrita por um rótulo a partir de uma sequência de imagens apresentadas por taxas de 113ms por imagem. Schyns e Oliva (1994) também relatam que indivíduos são capazes de reconhecer corretamente o tipo de cena (por exemplo, estrada, cidade) depois de um curto espaço de apresentação (150 ou mesmo 30ms). Além disso, as pessoas parecem fixar os olhos nas partes mais informativas da cena, o que requer a compreensão prévia da essência (HENDERSON; FERREIRA, 2004).

Os resultados das experiências apresentadas anteriormente mostram que o tempo necessário para reconhecer a cena é curto demais para permitir a análise detalhada de todos os seus elementos. Isso levanta a questão sobre as características de uma cena que as pessoas usam durante a primeira vista. Várias hipóteses têm sido propostas. Em primeiro lugar, uma análise do(s) objeto(s) poderia surgir e sugerir um tipo particular de uma cena (FRIEDMAN, 1979). Outra explicação dada por Biederman, Mezzanotte e Rabinowitz (1982) é que a essência é baseada em algumas características do nível de cena². A hipótese de que as características globais da cena³ são usadas no reconhecimento precoce foi confirmado pelos experimentos realizados por Schyns e Oliva (1994) nos quais filtros passa-baixa e passa-alta foram aplicados em imagens em escala de cinza de cenas. A cena representada por componentes de alta-frequência foi mais adequada quando as imagens eram apresentadas por períodos mais longos. Experimentos adicionais revelaram que as pessoas também são capazes de diferenciar alta e baixa frequência para o reconhecimento precoce (OLIVA; SCHYNS, 1997).

A cor ou textura pode ser também outra alternativa para facilitar o reconhecimento de cena. Nos experimentos de Oliva e Schyns (2000), pessoas são convidadas para nomear cenas apresentadas em cores normais, cores anormais e em escala de cinza. No caso em que a cor era de uma análise de categoria de cena, a presença de cores normais facilitou o reconhecimento, enquanto a presença de cores anormais tornou a tarefa mais difícil. Walker e Malik (2002) investigaram se o reconhecimento de cena pode ser explicado através de um modelo de textura. Os autores compararam os resultados de experimentos com seres

² Uma característica de nível de cena é um atributo que pode agregar estatísticas sobre características de nível inferior, tais como, porcentagem de cobertura de terra ou outras classificações descritivas, tais como diferenciar deserto, cidade, floresta tropical, etc.

³ São exemplos de características globais, o perímetro, o centro de massa, curvatura, área e momento de inércia.

humanos para validar o seu modelo de reconhecimento de textura e observar relações semelhantes.

Os resultados dos experimentos apresentados anteriormente indicam que informações globais e grosseiras são os primeiros detalhes a serem observados em uma cena e, posteriormente, analisam-se as informações locais. Embora as pessoas sejam capazes de extrair componentes de uma cena de alta-frequência em pouco tempo (OLIVA; SCHYNS, 2000), a maioria dos cientistas concorda que a cena é decomposta durante o processo de reconhecimento, ou seja, quanto mais se olha para uma cena, mais informações são extraídas (NAVON, 1977).

Outra questão interessante é quanto ao processamento das cenas naturais em relação aos outros estímulos, tais como rostos ou objetos. A hipótese desse processamento ser realizado separadamente é apoiada pelo fato de que há inúmeros casos descritos na literatura de pacientes que eram incapazes de reconhecer lugares, mesmo sendo capazes de identificar objetos e compreender as relações espaciais entre pontos diferentes (EPSTEIN et al., 1999). Os experimentos realizados por Epstein e Kanwisher (EPSTEIN; KANWISHER, 1998) mostram que não existe uma área distinta dentro do córtex hipocampal humano que responde na ressonância magnética funcional a imagens de cenas de ambientes internos e externos, assim como de pontos de referência (EPSTEIN et al., 1999). Essa região, denominada pelos autores de área para-hipocampal de lugares (*parahippocampal place area*, PPA), não responde ou responde fracamente quando a pessoa examinada vê imagens de rostos, objetos ou outros estímulos visuais, sem contexto espacial. Experiências adicionais demonstraram que a PPA responde a locais conhecidos e desconhecidos de uma maneira semelhante, e que é mais fortemente ativada por novidades do que por visões repetidas (EPSTEIN et al., 1999). A função precisa da PPA ainda é incerta. Hipóteses atuais questionam se essa região é envolvida na percepção de cena, participa da codificação da informação topográfica na memória ou ambos. Em conjunto, as experiências sugerem que existem unidades separadas no cérebro dedicadas ao processamento de cena.

2.3.2 Representação da Cena

Na seção anterior, foram descritos experimentos realizados por Potter (1976) demonstrando que as pessoas são capazes de detectar e compreender cenas complexas apresentadas em sequência com taxas de aproximadamente 113ms por imagem. No entanto, as mesmas experiências revelaram que com esse mesmo ritmo, os indivíduos foram incapazes de lembrar as imagens e que um atraso adicional é necessário para a consolidação da memória. Segundo Simons e Levin (1997), as pessoas não codificam muito do que eles vêem. Isto levanta questões sobre quais informações sobre uma cena são codificadas na memória e como elas são representadas.

Para responder à primeira pergunta, Aginsky e Tarr (2000) investigam que algumas

propriedades visuais de uma cena tais como cor, posição e presença do objeto exigem atenção especial quanto à codificação. Nos seus experimentos, indivíduos foram orientados a detectar mudanças em imagens com e sem avisos indicando as modificações. Os experimentos demonstraram que os avisos não tiveram influência sobre o tempo de detecção em mudanças de posição e presença, ao passo que o tempo de detecção foi menor para as alterações de cor quando os avisos eram fornecidos. Isso sugere que as propriedades que ajudam a determinar a disposição espacial de uma cena são codificadas com maior eficiência que as propriedades da superfície.

O problema da representação da cena na memória também foi estudado a fim de se descobrir se é algo específico ou invariante. Experimentos realizados por Christou e Bühlhoff (1997) mostraram que imediatamente após o treino, é mais provável que a representação possua um ponto de vista específico. Durante o treinamento, um modelo de computador de um ambiente interno com um conjunto limitado de direções foi apresentado a algumas pessoas. Na fase de teste, observou-se a capacidade dos observadores ao tentar reconhecer o mesmo ambiente com direções desconhecidas. Os observadores conseguiram reconhecer ambas situações, no entanto cenários familiares foram reconhecidos mais rapidamente e com maior frequência. No entanto, uma análise adicional sugere que representações específicas podem se tornar parcialmente invariantes (EPSTEIN; HIGGINS; THOMPSON-SCHILL, 2005).

2.4 Localização e Mapeamento

O mapeamento adotado é outro tópico importante no problema de reconhecimento de locais. Os métodos de mapeamento e localização de robôs móveis podem ser agrupados em dois paradigmas centrais (ENRICO et al., 2000): *i*) geométrico, tais como os trabalhos apresentados em (ARBEITER et al., 2012; NAGLA; UDDIN; SINGH, 2012; ZHANG; LIU; TAN, 2015) e *ii*) topológicos, ver (HAFAZ; LOO, 2015; DAWOOD; LOO, 2015; GERSTMAYR-HILLEN et al., 2013; CHIN et al., 2016). Os mapas geométricos representam o espaço total de navegação em um sistema de coordenadas globais. Por outro lado, em vez de se concentrar em informações geométricas rígidas, os métodos topológicos representam o ambiente como um grafo. Os vértices são os locais e a conexão entre eles são simbolizados pelas arestas (RADY, 2016; WU et al., 2014). A Figura 3 mostra um exemplo de mapa geométrico e topológico, respectivamente. Os mapas topológicos são eficientes na tarefa de localização de robôs móveis. Esses tipos de mapas são compactos e simples, exigem menor espaço de memória do sistema computacional e, conseqüentemente, aceleraram as etapas durante a navegação (CHENG; CHEN; LIU, 2015).

Outras propostas são formadas pela combinação de ambos os métodos. Rady (2016) apresenta uma abordagem baseada em visão computacional para construção e localização

de mapas híbridos. Wu et al. (2014) utilizam *QR Code* para formação de um mapa semântico ⁴. Esse mapa é combinado com uma representação topológica para construção de um mapa híbrido.

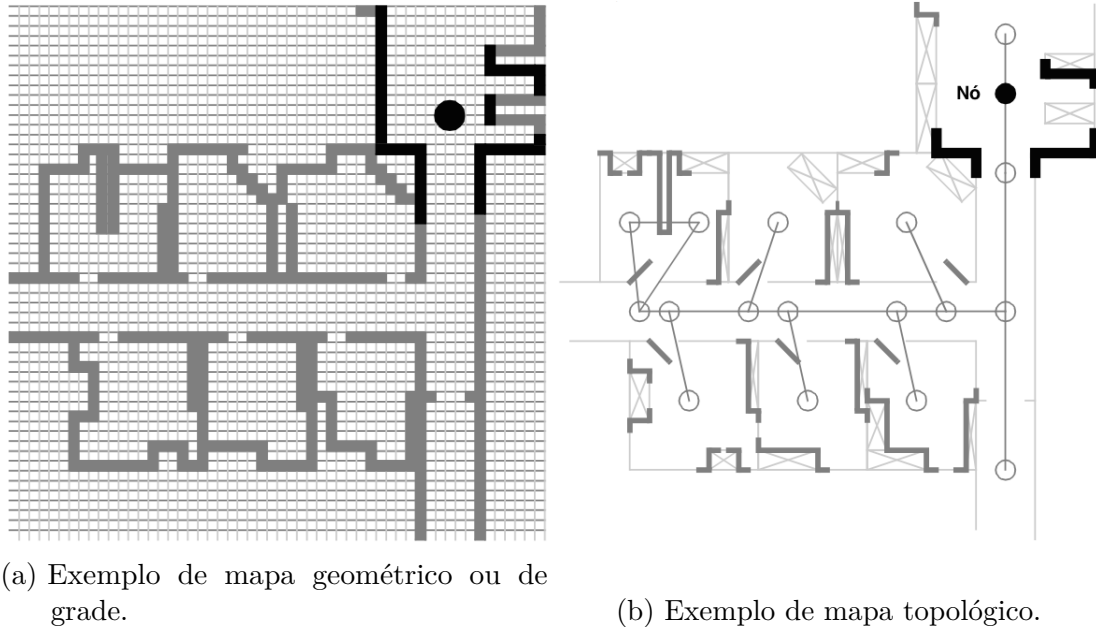


Figura 3 – Paradigmas de mapeamento e localização de robôs móveis.

2.5 Estrutura de um Sistema Típico de Reconhecimento de Padrões

O sistema de reconhecimento pode ser visto como um conjunto de componentes especializados que cooperam entre si através da troca de informações. A Figura 4 apresenta uma estrutura de um sistema típico de reconhecimento de padrões e mostra como os dados são passados entre os seus componentes. Os retângulos amarelos representam os componentes do sistema e correspondem às operações que integram as etapas de treinamento e classificação. As setas mostram a direção do fluxo de dados. Finalmente, os retângulos azuis descrevem os tipos de dados em todas as fases dos processos. As operações mais fundamentais, presentes em quase todos os sistemas de reconhecimento de padrões, são enquadradas com uma linha sólida. O diagrama é dividido em duas etapas que representam os processos de treinamento e classificação. Vários componentes foram marcados com uma linha sólida, uma vez que podem ser encontrados na maioria dos sistemas de reconhecimento de padrões.

Cada componente do sistema será apresentado a seguir em uma seção separada. Ademais, será discutido como cada elemento é implementado no sistema de reconhecimento

⁴ Pode-se definir um mapa semântico como uma representação de conhecimento hierárquica que possui, em um de seus níveis, um mapa espacial do ambiente.

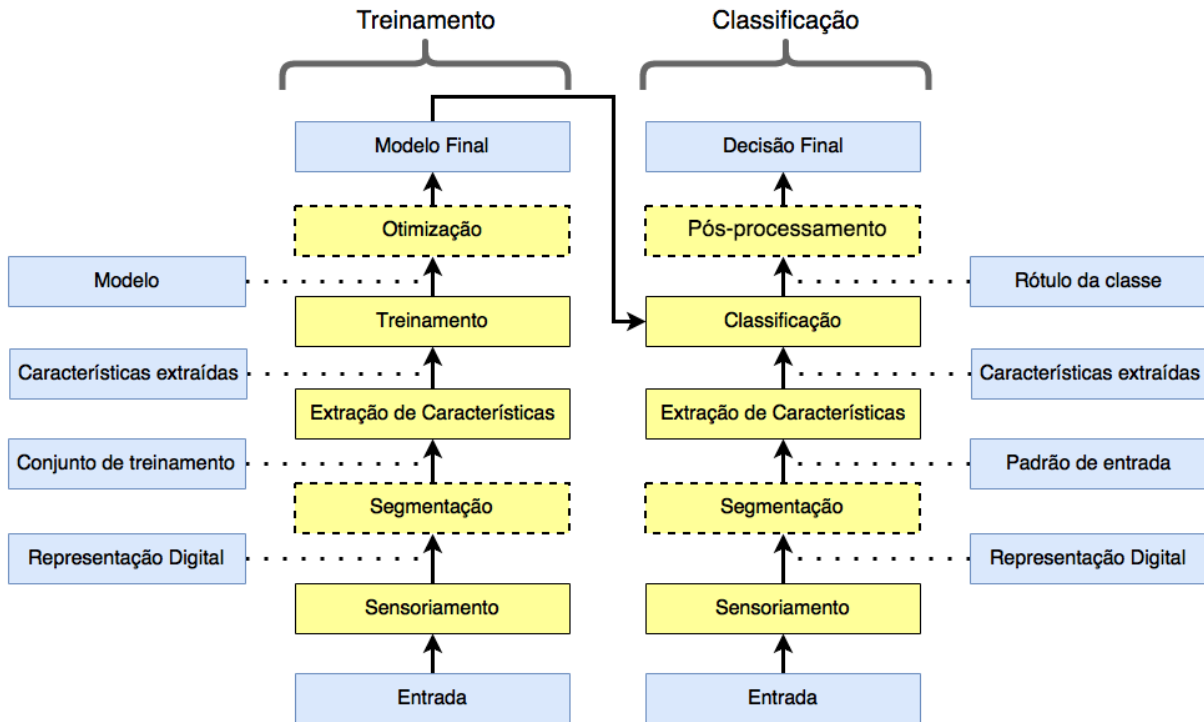


Figura 4 – Estrutura e fluxo de dados de um sistema típico de reconhecimento de padrões.

de locais apresentado nesta dissertação. As três primeiras operações são comuns para ambos os processos de treinamento e classificação e, portanto, serão primeiramente descritos.

2.5.1 Sensoriamento

A abrangência do termo padrão sugere a diversidade de estímulos que podem se tornar uma entrada de um sistema de reconhecimento de padrões. A fim de ser processada, essa informação deve ser detectada e armazenada em formato digital. Essa é uma tarefa para vários tipos de sensores. Embora a construção e o princípio do funcionamento desses sensores dependam fortemente da aplicação, na maioria dos casos eles podem ser considerados como conversores analógico-digital ou dispositivos de medição digital. Alguns exemplos são câmeras digitais, conversores AD de áudio, scanners a laser ou sensores de sonar. Há também aplicações que não necessitam de sensores especializados, devido os dados de entrada já estarem em um formato digital. A filtragem de *spam* é um exemplo desse tipo de aplicação.

Na maioria dos casos, os sensores podem ser caracterizados por vários parâmetros. No entanto, a importância destes parâmetros depende do problema e da construção do sensor. Alguns exemplos são: resolução, largura de banda e latência (fundamentais para sistemas de tempo real) e sensibilidade ou a relação sinal-ruído. É também desejável que o sensor seja capaz de suprimir a variabilidade inter-classe dos padrões (por exemplo, causada pela mudança de condições de iluminação).

Os dois tipos de sensores visuais mais usados no reconhecimento de locais com visão computacional são a câmera digital convencional e a omnidirecional (catadioptrica) (NAYAR, 1997). O primeiro tipo de câmera é o sensor visual mais comum, e, portanto, um sistema de reconhecimento que o adota pode ser empregado em diversas aplicações. Por outro lado, a câmera omnidireccional fornece um campo de visão horizontal de 360° , o que pode simplificar a tarefa de reconhecimento. No entanto, câmeras deste tipo costumam ser aplicadas quase que exclusivamente na localização de robôs móveis. A Seção 5.2 compara as imagens adquiridas usando câmeras de ambos os tipos. O sistema de reconhecimento de locais internos por visão computacional apresentado nessa dissertação foi avaliado com ambas as câmeras.

2.5.2 Segmentação e Pré-Processamento

Eventualmente, o padrão a ser reconhecido pode aparecer em algum fundo ou estar parcialmente sobreposto por outros elementos. Em tal situação, cada padrão individual necessita ser segmentado, antes de prosseguir com o processo de reconhecimento. Outro problema surge quando o padrão consiste em várias partes desconexas. Em tal caso, as partes segmentadas devem ser adequadamente combinadas de modo a formar uma entidade coerente. Esta operação é conhecida como agrupamento.

A segmentação e o agrupamento são problemas complexos e desafiadores em reconhecimento de padrões. Eles podem ser cruciais em determinadas aplicações como reconhecimento ótico de caracteres, reconhecimento de voz ou reconhecimento de objetos em ambientes reais. O último exemplo ilustra o quão difícil é o problema, uma vez que os objetos devem ser segmentados a partir de um fundo confuso. A segmentação não é necessária na tarefa de reconhecimento de locais. Dessa forma, a imagem completa forma um único padrão. Ademais, podem ser necessários outros processamentos nos dados de entrada, tais como filtragem de ruído ou normalização.

2.5.3 Extração de Características

O objetivo da extração de características é proporcionar uma nova representação do padrão de entrada que simplifique o problema de classificação. A nova representação deve ser tolerante à variabilidade inter-classe, além de enfatizar propriedades que são relevantes entre classes diferentes. Em outras palavras, deve consistir dos valores de atributos distintivos.

A divisão entre a extração de características e classificação é algo difícil de definir. No caso ideal, o processo de extração de características iria produzir uma representação que tornaria o problema de classificação trivial. Em tal caso, um classificador independente seria desnecessário. Por outro lado, pode-se imaginar uma tomada de decisão do classificador

diretamente dos dados brutos, sem transformação. Na prática, os algoritmos costumam cooperar entre si e a fronteira é definida por suas capacidades. Além disso, os classificadores possuem uma natureza universal, podendo ser utilizados para resolver vários problemas. Por outro lado, os métodos de extração de características servem como uma camada de adaptação e, geralmente, são mais adequados para uma tarefa em particular.

O desempenho do sistema de reconhecimento depende fortemente da qualidade do extrator de características. Portanto, a identificação correta das características distintivas do padrão é fundamental. No caso de reconhecimento de locais, por exemplo, é importante analisar descritores invariantes à translação, escala, iluminação, entre outros. O Capítulo 3 discute questões importantes dos extratores no contexto do reconhecimento de locais usando visão computacional. O mesmo capítulo também apresenta os métodos empregados na dissertação.

2.5.4 Classificação

O classificador é o elemento do sistema de reconhecimento que realiza o reconhecimento efetivo. Sua tarefa é de atribuir ao padrão de entrada, representado pelas características extraídas, uma das classes. A complexidade do problema de classificação depende de como os valores dos atributos diferem em cada classe em comparação com as divergências entre as classes. Isso resulta de vários fatores. Em primeiro lugar, a tarefa pode ser difícil em si. O reconhecimento de locais é um exemplo de um problema complexo devido à sua grande variabilidade dentro da classe. Em seguida, os dados de entrada podem ser modificados por ruídos. Finalmente, a complexidade depende do desempenho das operações de pré-processamento, tais como segmentação e extração de características. A qualidade do sensor também pode desempenhar um papel importante, uma vez que podem ser mais ou menos sensíveis a variações não desejadas (por exemplo, uma câmara pode ser equipada com controle automático de brilho).

Outras questões importantes relacionadas com o problema de classificação são descritos no Capítulo 4. O mesmo capítulo apresenta o estado da arte de diversos classificadores adotados na dissertação.

2.5.5 Pós-processamento

A tomada de decisão pelo classificador pode não ser a decisão final de um sistema de reconhecimento. Na verdade, em muitas situações, o desempenho e a robustez do sistema podem ser melhorados através da introdução de mecanismos suplementares. Esses mecanismos podem explorar outras fontes de informação ou apenas processar os dados provenientes de um único algoritmo de classificação.

O sistema de reconhecimento de padrões pode usar restrições *a priori*, informações

sobre o contexto em que ocorre o reconhecimento ou pode explorar um histórico de resultados anteriores. O sistema de reconhecimento de locais acoplado em um robô móvel pode, por exemplo, levar em consideração que o robô somente gira e não altera sua posição. Consequentemente, o robô deve estar ainda no mesmo lugar e a decisão pode ser realizada usando os dados atuais e anteriores. Tal sistema pode também fazer uso do mapa topológico do ambiente. Nesse caso, a robustez pode ser melhorada através de verificação de possíveis transições entre lugares (por exemplo, o robô não pode saltar a partir do escritório para a cozinha sem passar pelo corredor). A Seção 5.6 apresenta o suporte de decisão para localização proposto nessa dissertação.

A etapa de pós-processamento também é necessária se o sistema de reconhecimento é construído em torno de vários classificadores. Além disso, cada algoritmo de classificação pode sugerir várias hipóteses, mais prováveis. Em tal caso, os classificadores podem ser considerados como os peritos especializados em diferentes domínios. O sistema de reconhecimento de locais pode usar vários classificadores para categorizar imagens usando mais de um tipo de atributos. Ele pode também ser acoplado com outros algoritmos, tais como o reconhecimento de objetos.

2.5.6 Treinamento

De acordo com o esquema de aprendizagem supervisionada, o treinamento é realizado usando-se uma seleção de padrões rotulados representativos. O mesmo procedimento é seguido durante a aquisição do padrão tal como na classificação. Após a detecção dos dados de entrada, a segmentação pode ser realizada a fim de isolar os padrões individuais. Nesse momento, as amostras devem ser rotuladas por um “professor”.

O conjunto de padrões está, agora, pronto para ser utilizado na etapa de treinamento, bem como em outras partes do sistema ⁵. Como será discutido no Capítulo 4, é importante minimizar não apenas o erro das amostras no treinamento, mas também proporcionar a capacidade de generalização para novos padrões. A fim de atingir esses objetivos, as amostras podem ser divididas em dois subconjuntos: treino e teste. O algoritmo de treinamento tentará minimizar o erro de classificação no conjunto de treinamento, e o desempenho da generalização será avaliado utilizando o conjunto de teste. A divisão pode ser feita várias vezes, e em cada realização, diferentes pares de subconjuntos serão formados. Dessa forma, o desempenho final é avaliado com base em diversas observações. Tal método é conhecido como validação cruzada (ver Seção 5.7.2).

Até então, o treinamento é considerado como um processo realizado antes do sistema de reconhecimento ser utilizado. No entanto, em algumas aplicações pode ser mais adequado atualizar o conhecimento durante a execução do sistema. Esse tipo de aprendizagem

⁵ O processo de treinamento pode ser, por exemplo, utilizado para selecionar o “melhor” subconjunto das características distintivas dos padrões.

é conhecido como aprendizagem incremental, uma vez que permite a atualização do conhecimento do sistema incrementalmente, sem um retreino completo.

2.5.7 Otimização

Algumas aplicações necessitam que os sistemas de reconhecimento de padrões além de robustos e confiáveis, sejam eficientes e consumam poucos recursos. Essa é uma questão crucial, especialmente para sistemas em tempo real ou que realizam aprendizagem contínua. A fim de alcançar esses objetivos os modelos resultantes podem ser aprimorados ou métodos de otimização adicionais podem ser aplicados. Nessa dissertação, é feita uma comparação entre a Máquina de Aprendizagem Mínima (*Minimal Learning Machine*, MLM) e sua versão melhorada, a Máquina de Aprendizagem Mínima com Vizinhos Mais Próximos (MLM-*Nearest Neighbors*, MLM-NN).

2.6 Conclusão

Neste capítulo, foram discutidas as principais questões relacionadas com o reconhecimento de locais e percepção de ambiente no que diz respeito a sistemas de reconhecimento artificiais e humanos. Inicialmente, os termos de cena e local foram definidos com base na literatura. Em seguida, analisou-se o problema de reconhecimento de local, bem como as questões-chave e termos relacionados com a concepção e teste de um sistema de reconhecimento de lugar. Apresentou-se uma revisão de uma seleção de publicações sobre reconhecimento e percepção de cenas. Posteriormente, discutiu-se sobre os tipos de métodos de mapeamento e localização de robôs móveis. Finalmente, o reconhecimento de lugar foi apresentado como um caso especial de reconhecimento de padrões e sua complexidade foi apresentada. Da mesma forma, utilizou-se a estrutura de um sistema de reconhecimento de padrões típico, a fim de demonstrar a posição e o papel de cada parte do sistema dentro do processo de reconhecimento. As principais partes serão descritas em detalhe nos seguintes capítulos.

3 Extratores de Características

Neste capítulo, os métodos empregados na etapa de extração de características são descritos. A Seção 3.1 discute sobre o conceito de extração de características em imagens. Na Seção 3.2, o termo invariância é definido. Os Momentos Espaciais, Centrais e Invariantes de Hu são apresentados na Seção 3.3. A Seção 3.4 mostra os extratores baseados em textura, LBP e descritores de Haralick, adotados no trabalho. Por último, os descritores de pontos-chave, ou pontos de interesse, SIFT e SURF, são explanados na Seção 3.5.

3.1 Descrição de Características

Uma imagem digital pode conter uma elevada quantidade de dados. Uma figura de tamanho médio, por exemplo, consiste de 1024 *pixels* por linha e 768 por coluna. O número total de *pixels* pode ser ainda maior em imagens profissionais.

As imagens podem ser pensadas como uma função discreta com dimensão bastante elevada em um espaço $\mathbb{R}^N$, onde N é o número de *pixels*. Uma maneira de diminuir essa dimensão é reduzindo a escala da imagem. Essa abordagem pode ser alcançada, por exemplo, retirando-se os *pixels* similares.

No entanto, a redução da escala não ajuda significativamente a reduzir sua dimensão, uma vez que há uma vasta quantidade de *pixels* sem utilidade. A melhor maneira de lidar com esse problema é extraindo as características essenciais da imagem que podem representá-la por completo. As características detectadas podem significar termos abstratos, tais como: área, fronteira, número ou posição de bordas, cor, entre outros.

Os descritores de características devem ser robustos e invariantes às mudanças. Além disso, precisam ser extraídos em um prazo razoável e com menor complexidade possível. É importante ressaltar que não há fórmulas gerais para um descritor ideal, uma vez que a adequação está fortemente ligada à aplicação.

3.2 Invariância

O reconhecimento de padrões em imagens pode ser realizado, de modo geral, através da força bruta ou por descritores invariantes. Na abordagem da força bruta, é feito uma busca no espaço paramétrico de todas as possíveis degradações da imagem. Isso significa que o conjunto de treinamento deve conter não apenas as classes, mas também as rotações, mudanças de escala e versões de deformação. Claramente, esse procedimento possui uma

elevada complexidade computacional sendo praticamente inaplicável (FLUSSER; ZITOVA; SUK, 2009).

A abordagem com descritores invariantes é a mais promissora e amplamente utilizada. Sua ideia básica consiste em descrever os padrões por um conjunto de quantidades mensuráveis chamadas invariantes que são indiferentes a deformações particulares e que fornecem poder de discriminação suficiente para distinguir padrões de diferentes classes. As seguintes operações podem ser invariantes em um descritor: translação, rotação, escala, perspectiva, transformação das coordenadas do espaço, convolução, cor, entre outras.

Do ponto de vista matemático, Γ é uma função definida sobre o espaço de todas as funções de imagens admissíveis que não mudam seus valores sobre um operador de degradação D , o qual satisfaz a condição $\Gamma(f) = \Gamma(D(f))$ para qualquer função de imagem f . Esta propriedade é chamada invariância. Na prática, a fim de permitir o impacto da segmentação imperfeita, variabilidade intraclasses e ruído é, normalmente, formulado uma restrição mais fraca: $\Gamma(f)$ que não deve ser significativamente diferente de $\Gamma(D(f))$. Outra propriedade importante de Γ é a discriminabilidade. Para padrões pertencentes a diferentes classes, Γ deve ter valores consideravelmente distintos. Evidentemente, essas duas condições são antagonistas - quanto maior a invariância, menor o poder de discriminação e vice-versa. A escolha da propriedade entre invariância e discriminação é uma importante tarefa no reconhecimento de padrões baseados em características (FLUSSER; ZITOVA; SUK, 2009).

3.3 Descritores baseados em Momentos

Os momentos são amplamente aplicados em tarefas de análise de imagens, tais como reconhecimento de padrões, classificação de objetos, estimativa de postura, localização *indoor*, codificação e reconstrução de imagens. Um conjunto de momentos computados de uma imagem digital, geralmente, representa características globais da forma da imagem e fornecem informações relevantes sobre diferentes tipos de características geométricas da mesma. A capacidade de representação de características dos momentos tem sido amplamente utilizada em diversas áreas de visão computacional e robótica. O acelerado avanço das pesquisas em processamento de imagens propiciou o desenvolvimento de vários tipos de momentos, cada um com suas próprias vantagens em áreas específicas. Os Momentos Espaciais, por exemplo, foram os primeiros aplicados a imagens e são computacionalmente simples (MUKUNDAN; RAMAKRISHNAN, 1988). Momentos de Hu são invariantes à translação, mudança de escala, espelhamento e rotação (GONZALEZ; WOODS, 2010).

As propriedades dos momentos baseiam-se em conceitos de Estatística e Mecânica. Momentos de ordem zero, um e dois de uma função de densidade de probabilidade representam a probabilidade total, a esperança e variância, respectivamente. Na Mecânica,

esses momentos em uma distribuição espacial de massa fornecem a massa total, posição do centróide e valores de inércia. Esses atributos podem ser utilizados para construir vetores de características invariantes à translação, rotação e escala. Enquanto momentos de ordem entre zero e três representam características mais gerais da imagem, momentos de alta ordem costumam ter um maior nível de detalhamento da imagem, porém são mais sensíveis a ruídos (MUKUNDAN; RAMAKRISHAN, 1988).

3.3.1 Definição Geral dos Momentos

Segundo Flusser, Zitova e Suk (2009), momentos são quantidades escalares usados para caracterizar uma função e para capturar suas características significativas. Eles são amplamente empregados na Estatística para a descrição da forma de uma função de densidade de probabilidade e no clássico corpo rígido para medir a distribuição de massa em um corpo. Do ponto de vista matemático, momentos são projeções de uma função em uma base polinomial.

O *momento geral* M_{pq} de uma imagem $f(x, y)$ é definido como (FLUSSER; ZITOVA; SUK, 2009)

$$M_{pq} = \int \int_D p_{pq}(x, y) f(x, y) dx dy \quad (3.1)$$

em que p e q são inteiros não negativos, $r = p + q$ é chamado a *ordem* do momento e $p_{00}(x, y), p_{10}(x, y), \dots, p_{kj}(x, y)$, são funções de base polinomiais definidas em D .

3.3.2 Momentos Espaciais

Os Momentos Espaciais m_{pq} de uma imagem $f(x, y)$ podem ser definidos como

$$m_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x, y) dx dy \quad (3.2)$$

em que p e q são inteiros não negativos e a imagem $f(x, y)$ possui uma integral finita não nula .

Os Momentos Espaciais de baixa ordem possuem um significado intuitivo: m_{00} é a “massa” da imagem (para imagens binárias, m_{00} é a área do objeto); m_{10}/m_{00} e m_{01}/m_{00} definem o centro de gravidade ou centroide da imagem. Os momentos de segunda ordem m_{20} e m_{02} descrevem a “distribuição de massa” da imagem quanto aos eixos das coordenadas. Eles são também conhecidos como momentos de inércia na Mecânica.

Se a imagem for considerada uma função de densidade de probabilidade, m_{10} e m_{01} serão valores médios. No caso da média zero, m_{20} e m_{02} são variâncias das projeções horizontais e verticais e m_{11} é a covariância entre eles. Deste modo, os momentos de segunda

ordem definem a orientação da imagem. Segundo Flusser, Zitova e Suk (2009), Momentos Espaciais de segunda ordem podem ser usados para encontrar a posição normalizada da imagem. Na Estatística, duas características dos momentos de alta ordem costumam ser usadas: *skewness* e *kurtosis*. A assimetria da projeção horizontal é definida como $m_{30}/\sqrt{m_{20}^3}$ e da projeção vertical como $m_{03}/\sqrt{m_{02}^3}$. A assimetria permitem distinguir as projeções simétricas. Se a projeção é simétrica em relação a média, então a assimetria é igual a zero. A curtose mede os “picos” das funções de densidade de probabilidade e também é definida para cada projeção separadamente: a curtose horizontal como m_{40}/m_{20}^2 e a curtose vertical como m_{04}/m_{20}^2 .

A representação discreta da Equação 3.2 é mostrada a seguir:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y). \quad (3.3)$$

3.3.3 Momentos Centrais

Algumas aplicações exigem momentos invariantes à translação. Essa propriedade pode ser alcançada através dos Momentos Centrais que são computados a partir dos Momentos Estatísticos como mostra a seguinte equação. Considere

$$\mu_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy \quad (3.4)$$

em que $\bar{x} = m_{10}/m_{00}$ e $\bar{y} = m_{01}/m_{00}$ representam as coordenadas do centro de gravidade da imagem. Da Equação 3.4, pode-se inferir que

$$\begin{aligned} \mu_{00} &= m_{00} \\ \mu_{10} &= \mu_{01} = 0 \end{aligned} \quad (3.5)$$

Alternativamente, a representação discreta dos Momentos Centrais é dada por

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (3.6)$$

Os Momentos Centrais de primeira e segunda ordem podem ser derivados diretamente dos Espaciais como mostra a seguir

$$\mu_{pq} = \frac{m_{pq}}{m_{00}} - \left(\frac{m_{10}}{m_{00}} \right)^p \quad (3.7)$$

Nesse contexto, os Momentos Centrais de baixa ordem podem ser obtidos a partir do Momentos Espaciais na Equação 3.7 por

$$\begin{aligned}
 \mu_{10} &= \frac{m_{10}}{m_{00}} - \left(\frac{\mu_{10}}{\mu_{00}} \right) = 0 \\
 \mu_{01} &= \frac{m_{01}}{m_{00}} - \left(\frac{\mu_{01}}{\mu_{00}} \right) = 0 \\
 \mu_{20} &= \frac{m_{20}}{m_{00}} - \left(\frac{\mu_{10}}{\mu_{00}} \right)^2 = \left(\frac{\mu_{20}}{\mu_{00}} \right) - \bar{x}^2 \\
 \mu_{02} &= \frac{m_{02}}{m_{00}} - \left(\frac{\mu_{01}}{\mu_{00}} \right)^2 = \left(\frac{\mu_{02}}{\mu_{00}} \right) - \bar{x}^2 \\
 \mu_{11} &= \frac{m_{11}}{m_{00}} - \left(\frac{\mu_{10}}{\mu_{00}} \right) * \left(\frac{\mu_{01}}{\mu_{00}} \right) = \left(\frac{\mu_{11}}{\mu_{00}} \right) - \bar{x} * \bar{y}
 \end{aligned} \tag{3.8}$$

A mudança nos Momentos Centrais em relação aos Momentos Espaciais determina sua invariância a translação. Essa propriedade significa que independente da localização do objeto na imagem, seus momentos serão computados como se o mesmo estivesse no centro da região.

3.3.4 Momentos Centrais Normalizados

A desvantagem dos Momentos Centrais e Espaciais é sua dependência sobre o tamanho do objeto. Esse é um problema quando tentamos comparar objetos que se encontram a diferentes distâncias. Para permitir tal comparação, é preciso tornar os momentos invariantes à escala. Normalmente, a área do objeto é utilizada como fator de escala. Dessa forma, os Momentos Centrais μ_{pq} são divididos por potências da área para obter os Momentos Centrais Normalizados η_{pq} .

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\alpha} \tag{3.9}$$

em que $\alpha = (p + q)/2 + 1$.

A principal vantagem do Momentos Centrais Normalizados é sua invariância em relação ao tamanho do objeto.

3.3.5 Momentos Invariantes de Hu

Hu (1962) propôs um método para descrever formas através de sete momentos que são invariantes à escala, rotação e translação. Considerando os Momentos Centrais

Normalizados η_{pq} definidos na Equação 3.9, os sete Momentos de Hu, também conhecidos como Momentos Invariantes, são definidos a seguir.

$$\begin{aligned}
\phi_1 &= \eta_{20} + \eta_{02} \\
\phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
\phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
\phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
\phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{03} - \eta_{12})^2] - \\
&\quad 3(\eta_{21} + \eta_{03})^2 + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
\phi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
&\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
\phi_7 &= (3\eta_{12} - \eta_{30})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
&\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
\end{aligned} \tag{3.10}$$

Os sete momentos são invariantes à espalhamento, ou seja, duas formas idênticas porém invertidas terão valores iguais de similaridade. É possível inferir que quando um objeto é rotacionalmente simétrico, os sete Momentos de Hu serão iguais a zero.

3.4 Descritores baseados em texturas

3.4.1 Padrões Binários Locais

Os Padrões Binários Locais (*Local Binary Patterns*, LBP) são simples e eficientes descritores de texturas. O operador LBP original, formulado por Ojala, Pietikainen e Harwood (1994), atribui um rótulo (representado por um número binário) a cada *pixel* analisado. Para atribuir esse rótulo, é verificado se cada vizinho do *pixel* em análise, em um raio considerado, excede, ou não, um limiar (*threshold*) que é determinado pelo valor do *pixel* central, cujo rótulo está sendo determinado. Cada vizinho que extrapolar esse limiar (*threshold*), ou se igualar a ele, será marcado com o dígito binário 1, caso contrário, receberá 0. A junção desses dígitos binários representará o *pixel* central. Faz-se necessário seguir uma lógica para determinar qual o dígito mais significativo, em outras palavras, a partir de que vizinho o número começará a ser formado, visto que a vizinhança é circular. Também é preciso observar que quando se determina o dígito mais significativo, automaticamente está sendo atribuído um peso a cada um dos dígitos. A soma da multiplicação de cada dígito pelo seu respectivo peso retornará o que é referenciado como código LBP. Como a vizinhança consiste de 8 *pixels*, um total de $2^8 = 256$ padrões podem ser codificados. A ilustração do processo é mostrada na Figura 5.

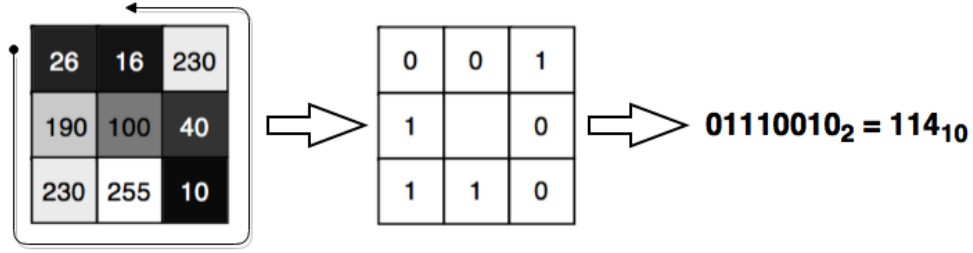


Figura 5 – Para uma vizinhança 3×3 na imagem, o padrão binário local é determinado pela comparação do *pixel* central com seus vizinhos na ordem fixada atribuída pela seta. O padrão binário gera um número que é computado no histograma LBP.

Mais formalmente, consideremos uma imagem em escala de cinza $\mathbf{I}$. Seja g_c o valor de intensidade de um *pixel* arbitrário $g_c = i(x, y)$, g_p para $p = 0, \dots, 7$ denota o valor de intensidade na vizinhança espacial 3×3 de (x, y) . O rótulo LBP de padrão de intensidade da vizinhança espacial do *pixel* (x, y) é

$$LBP = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (3.11)$$

onde $P = 8$ e $s(\cdot)$ é a função degrau:

$$s(x) = \begin{cases} 1, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases} \quad (3.12)$$

O operador LBP pode ser facilmente generalizado para qualquer vizinhança circular P com raio R . O padrão g_p é então definido como:

$$g_p = \mathbf{I}(x_p, y_p) \text{ para } p = 0, \dots, P - 1 \quad (3.13)$$

em que

$$x_p = x + R \cos\left(\frac{2\pi p}{P}\right), \quad (3.14)$$

$$y_p = y + R \sin\left(\frac{2\pi p}{P}\right), \quad (3.15)$$

os valores de g_p são interpolados. A Equação 3.11 é, portanto, utilizada para computar os rótulos do LBP.

Segundo Pietikäinen et al. (2011), uma das propriedades mais importante do operador LBP, em aplicações do mundo real, é a sua invariância. Essa característica é percebida quando acontecem mudanças monotônicas nos níveis de cinza, que podem ser

causadas por alterações na iluminação do ambiente, no momento da captura da imagem. Outro atributo igualmente importante é a sua simplicidade computacional, que torna possível enfrentar desafios nos quais as imagens precisam ser analisadas em tempo real.

Para um determinado padrão LBP, há oito números binários diferentes que são formados diante da escolha do dígito como mais significativo. Obviamente, alguns vizinhos serão mais importantes que outros, e, por conseguinte, é possível perceber que essa representação é sensível à rotação (PETROU; SEVILLA, 2006). Diante desse problema, uma estratégia adequada para remover essa invariância consiste em escolher o menor número dentre todos os que podem ser formados pela variação do dígito mais significativo.

3.4.1.1 Padrões Uniformes

A abordagem discutida até o momento não trata de redução de dimensionalidade, pois o operador LBP produz uma imagem do mesmo tamanho da original. É importante, no entanto, ressaltar que essa imagem não consiste de valores de intensidades, mas sim de padrões. Uma prática comum é descartar a localização espacial de padrões e preservar apenas os mais frequentes. Embora o procedimento anterior reduza o espaço de característica, a dimensionalidade ainda será bastante elevada.

Pietikäinen et al. (2011) propuseram uma metodologia, denominada Padrões Uniformes (*Uniform Patterns*), para reduzir, significativamente, a dimensionalidade gerada pelo operador LBP original. Seja $(LBP)_b$ o valor binário de um rótulo LBP, a medida de uniformidade $U(LBP)$ é o número de transições *bit a bit* de 0 para 1, e vice-versa, em $(LBP)_b$, que é considerada circular. O padrão uniforme é um padrão binário local que tem a sua medida de uniformidade no máximo igual a 2. Na Figura 6, o valor binário do primeiro padrão é 11001100 e possui quatro transições (não uniforme). Em contrapartida, o valor binário do segundo é 11000111 e há, apenas, duas transições (uniforme). Portanto, o número de rótulos diferentes para a vizinhança de P pixels é $P(P-1)+3$. Uma vizinhança 3×3 possui 59 padrões únicos (58 para os padrões uniformes e 1 para os não uniformes). Ojala, Pietikainen e Maenpää (2002) mostrou que uma vizinhança 3×3 de padrões uniformes representa, aproximadamente, 90% de todos os padrões da textura. A Figura 7 mostra uma aplicação do LBP com padrões uniformes aplicado ao problema de localização de robôs móveis.

3.4.2 Matriz de Coocorrência e os Descritores de Haralick

A Matriz de Coocorrência (*Gray-Level Co-occurrence Matrix*, GLCM) é uma técnica utilizada dentro da área de análise de texturas, que foi desenvolvida pelo pesquisador Robert M. Haralick (HARALICK; SHANMUGAM; DINSTEN, 1973). A GLCM é um método estatístico de segunda ordem, pois considera as coocorrências existentes entre

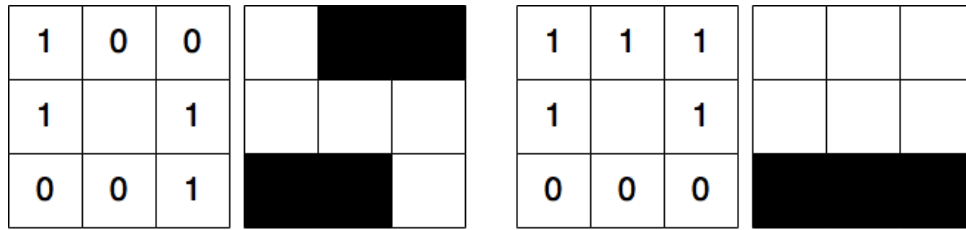
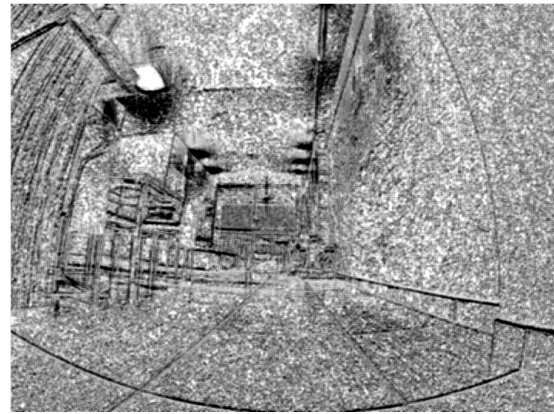


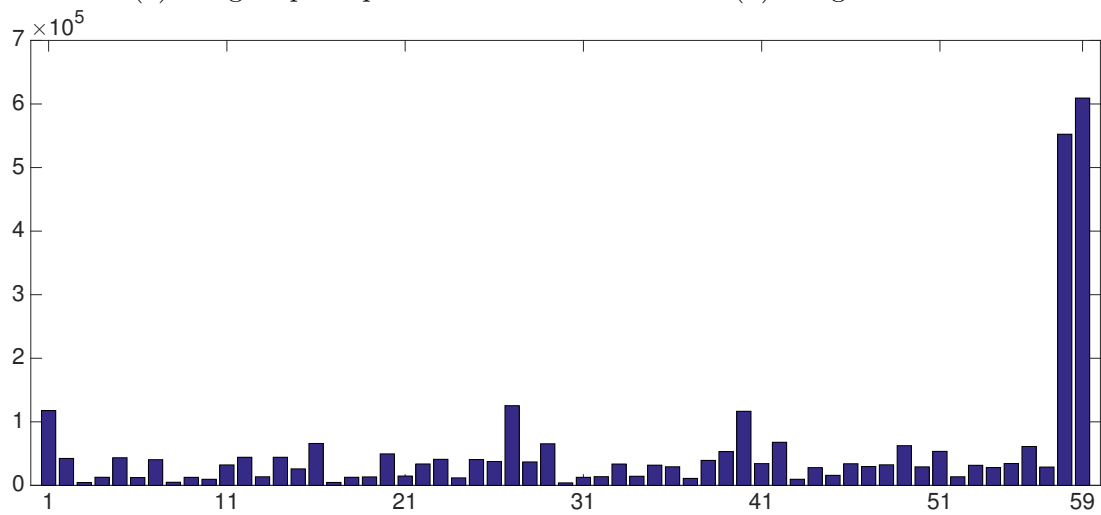
Figura 6 – Diferença entre os padrões não uniforme e uniforme. Da esquerda para a direita; a primeira e segunda matriz mostram um exemplo de um padrão não uniforme. A terceira e quarta matriz exibem características do LBP uniforme. O padrão uniforme possui, apenas, duas regiões conectadas.



(a) Imagem principal



(b) Imagem LBP



(c) Histograma da imagem LBP

Figura 7 – Exemplo do LBP com padrões uniformes aplicado ao problema de localização de robôs móveis: (a) Imagem de entrada. (b) Resultado da aplicação do LBP com contraste melhorado para propósitos de visualização e (c) histograma da imagem após utilizar o LBP.

pares de *pixels*. Isto é, não há uma análise individual dos *pixels*, mas sim de um conjunto de *pixels* relacionados através de algum padrão.

A GLCM é uma matriz quadrada que mantém informações das intensidades relativas dos *pixels* em uma dada imagem. Nessa técnica, são calculadas as probabilidades de co-ocorrências entre dois níveis de cinza, dada uma certa distância D e uma orientação θ . Este pode assumir os valores 0, 45, 90, 135, 180, 225, 270, 315 como mostra a Figura 8. Consideremos uma imagem com L níveis de cinza. Na matriz de co-ocorrência $\mathbf{G}_D^\theta$ de ordem $L \times L$, o elemento $g(i, j)$ é o número de ocorrências de pares de níveis de cinza i e j com distância D e que seguem a direção θ .

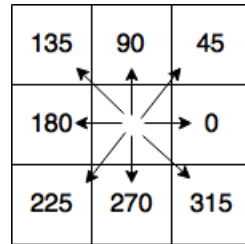


Figura 8 – Direções de adjacência da GLCM.

A Figura 9 mostra um exemplo de uma matriz de coocorrência $\mathbf{G}$ com distância $D = 1$ e direção horizontal $\theta = 0$, ou seja, o vizinho de um *pixel* é definido como um *pixel* imediatamente à sua direita. Observamos que o elemento $(1, 1)$ de $\mathbf{G}$ é 1, porque há apenas uma ocorrência na imagem $\mathbf{I}$ de um *pixel* com valor 1 com outro *pixel* avaliado com 1 imediatamente à sua direita. Similarmente, o elemento na linha 6 e coluna 2 da matriz $\mathbf{G}$ tem valor 3, o que indica que há três ocorrências em $\mathbf{I}$ de um *pixel* com valor de 6 e com um *pixel* de valor 2 imediatamente à sua direita. Esse processo é repetido até que a matriz GLCM esteja completa.

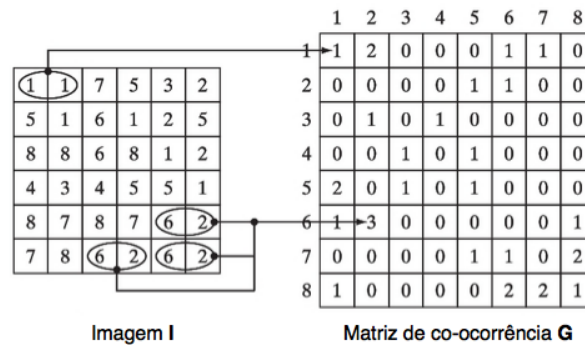


Figura 9 – GLCM com distância $D = 1$ e direção $\theta = 0$ de uma imagem $\mathbf{I}$ (GONZALEZ; WOODS, 2010).

A matriz de coocorrência pode ser normalizada pelo número total de pares de *pixels*, N , o qual é igual à soma dos elementos de $\mathbf{G}$. Assim, a quantidade $p(i, j) = g(i, j)/N$ é uma estimativa que um par de pontos tenham valores (z_i, z_j) (GONZALEZ; WOODS, 2010).

Uma análise computacional da obtenção da matriz GLCM a partir da representação matricial da imagem mostra que o relacionamento escolhido (*offset*) não interfere na complexidade do algoritmo. Para um dado *offset*, a imagem deve ser percorrida *pixel* a *pixel* para verificar se algum *pixel* possui o relacionamento estabelecido e a quantidade deve ser incrementada na matriz GLCM. Assim, se a imagem possui M *pixels*, a complexidade será $O(M)$, independente do *offset* adotado.

Haralick, Shanmugam e Dinstein (1973) descreveram 14 medidas estatísticas a partir da matriz de coocorrência que podem ser usadas para representar características presentes em texturas. Esses descritores (f_1 a f_{14}) foram definidos conforme as equações abaixo e são comentados a seguir.

- $p(i, j)$: (i, j) -ésima entrada em uma matriz normalizada de dependência espacial de níveis de cinza $g(i, j)/N$;
- $p_x(i)$: i -ésima entrada em uma matriz de probabilidade marginal obtida pela soma das linhas de $p(i, j)$, $\sum_{j=1}^L p(i, j)$;
- L : Número de níveis de cinza na imagem quantizada;
- $\sum_i$ e $\sum_j$: forma simplificada de $\sum_{i=1}^L$ e $\sum_{j=1}^L$, respectivamente;
- $p_{x+y}(k) = \sum_{i=1}^L \sum_{\substack{j=1 \\ i+j=k}}^L p(i, j)$ para $k = 2, 3, \dots, 2L$;
- $p_{x-y}(k) = \sum_{i=1}^L \sum_{\substack{j=1 \\ |i-j|=k}}^L p(i, j)$ para $k = 0, 1, \dots, 2L - 1$.
- Segundo Momento Angular:

$$f_1 = \sum_i \sum_j [p(i, j)]^2. \quad (3.16)$$

- Contraste:

$$f_2 = \sum_{k=0}^{L-1} k^2 [p_{x-y}(k)]. \quad (3.17)$$

- Correlação:

$$f_3 = \frac{\sum_i \sum_j (i, j) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}, \quad (3.18)$$

onde μ_x, μ_y, σ_x e σ_y são as médias e desvios padrões de p_x e p_y .

- Soma dos quadrados: Variância

$$f_4 = \sum_i \sum_j (i - \mu)^2 p(i, j). \quad (3.19)$$

- Momento da diferença inversa:

$$f_5 = \sum_i \sum_j \frac{1}{1 + (i + j)^2} p(i, j) \quad (3.20)$$

- Média da soma:

$$f_6 = \sum_{i=2}^{2L} i p_{x+y}(i). \quad (3.21)$$

- Variância da soma:

$$f_7 = \sum_{i=2}^{2L} (i - f_6)^2 p_{x+y}(i). \quad (3.22)$$

- Entropia da soma:

$$f_8 = - \sum_{i=2}^{2L} p_{x+y}(i) \log [p_{x+y}(i)]. \quad (3.23)$$

Desde que alguma das probabilidades possa ser zero e o $\log(0)$ não é definido, é recomendado que o termo $\log(p + \epsilon)$, onde ϵ é uma constante arbitrária pequena, seja usado no lugar de $\log(p)$.

- Entropia:

$$f_9 = - \sum_i \sum_j p(i, j) \log [p(i, j)]. \quad (3.24)$$

- Variância da diferença:

$$f_{10} = \text{variância de } p_{x+y}. \quad (3.25)$$

- Entropia da diferença:

$$f_{11} = - \sum_{i=0}^{L-1} p_{x-y} \log [p_{x-y}(i)]. \quad (3.26)$$

- Medidas de informação da correlação:

$$f_{12} = \frac{HXY - HXY1}{\max [HX, HY]}, \quad (3.27)$$

$$f_{13} = 1 - \exp [-2.0(HXY2 - HXY)]^{1/2}, \quad (3.28)$$

onde:

- $HXY = - \sum_i \sum_j p(i, j) \log [p(i, j)];$
- HX e HY são entropias de p_x e p_y ;
- $HXY1 = - \sum_i \sum_j p(i, j) \log [p_x(i)p_y(j)];$
- $HXY2 = - \sum_i \sum_j p_x(i)p_y(j) \log [p_x(i)p_y(j)].$

- Coeficiente de correlação máximo:

$$f_{14} = (\text{segundo maior autovalor de } \mathbf{Q})^{1/2}, \quad (3.29)$$

onde:

$$\mathbf{Q} = \sum_k \frac{p(i, k)p(j, k)}{p_x(i)p_y(k)}. \quad (3.30)$$

O segundo momento angular, ou energia, mede a uniformidade da textura, ou seja, a quantidade de repetições de pares de *pixels*. Essa medida terá valores altos quando os valores da matriz de coocorrência estiverem concentrados, ou seja, quando houver grande repetição na variação dos níveis de cinza.

A diferença entre o maior e o menor valor de um determinado conjunto contíguo de *pixels* é definida como frequência espacial. A expressão do contraste indica que uma imagem com valor baixo nesse parâmetro não é necessariamente caracterizada por uma distribuição de tons de cinza estreita, mas sim por uma baixa frequência espacial, possuindo valores da matriz de co-ocorrência concentrados próximos à diagonal principal.

A entropia mede a desordem da imagem. Portanto, essa medida assume valores elevados quando a imagem possui textura não uniforme. A entropia é inversamente relacionada ao momento angular.

A variância é uma medida da heterogeneidade da textura. O valor da variância aumenta quando os valores dos tons de cinza diferem de sua média.

A correlação mostra a dependência linear dos tons de cinza em uma imagem. Altos valores de correlação implicam uma relação linear entre os níveis de cinza dos pares de *pixels*.

A homogeneidade, também conhecida como Momento da Diferença Inversa, mede, como sugere o nome, a uniformidade de uma imagem. Essa medida é sensível a valores próximos à diagonal da matriz de co-ocorrência e, portanto, ao baixo contraste da textura. A homogeneidade está inversamente relacionada com o contraste e a energia da imagem.

3.5 Descritores de pontos-chave

Descritores de pontos-chave extraem características locais da imagem em alguma distribuição de pontos, também chamados de pontos-chave. A ideia por trás dessa abordagem consiste no fato de que os pontos-chave detectados estão localizados em regiões, relativamente estáveis e confiáveis, que possuem informação relevante sobre o conteúdo da imagem local. Dessa forma, a performance durante a classificação de padrões depende da confiabilidade, acurácia e repetibilidade da detecção dos pontos-chave para cada classe de padrões e da quantidade de informação relevante adquirida por eles. Algumas técnicas comuns na detecção de pontos de interesse incluem Förster (FÖRSTNER, 1994), Laplaciano

(LINDBERG, 1998), Diferença de Gaussianas (*Difference of Gaussians*, DoG) (LOWE, 2004), Harris-Laplace (MIKOLAJCZYK; SCHMID, 2004) além de algumas combinações entre eles. Assim, vetores de características podem ser obtidos através da computação sobre as regiões da imagem local adjacentes aos pontos-chave encontrados. A maior vantagem desses métodos é sua compacidade de representação, visto que a dimensão dos descritores de características costuma ser significativamente menor que do número total de *pixels* na imagem. No entanto, esses métodos possuem limitações para classes gerais de objetos devido a maioria serem projetados para casos específicos. Os dois métodos mais conhecidos são a Transformada de Características Invariante à Escala (*Scale-Invariant Feature Transformation*, SIFT) e as Características Robustas Aceleradas (*Speeded Up Robust Features*, SURF) explicados nas próximas seções.

3.5.1 SIFT

A Transformada de Características Invariante à Escala (*Scale-Invariant Feature Transformation*, SIFT) é um descritor invariante à escala e rotação desenvolvida por (LOWE, 2004). As características extraídas pelo SIFT estão bem localizadas nos domínios do espaço e da frequência, reduzindo assim a possibilidade de interrupção por oclusão, desordem ou ruído. A técnica é parcialmente invariante à mudança de iluminação e mudança no ponto de vista. As características extraídas são distintivas o que permite a sua comparação com um grande banco de dados de características de muitas imagens. Possibilitando assim, uma base para o reconhecimento de objetos e cenas.

O processo de extração do SIFT consiste em quatro etapas:

1. **Detecção de extremos (espaço escala).** Nesta primeira etapa é feita a procura para todas as escalas e localizações de uma imagem. Isto é realizado utilizando-se a Diferença de filtros Gaussianos (*Differences of Gaussians*, DoG) (Figura 10) de modo a se identificar pontos de interesse invariáveis à escala e orientação. A convolução de uma imagem $I(x, y)$ com filtros Gaussianos em diferentes escalas σ é dado na Equação 3.31.

$$L(x, y, \sigma) = G(x, y, \sigma) \star I(x, y) \quad (3.31)$$

A diferença entre duas imagens Gaussianas com escalas $k\sigma$ e σ é apresentada na Equação 3.32.

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (3.32)$$

2. **Localização dos *keypoints* (pontos-chaves).** Para cada localização, um modelo detalhado é ajustado de modo a se determinar a localização e a escala. Os pontos-chaves são selecionados baseando-se em medidas de estabilidade.

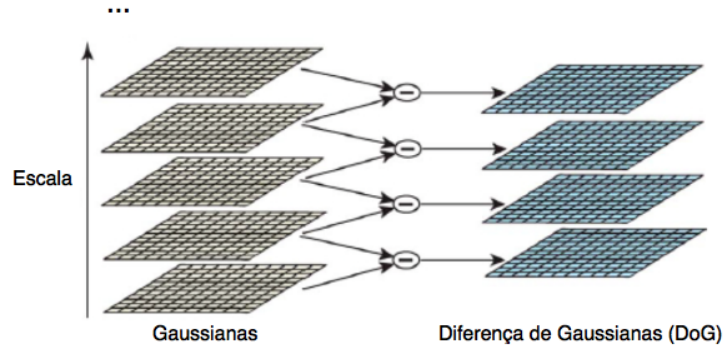


Figura 10 – Criação de uma imagem resultante da diferença de Gaussianas (DoG) (LOWE, 2004).

3. **Atribuição de orientação.** Uma ou mais orientações são definidas para cada localização de *keypoint* através dos gradientes locais da imagem. Todas as operações futuras serão realizadas sobre os dados da imagem transformados em relação à orientação, à escala e à localização de cada característica. Desta forma obtém-se invariância a estas transformações. A magnitude do gradiente $m(x, y)$ e a orientação $\theta(x, y)$ são computados usando a Equação 3.33.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \arctan \left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right). \quad (3.33)$$

4. **Descritor dos *keypoints*.** Os gradientes locais da imagem são medidos na região vizinha de cada *keypoint* na escala selecionada. Essas medidas são então transformadas para uma representação que permite níveis significativos de distorção e mudança de iluminação. Se uma orientação do *keypoint* é selecionada, o descritor de característica é computado como um conjunto de histogramas de orientação em uma vizinhança 4×4 , como mostra a Figura 11. Para facilitar a visualização, são ilustradas somente 2×2 subregiões.

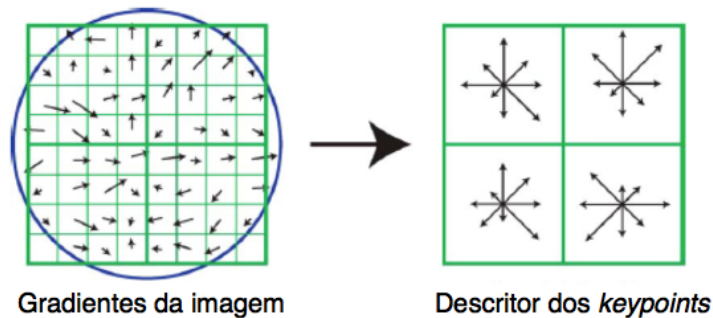


Figura 11 – Descritor das características do SIFT (LOWE, 2004).

O histograma possui 8 faixas, assim um vetor de características SIFT tem $4 \times 4 \times 8 = 128$ elementos.

Uma outra etapa importante é a correspondência entre características (*feature matching*) cujo objetivo é encontrar correspondências entre dois conjuntos de descritores. A estratégia de busca pode ser exaustiva ou restrita. Ao calcular a pontuação de dissimilaridade entre cada par de descritores, uma matriz de pontuação é construída. Para cada ponto de característica no quadro $k - 1$, a melhor correspondência na quadro k é escolhido. Finalmente, com base na estratégia de correspondência, alguns dos “melhores” correspondências podem ser descartadas devido à ambiguidade ou consistência.

3.5.2 SURF

O descritor chamado de Características Robustas Aceleradas (*Speeded Up Robust Features*, SURF), proposto por (BAY et al., 2008), é um método rápido e robusto inspirado no SIFT para descrever pontos-chave em imagens. Sua baixa complexidade computacional é alcançada através das imagens integrais introduzida por (VIOLA; JONES, 2004).

No SURF, a detecção dos pontos de interesse, ou pontos-chave, é realizada através da aproximação da matriz Hessiana. O conceito da imagem integral é utilizado com o seguinte propósito: a inserção de uma imagem integral $\mathbf{I}_\Sigma(x)$ em uma localização $x = (x, y)^T$ representa a soma de todos os *pixels* na imagem de entrada $\mathbf{I}$ com uma região retangular formada pela origem e x como mostra a Figura 12.

$$\mathbf{I}_\Sigma(x) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} (\mathbf{i}, \mathbf{j})$$

Dada um ponto $x = (x, y)$ em uma imagem $\mathbf{I}$, a matriz Hessiana $\mathbf{H}(x, \sigma)$ na posição x e escala σ é definida na Equação 3.34.

$$\mathbf{H} = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad (3.34)$$

onde $L_{xx}(x, \sigma)$ é a convolução da derivada de segunda ordem Gaussiana $\frac{\partial^2}{\partial x^2} g(\sigma)$ na imagem $\mathbf{I}$ e no ponto x .

A matriz Hessiana, no SURF, é aproximada utilizando filtros espaciais com máscara 9×9 . A Figura 13 mostra um exemplo dessas aproximações. Respostas semelhantes são computadas através da Equação 3.35.

$$\det(\mathbf{H}_{\text{aproximada}}) = D_{xx}D_{yy} - (wD_{xy})^2, \quad (3.35)$$

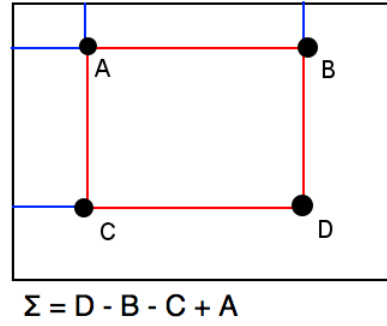


Figura 12 – Com imagens integrais, apenas quatro operações são necessárias para calcular a soma de intensidades em uma área retangular na imagem (BAY et al., 2008).

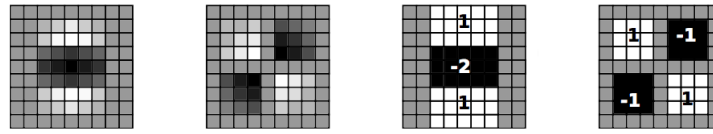


Figura 13 – Da esquerda para direita: a derivadas Gaussianas de segunda ordem discretizadas L_{yy} e L_{xy} e aproximações do SURF para L_{yy} e L_{xy} (EVANS, 2009).

onde D_{xx} , D_{yy} e D_{xy} denotam aproximações das derivadas Gaussianas parciais de segunda ordem nas direções x , y e xy . w é um peso usado para balancear a expressão. Isso é importante para a conservação da energia entre os *kernels* Gaussianos e as aproximações dos *kernels* Gaussianos, no geral $w \approx 0,912 \dots \cong 0,9$ (BAY et al., 2008).

A representação do espaço-escala é necessária para detectar características em vários tamanhos. Assim como em DoG, o espaço-escala é dividido em oitavas. Uma oitava é um conjunto de imagens filtradas com o aumento do tamanho do *kernel*. Em cada oitava, o filtro é escalado pelo fator de 2. O cálculo rápido do filtro permite que o SURF aumente os filtros no lugar de diminuir a escala da imagem.

Finalmente, as características locais são selecionadas como máximos locais na vizinhança $3 \times 3 \times 3$ no espaço-escala. Um método rápido para a supressão não máxima, proposto por Neubeck e Gool (2006), é utilizado para localizar esses pontos extremos.

3.6 Conclusão

Esse capítulo discute o conceito de extração de características em imagens além de explicar o funcionamento de métodos estado da arte adotados na dissertação, sendo abordado o uso dos momentos, extratores baseados em textura e descritores de pontos-chave.

4 Reconhecimento de Padrões

Neste capítulo, são abordados aspectos relacionados ao reconhecimento de padrões. O classificador Bayesiano é descrito na Seção 4.1. Em seguida, os classificadores que se baseiam na Teoria de Aprendizado Estatístico, SVM e LSSVM, são discutidos na Seção 4.2 e 4.3, respectivamente. A rede neural artificial MLP é apresentada na Seção 4.4. Finalmente, os classificadores baseados em distância, MLM e MLM-NN, são discutidos nas Seções 4.5 e 4.6, respectivamente.

Segundo Bishop (2006), o reconhecimento de padrões tem sua origem na engenharia, enquanto o aprendizado de máquina cresceu na Ciência da Computação. No entanto, essas atividades podem ser vistas como duas facetas do mesmo campo. Para Theodoridis e Koutroumbas (2008), o reconhecimento de padrões é uma área de estudo cujo objetivo principal é a classificação de padrões dentro de um número de categorias ou classes. Dependendo da aplicação, esses padrões podem ser imagens, formas de onda de sinal ou algum tipo de medição que precisa ser classificada. O projeto de um sistema de reconhecimento de padrões envolve, essencialmente, três etapas:

- aquisição de dados (extração de características) e pré-processamento de dados (seleção de características mais discriminativas);
- representação de dados;
- tomada de decisão (construção de um classificador).

O tipo de domínio é determinante para a escolha da técnica de pré-processamento, do esquema de representação e do modelo de tomada de decisão. Um problema bem definido e suficientemente detalhado é caracterizado por pequenas variações intra-classes (pequenas variações para padrões da mesma classe) e grandes variações inter-classes (grandes variações entre padrões de classes distintas). Por consequência conduz a uma representação compacta das amostras e simplifica a tomada de decisão.

A representação de uma amostra, padrão ou vetor $\mathbf{x} = (x_1, x_2, x_3, \dots, x_D)^\top$ é dado por um conjunto de características ou atributos no espaço D -dimensional. Um dos problemas fundamentais no reconhecimento de padrões é o de proporcionar uma resposta razoável para uma determinada entrada, daí então considerar o vetor $\mathbf{x}$ pertencente a uma classe Y_i , tal que $Y_1, Y_2, \dots, Y_K$ baseado nos valores de seus D atributos. Essa tarefa é denominada classificação. No entanto, é importante ressaltar que o reconhecimento de padrões é um problema mais geral que engloba outros tipos de saída, tal como regressão (BISHOP, 2006).

O reconhecimento de padrões pode ser dividido em: aprendizagem supervisionada e aprendizagem não supervisionada. Na primeira, os padrões de treinamento apresentam-se com seus rótulos (*labels*) que indicam a qual classe pertencem. Na aprendizagem não supervisionada, por outro lado, não se usa ou não há rótulos para o treinamento e o número de classes, muitas vezes, deve ser também encontrado. Uma terceira categoria de reconhecimento de padrões é a aprendizagem semisupervisionada, que consiste na combinação das outras duas. Nessa, a aprendizagem é baseada em dados rotulados e não rotulados. Normalmente, esse método é usado quando a tarefa de se obter dados de treinamento rotulados é demorada (DUDA; HART; STORK, 2001; BISHOP, 2006).

4.1 Classificador Bayesiano

Dado o problema de estimar uma classe desconhecida $y \in \{Y_1, Y_2, \dots, Y_K\}$ de um vetor de características $\mathbf{x} = (x_1, \dots, x_D) \in \mathbb{R}^D$, o classificador Bayesiano aprende o modelo da probabilidade conjunta $P(\mathbf{x}, y)$ dos dados de treinamento e prediz a classe y mais provável através da regra de Bayes calculando $P(y|\mathbf{x})$. Mais precisamente, aplicando a regra de Bayes, a probabilidade *a posteriori* $P(y = Y_k|\mathbf{x})$ para a classe Y_k pode ser definida a seguir (THEODORIDIS; KOUTROUMBAS, 2008; DUDA; HART; STORK, 2001; BISHOP, 2006).

$$P(y = Y_k|\mathbf{x}) = \frac{P(\mathbf{x}|y = Y_k)P(y = Y_k)}{\sum_{j=1}^K P(\mathbf{x}|y = Y_j)P(y = Y_j)} \quad (4.1)$$

em que Y_k denota o k -ésimo valor possível para y e o somatório no denominador é sobre todos os valores de y . A distribuição $P(y)$ fornece a probabilidade *a posteriori* de y e $P(\mathbf{x}|y)$ modela a função de densidade de probabilidade condicional da classe ou a probabilidade de $\mathbf{x}$ dado y . Ambas as distribuições podem ser estimadas utilizando os dados de treinamento. Essas estimativas são utilizadas na regra de Bayes para determinar $P(y|\mathbf{x}_{novo})$ para uma nova amostra sem classe $\mathbf{x}_{novo}$. O classificador Bayesiano gera uma função que atribui uma classe $\hat{y}$ à amostra $\mathbf{x}_{novo}$ com base na regra de classificação apresentada na Equação 4.1 conhecida como máximo *a posteriori* (*maximum a posteriori*, MAP).

$$\hat{y} = \arg \max_{Y_k} P(y = Y_k|\mathbf{x}), \quad k \in \{1, \dots, K\}. \quad (4.2)$$

Essa regra de decisão pode ser simplificada, basta que o denominador da Equação 4.1 seja desconsiderado, uma vez que, para uma dada amostra $\mathbf{x}_{novo}$, esse termo é constante. Em outras palavras, a regra de decisão escolhe a classe de uma nova amostra através da probabilidade *a posteriori* máxima. A versão logarítmica do MAP é apresentado na

Equação 4.3.

$$\hat{y} = \arg \max_{Y_k} \{\log P(y = Y_k) + \log P(\mathbf{x} = \mathbf{x}_{novo} | y = Y_k)\}, k \in \{1, \dots, K\} \quad (4.3)$$

Uma das funções de densidade de probabilidade mais comumente adotadas é a Gaussiana ou Normal. A generalização multivariada de uma distribuição Gaussiana no espaço D -dimensional para a classe Y_k é apresentada a seguir.

$$P(\mathbf{x} | y = Y_k) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) \right\} \quad (4.4)$$

em que $\boldsymbol{\mu}_k$ é o vetor de médias e Σ_k a matriz de covariância de tamanho $D \times D$. Logo, $|\Sigma_k|$ e Σ_k^{-1} são o determinante e a inversa da matriz de covariância, respectivamente.

4.2 Máquinas de Vetores Suporte

As Máquinas de Vetores de Suporte (*Support Vector Machines*, SVMs) são embasadas na Teoria de Aprendizado Estatístico, desenvolvida por Vapnik (1998). Essa teoria estabelece uma série de princípios que devem ser seguidos na obtenção de classificadores com boa generalização. Enquanto técnicas tradicionais para reconhecimento de padrões são baseadas na minimização do risco empírico— isto é, tenta-se otimizar o desempenho sobre o conjunto de treinamento —, SVMs minimizam o risco estrutural, o qual tem relação com a minimização do erro nos padrões de teste (ROCHA NETO, 2011).

A ideia básica de uma SVM é separar as classes com superfícies que maximizem a margem entre as mesmas, ou seja, encontrar um hiperplano de separação ótima. Os padrões de entrada são convertidos para um vetor de características de alta dimensionalidade, cujo objetivo é separar linearmente os padrões no espaço. Uma vez que, o espaço adequado de características é definido, a SVM seleciona o hiperplano ótimo, o qual corresponde à maior distância de seus padrões no conjunto de treinamento. Esses padrões são chamados de Vetores de Suporte (*Support Vector*) (ROCHA NETO, 2011).

Embora, a SVM tenha sido, originalmente, formulado para resolver problemas binários, há algumas formulações do SVM para problemas multiclasse (CRAMMER et al., 2001). No entanto, devido sua complexidade, outras abordagens como um contra um (DUAN; KEERTHI, 2005), um contra todos (DUAN; KEERTHI, 2005) são algumas mais utilizadas.

4.2.1 Problema Primal

A tarefa de classificação em problemas linearmente separáveis consiste em encontrar um hiperplano que divida corretamente as amostras de cada classe. Esse tipo de problema possui diversas soluções, como mostra a Figura 14. Mais formalmente, as soluções para esse tipo de problema podem ser apresentadas através da equação de um hiperplano,

$$\mathbf{w}^T \mathbf{x} + b = 0, \quad (4.5)$$

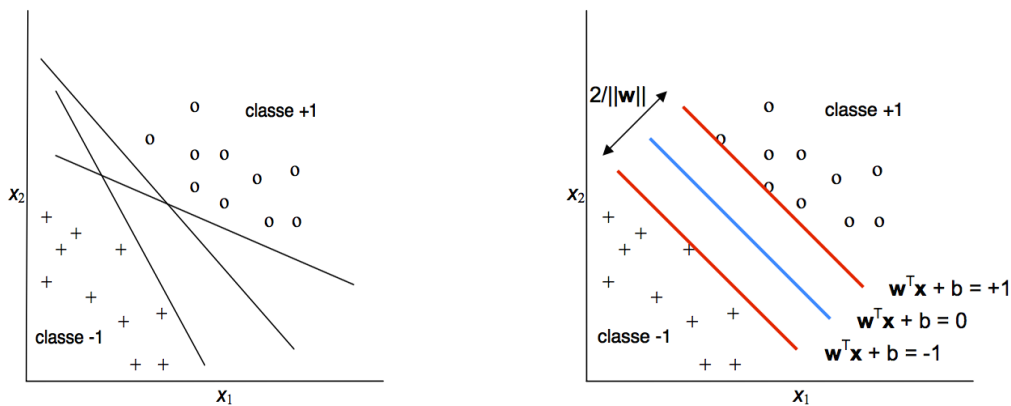
em que $\mathbf{w}$ é o vetor de pesos e b é o viés (*bias*). A categoria de hiperplanos que separam as classes devem satisfazer as seguintes regras:

$$\begin{cases} \mathbf{w}^T \mathbf{x}_i + b \geq a & \text{se } y_i = +1, \\ \mathbf{w}^T \mathbf{x}_i + b \leq -a & \text{se } y_i = -1. \end{cases} \quad (4.6)$$

em que $a > 0$ e $i = 1, \dots, N$. Assim, as restrições são satisfeitas para todos os padrões do conjunto de treinamento.

Dentre os hiperplanos possíveis, é desejado encontrar o ótimo que maximize a margem de separação entre o hiperplano ótimo e o padrão de treinamento mais próximo, os vetores de suporte. O hiperplano e a margem de separação ótima são apresentados na Figura 14. A margem de separação ótima é dada por $\rho = 2/\|\mathbf{w}\|$ e o hiperplano ótimo é representado por

$$\mathbf{w}_o^T \mathbf{x} + b_o = 0. \quad (4.7)$$



(a) Diversos hiperplanos capazes de resolver o problema de classificação binária. (b) Margem (vermelho) e hiperplano (azul) de separação ótima.

Figura 14 – Problema de classificação binária.

Em problemas de classificação de padrões, classificadores SVM utilizam a função sinal, ou seja

$$f(\mathbf{x}) = \text{sign}(\mathbf{w}_o^T \mathbf{x} + b_o). \quad (4.8)$$

A formulação do problema de obtenção do hiperplano ótimo pode começar a ser realizada com a maximização da margem de separação ρ ou, de forma equivalente, com a minimização da norma $\|\mathbf{w}\|$ dos vetores de pesos. Portanto, o hiperplano que separa os dados de entrada pode ser descrito como um que minimize

$$\tau(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w}, \quad (4.9)$$

satisfazendo as restrições

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq +1, \forall i. \quad (4.10)$$

Embora muitos problemas reais não sejam linearmente separáveis, utilizar essa suposição inicial é importante para apresentar de forma mais simples a ideia que fundamenta a teoria das SVMs. Posteriormente, a restrição relacionada à linearidade do problema é removida. Quando o problema de otimização é formulado com base nesta imposição, o classificador resultante é denominado SVM com margem rígida.

Nem sempre há um hiperplano que separe sem erros todos os padrões pertencentes às duas classes. Especialmente, quando existe sobreposição entre os dados que compõem as classes. O sobre-ajustamento da superfície de decisão aos dados é outro motivo que deve ser levado em consideração para adotar uma flexibilidade na margem. Diante disso, foi formulado um relaxamento das restrições dos classificadores SVM com margens rígidas com base em um limiar, ou seja

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, N, \quad (4.11)$$

em que $\xi_i \geq 0, i = 1, \dots, N$. Os limiares ξ_i são chamados de variáveis de folga (*slack variables*).

O problema de otimização, agora, é dado por

$$\min \tau(\mathbf{w}, \xi) = \min \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i \right\}, \quad (4.12)$$

sujeito a

$$\begin{cases} y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i & , i = 1, \dots, N \\ \xi_i \geq 0 & , i = 1, \dots, N. \end{cases} \quad (4.13)$$

em que a constante C é um parâmetro que faz uma regularização entre o primeiro ($\mathbf{w}^\top \mathbf{w}$) e o segundo termo ($C \sum_{i=1}^N \xi_i$) da função-custo.

4.2.2 Formulação Dual

A solução do problema de otimização (primal) é determinada pelo ponto de sela da função lagrangeana $L(\mathbf{w}, b, \xi, \alpha, \beta)$, a qual deve ser minimizada em relação a $\mathbf{w}, b$ e ξ e maximizada em relação a α e β , sendo definida da seguinte maneira

$$L(\mathbf{w}, b, \xi, \alpha, \beta) = \frac{1}{2} \mathbf{w}^\top \mathbf{w} + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i (y_i (\mathbf{x}_i^\top \mathbf{w} + b) - 1 + \xi_i) - \sum_{i=1}^N \beta_i \xi_i, \quad (4.14)$$

em que todos os elementos dos conjuntos $\alpha = \{\alpha_i\}_{i=1}^N, \beta = \{\beta_i\}_{i=1}^N$ e $\xi = \{\xi_i\}_{i=1}^N$ possuem valores não-negativos.

O problema de otimização dual para o classificador SVM com margem flexível é dado por

$$\max L(\alpha) = \max \left\{ \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j \right\}, \quad (4.15)$$

sujeito a

$$\begin{cases} \sum_{i=1}^N \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C \end{cases} \quad \text{para } i = 1, \dots, N \quad (4.16)$$

4.2.3 O Truque do *Kernel*

Mesmo com a margem flexível, o SVM sofre com algumas situações em que o problema não é resolvido de forma satisfatória. Um exemplo deste tipo de problema é ilustrado na Figura 15, que traz um problema de classificação binário não-linearmente separável.

O truque do *kernel* (*kernel trick*) possibilita que o espaço original seja mapeado em um espaço de alta dimensão chamado espaço de características, onde os dados podem ser linearmente separáveis. Na Figura 16 é possível visualizar um exemplo de mapeamento de um espaço de baixa dimensão (bidimensional) para um espaço de dimensão mais alto

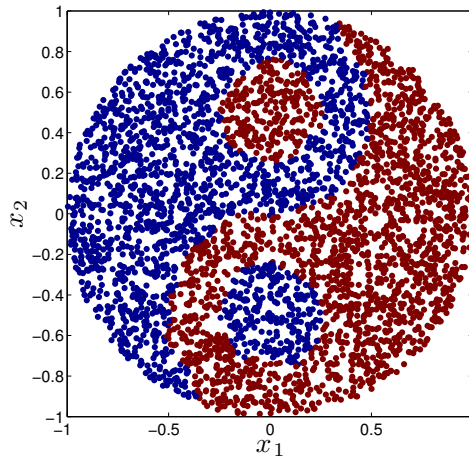


Figura 15 – Exemplo de um problema que não pode ser resolvido de forma satisfatória por um classificador linear (SOUZA JÚNIOR et al., 2015).

(tridimensional). No espaço de dimensão mais alta é possível separar os dados com uma superfície linear. Na prática são utilizados espaços de características com dimensões muito altas.

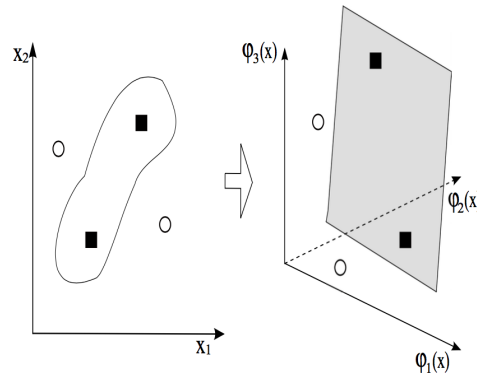


Figura 16 – Mapeamento de dados para um espaço de características de mais alta dimensão através do truque do *kernel* (ROCHA NETO, 2011).

Utilizando o truque do *kernel*, a classe para um dado vetor $\mathbf{x}$ pode ser determinada a partir da seguinte equação,

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^N \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b\right). \quad (4.17)$$

em que $K(\mathbf{x}, \mathbf{x}_i) = \varphi(\mathbf{x}^\top)\varphi(\mathbf{x}_i)$ é uma função de *kernel* que satisfaz o teorema de Mercer (MERCER, 1909). Os *kernels* mais comumente utilizados são apresentados na Tabela 3.

Tabela 3 – *Kernels* comumente utilizados.

<i>Kernel</i>	Descrição
Linear	$K(\mathbf{x}, \mathbf{x}_i) = \mathbf{x}_i^\top \cdot \mathbf{x}$
Polinomial	$K(\mathbf{x}, \mathbf{x}_i) = (\mathbf{x}_i^\top \cdot \mathbf{x} + 1)^d$, em que d é o grau do polinômio.
Gaussiano (RBF)	$K(\mathbf{x}, \mathbf{x}_i) = \exp\{-\ \mathbf{x} - \mathbf{x}_i\ ^2/\sigma^2\}$, em que σ é uma constante.

4.3 Máquinas de Vetores de Suporte por Mínimos Quadrados

Suykens, Gestel e Brabanter (2002) propuseram uma modificação do SVM chamado Máquinas de Vetores de Suporte por Mínimos Quadrados (*Least Squares Support Vector Machines*, LSSVMs). O problema de otimização da LSSVM difere do classificador SVM em dois aspectos. A primeira mudança está na substituição da restrições de desigualdade por restrições de igualdade. A outra mudança refere-se à incorporação na função custo da soma das variáveis de folga ao quadrado ponderada por uma constante de regularização γ . Nesse caso, ξ_i é trocado por ξ_i^2 . Esta constante tem uma interpretação semelhante à constante C exposta no problema do classificador SVM (ROCHA NETO, 2011). Dessa forma, o problema de otimização pode ser resolvido de uma maneira mais simples, a partir de um sistema de equações lineares. A função de otimização do classificador LSSVM torna-se então,

$$\min \tau(\mathbf{w}, \xi) = \frac{1}{2} \mathbf{w}^\top \mathbf{w} + \gamma \frac{1}{2} \sum_{i=1}^N \xi_i^2, \quad (4.18)$$

sujeito a

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) = 1 - \xi_i, i = 1, \dots, N. \quad (4.19)$$

em que $\tau(\mathbf{w}, \xi)$ representa a função custo que deve ser minimizada. Nessa formulação, as variáveis de folga podem assumir valores negativos, diferente do que ocorre no SVM.

A simplificação do processo de treinamento da LSSVM é uma de suas vantagens. Porém, uma das limitações provenientes dessa formulação refere-se à perda da natureza esparsa na solução do problema. Como os multiplicadores de Lagrange tornam-se não-nulos, após o processo de aprendizagem, é necessário armazenar todos os padrões de treinamento, bem como os multiplicadores de Lagrange associados (ROCHA NETO, 2011).

Seguindo o raciocínio semelhante ao das SVMs juntamente com a simplificação citadas anteriormente, pode-se formular uma sistema de equações lineares, $\mathbf{Ax} = \mathbf{B}$, com propósito de representar o problema dos classificadores LSSVM,

$$\begin{bmatrix} 0 & \mathbf{y}^\top \\ \mathbf{y} & \Omega + \gamma^{-1} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix} \quad (4.20)$$

em que $\Omega_{i,j} = y_i y_j \mathbf{x}_i^T \mathbf{x}_j$, $y = [y_1 y_2 \dots y_N]^T$, $\alpha = [\alpha_1 \alpha_2 \dots \alpha_N]^T$ e $\mathbf{1} = [1 1 \dots]^T$, tal que o tamanho do vetor $\mathbf{1}$ é igual a N . Ao utilizar o truque do *kernel*, redefini-se $\Omega_{i,j}$ para $\Omega_{i,j} = y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$. Dessa forma, a saída do classificador LSSVM é computada por

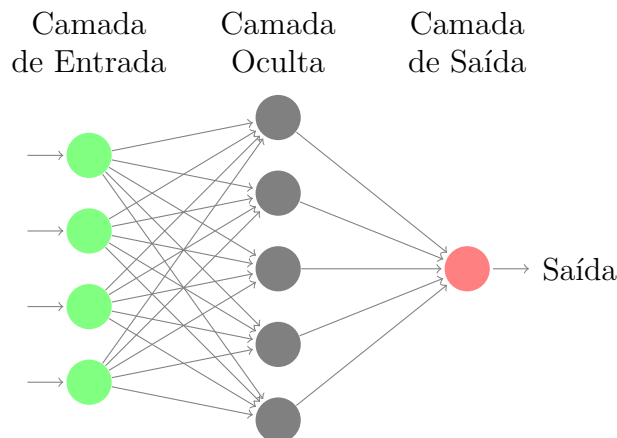
$$f(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^L \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b \right) \quad (4.21)$$

4.4 Perceptron de Múltiplas Camadas

O Perceptron de Múltiplas Camadas (*Multilayer Perceptron*, MLP) utiliza o paradigma supervisionado de aprendizagem, caracterizado pela existência de um supervisor externo, cujo papel central é fornecer as entradas à rede e comparar as saídas geradas com um resultado esperado, para posterior ajuste de parâmetros, a partir dos erros gerados. A minimização do erro decorre de maneira incremental até a convergência, definida por um critério de parada pré-estabelecido. Assim, o vetor contendo os dados a serem processados é apresentado à rede a partir da camada de entrada, gerando estímulos que são propagados adiante nas demais camadas. As saídas de cada neurônio das camadas escondidas comporão as entradas da camada subsequente. Desta forma, por cada conexão entre neurônios, transcorrem estímulos que são ponderados por seus respectivos pesos, determinando, a partir de uma função de ativação, os estímulos que chegam aos neurônios conectados a sua saída. Seguindo este procedimento, tem-se então uma resposta final na camada de saída para o vetor como entrada.

As redes MLP adotam múltiplas camadas, alimentadas para frente (*feedforward*), sendo a primeira a camada de entrada, esta seguida pelas camadas escondidas e, por fim, uma camada de saída. A Figura 17 ilustra um modelo MLP com uma camada escondida. A arquitetura da rede pode ser melhor descrita pelo número de camadas, a função de ativação e o número de neurônios por camada.

Figura 17 – Modelo esquemático de uma rede MLP com uma camada oculta.



A escolha do número de camadas é dada pelo problema em questão e em cada camada ocorre uma fração do processamento como um todo, onde o espaço é dividido em regiões de decisão. Desta forma, a rede é capaz de trabalhar com problemas que não sejam linearmente separáveis, desde que haja neurônios não-lineares nas camadas escondidas. Papel importante cabe às camadas intermediárias, que criam uma codificação interna para os padrões apresentados na entrada. Um número suficientemente grande de camadas é possível formar representações para qualquer conjunto de entrada (BRAGA, 2007). Entretanto a grande maioria dos problemas raramente necessita de mais de uma camada intermediária.

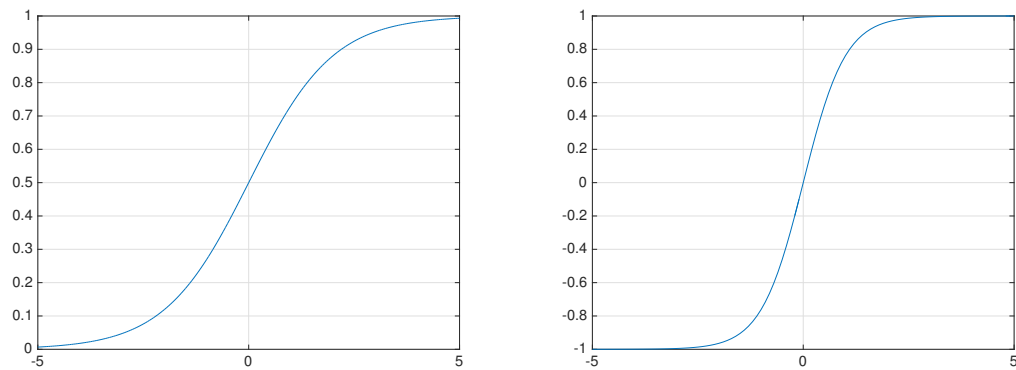
Não há uma regra geral bem explicitada na literatura sobre a definição do número de neurônios para a resolução de um determinado problema. Esta modelagem deverá ser definida em função da complexidade do problema, de maneira que quanto maior a complexidade, um maior número de neurônios é exigido, atentando sempre para que a rede não perca sua capacidade de generalização caso um número excessivo de neurônios escondidos seja utilizado. Um ponto relevante neste caso é que a complexidade do problema a ser tratado não é conhecido de antemão, e o modelo a ser utilizado deve ter complexidade proporcional ao problema.

A rede emprega em cada unidade de processamento uma função de ativação. A função de ativação utilizada nas camadas intermediárias deve ser não linear, para que seja incorporado o tratamento de não-linearidade à rede. Caso as funções sejam puramente lineares, a MLP terá funcionamento equivalente a uma única camada, visto que sucessivas transformações lineares equivalem a uma única transformação linear (BRAGA, 2007). As funções sigmóides são contínuas, diferenciáveis em qualquer ponto, monotônicas estritamente crescentes, possuindo certa similaridade com funções lineares, o que as tornam ideais para a rede MLP. As funções de ativação mais utilizadas no modelo MLP são funções sigmoidais, tais como a função logística e a função tangente hiperbólica, que devem estar presentes em pelo menos uma das camadas escondidas. A diferença entre essas duas funções é quanto ao intervalo de suas saídas, $[0, 1]$ na primeira e $[-1, 1]$ na segunda. As funções logística e a tangente hiperbólica são dados, respectivamente, por

$$\phi(x) = \frac{1}{1 + \exp(-x)}, \quad e \quad (4.22)$$

$$\phi(x) = \tanh(x) = \frac{1 + \exp(-x)}{1 - \exp(-x)}. \quad (4.23)$$

Enquanto, os gráficos para tais funções são apresentados na Figura 18.



(a) Função logística.

(b) Função tangente hiperbólica.

Figura 18 – Gráficos das funções logística (a) e tangente hiperbólica (b).

4.4.1 Algoritmos de Treinamento

O treinamento é o processo de determinar os valores ótimos dos pesos para uma RNA. Isso é feito definindo a função de erro total entre a saída da rede e a desejada. O algoritmo de treinamento mais conhecido é o *Backpropagation*. Além dele, é apresentado o algoritmo de Levenberg-Marquardt, adotado no trabalho.

4.4.1.1 Algoritmo de Retropropagação do Erro

O algoritmo de retropropagação do erro, conhecido como *Backpropagation* (BP) é um dos métodos mais utilizados para treinamentos de redes neurais artificiais. Ele consegue propagar a informação de erro da camada de saída para as camadas escondidas, podendo ser visto como uma generalização do filtro adaptativo do Mínimo Quadrático Médio (*Least Mean Square*, LMS).

Basicamente, o algoritmo de retropropagação do erro consiste em duas etapas de processamento: computação para a frente (propagação) e computação para trás (retropropagação) (HAYKIN; ENGEL, 2007). O primeiro é equivalente ao *feedforward*, em que as entradas são processadas pelas camadas em sequência, e a saída é obtida na última camada. A saída obtida é então subtraída da saída esperada, gerando a informação de erro. A informação de erro é então propagada na direção contrária, ajustando os pesos no processo, vindo daí o nome *backpropagation*. Essa propagação do erro é feita usando a generalização da regra delta, que usa derivadas parciais para calcular os ajustes nos valores dos pesos de camada. Por isso, as funções de ativação dos neurônios devem ser sempre deriváveis (ROJAS, 1996).

4.4.1.2 Levenberg-Marquardt

Embora o algoritmo retropropagação do erro com gradiente descendente seja o mais amplamente usado, ele sofre de algumas limitações: entre elas pode ser destacada

sua velocidade de convergência e os problemas com os mínimos locais. Para melhorar o desempenho do processo de treinamento da rede, foram propostos outros métodos de treinamento, e um dos mais conhecidos é o algoritmo Levenberg-Marquardt (LM). Ele foi desenvolvido de forma independente por Kenneth Levenberg e Donald Marquardt e fornece uma solução numérica para o problema de minimizar a função não linear com convergência rápida e estável. É difícil saber qual algoritmo de aprendizado é o mais adequado para um determinado problema devido vários fatores, tais como: tamanho da rede, objetivo do treinamento e complexidade computacional.

O algoritmo LM é uma combinação do gradiente descendente e do algoritmo de Gauss-Newton (GN). Ele supera a lenta convergência do método gradiente descendente e corrige a única aproximação quadrática da função de erro do algoritmo de GN (WILAMOWSKI; IRWIN, 2011). O método gradiente descendente foi introduzida primeiro, enquanto algoritmo de GN é baseado no método de Newton pressupondo que todos os componentes do gradiente na regra delta são funções de pesos e todo os pesos são linearmente independentes. Assim, o vetor de gradiente pode ser expresso como o algoritmo de segunda ordem, onde a matriz Hessiana $\mathbf{H}$ e a matriz Jacobiana $\mathbf{J}$ são utilizadas para simplificação. A regra de atualização do algoritmo de GN é,

$$\mathbf{w}_{i+1} = \mathbf{w}_i - (\mathbf{J}_i^T \mathbf{J}_i)^{-1} \mathbf{J}_i \mathbf{e}_i, \quad (4.24)$$

em que $\mathbf{e}$ é o vetor de erros. O algoritmo LM torna a matriz Hessiana $\mathbf{J}^T \mathbf{J}$ inversível através da aproximação $\mathbf{H} \approx \mathbf{J}^T \mathbf{J} + \mu \mathbf{I}$, em que μ é o coeficiente de combinação. Se μ é muito grande, o algoritmo LM é igual ao método gradiente descendente, por outro lado, se μ é muito pequeno, o algoritmo é próximo do GN.

4.5 Máquina de Aprendizagem Mínima

A Máquina de Aprendizagem Mínima (*Minimal Learning Machine*, MLM), proposto por Souza JÚnior et al. (2015), é um método recentemente proposto para aprendizagem supervisionada. A ideia básica por trás do MLM é a existência de um mapeamento entre configurações geométricas dos pontos no espaço de entrada e saída. Tal mapeamento pode ser reconstruído pela aprendizagem de um modelo de regressão linear multireposta entre distâncias. Sob estas condições, para um ponto de entrada com configuração conhecida no espaço de entrada, sua configuração correspondente no espaço de saída pode ser estimada depois de aprender um modelo linear simples entre matrizes de distância da entrada e saída. O resultado estimado é então usado para localizar o ponto de saída e, assim, fornecer uma estimativa para resposta. Na sua formulação básica, o MLM se assemelha a um método não-supervisionado clássico de redução de dimensionalidade, a principal diferença é que a configuração de saída no MLM é conhecida previamente.

4.5.1 Formulação

Dado um conjunto de N pontos de entrada $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^N$, com $\mathbf{x}_i \in \mathbb{R}^D$, e o conjunto de saídas correspondes $\mathcal{Y} = \{\mathbf{y}_i\}_{i=1}^N$, com $\mathbf{y}_i \in \mathbb{R}^S$. Assumindo a existência de um mapeamento $f: \mathcal{X} \rightarrow \mathcal{Y}$ entre o espaço de entrada e saída, queremos estimar f a partir dos dados com o modelo

$$\mathbf{Y} = f(\mathbf{X}) + \mathbf{R}. \quad (4.25)$$

As colunas da matriz $\mathbf{X}$ correspondem às D entradas e as linhas às N observações. Igualmente, as colunas da matriz $\mathbf{Y}$ correspondem às S saídas, e as linhas às N observações. As colunas da matriz $\mathbf{R}$ correspondem aos vetores de resíduo.

Dado que o espaço de entrada $\mathcal{X}$ é bem amostrado e f é suave, espera-se que para cada par de pontos de entrada $(\mathbf{x}_i, \mathbf{x}_j)$ e para todo $\epsilon_y > 0$, existe um $\epsilon_x > 0$ tal que para $d(\mathbf{x}_i, \mathbf{x}_j) < \epsilon_x$ vale $\delta(f(\mathbf{x}_i)), \delta(f(\mathbf{x}_j)) < \epsilon_y$ onde $d(\cdot, \cdot)$ e $\delta(\cdot, \cdot)$ são funções de distância em $\mathcal{X}$ e $\mathcal{Y}$, respectivamente. Sob esta condição, precisa-se reconstruir o mapeamento entre as matrizes de distância de entrada $\mathbf{D}_x$ e as correspondentes matrizes de distância de saída Δ_y .

4.5.1.1 Regressão entre Distâncias

Para uma seleção de pontos de entrada de referência $\mathcal{R} = \{\mathbf{m}_k\}_{k=1}^K$ com $\mathcal{R} \subseteq \mathcal{X}$ e correspondentes saídas $\mathcal{T} = \{\mathbf{t}_k\}_{k=1}^K$ com $\mathcal{T} \subseteq \mathcal{Y}$, é definido $\mathbf{D}_x \in \mathbb{R}^{N \times K}$ de tal forma que sua k -ésima coluna contém as distâncias $d(\mathbf{x}_i, \mathbf{m}_k)$ entre os $\mathbf{x}_i$, $i = 1, \dots, N$ pontos de entrada e o k -ésimo ponto de referência $\mathbf{m}_k$. Analogamente, é definido $\Delta_y \in \mathbb{R}^{N \times K}$ de forma que sua k -ésima coluna contém as distâncias $\delta(\mathbf{y}_i, \mathbf{t}_k)$ entre os pontos de saída $\mathbf{y}_i$ e a saída $\mathbf{t}_k$ do k -ésimo ponto de referência. O modelo de regressão multiresposta associado para estimar g é portanto

$$\Delta_y = g(\mathbf{D}_x) + \mathbf{E}, \quad (4.26)$$

onde as colunas da matriz $\mathbf{D}_x$ correspondem aos K vetores de entrada e as colunas da matriz Δ_y correspondem aos K vetores de resposta. As colunas da matriz $\mathbf{E}$ correspondem aos K resíduos.

Assume-se que o mapeamento g entre distâncias de entrada e saída é linear, então o modelo de regressão multiresposta entre as matrizes de distância torna-se

$$\Delta_y = \mathbf{D}_x \mathbf{B} + \mathbf{E}. \quad (4.27)$$

As colunas da matriz de regressão $\mathbf{B}$ de dimensão $K \times K$ correspondem aos coeficientes para as K respostas. A matriz $\mathbf{B}$ pode ser estimada a partir dos dados através da minimização da soma dos quadrados dos resíduos com a função de custo:

$$\text{RBS}(\mathbf{B}) = \sum_{k=1}^K \sum_{i=1}^N (\delta(\mathbf{y}_i, \mathbf{t}_k) - h_k(d(\mathbf{x}_i, \mathbf{m}_k)))^2 \quad (4.28a)$$

$$= \text{tr}((\mathbf{\Delta}_y - \mathbf{\Delta}_x \mathbf{B})^\top (\mathbf{\Delta}_y - \mathbf{D}_x \mathbf{B})) \quad (4.28b)$$

Sob as condições normais onde o número de equações em (4.27) é maior do que o número de incógnitas, o problema é sobredeterminado e, geralmente, sem solução. Isto corresponde ao caso em que o número de pontos de referência selecionada é menor que o número de pontos disponível para resolver o modelo (*e.g.*, $K < N$) sendo necessário contar com a solução fornecida pela estimativa dos mínimos quadrados:

$$\hat{\mathbf{B}} = (\mathbf{D}_x^\top \mathbf{D}_x)^{-1} \mathbf{D}_x^\top \mathbf{\Delta}_y. \quad (4.29)$$

Por outro lado, se na Equação 4.27 o número de equações é igual ao número de incógnitas (i.e., todos os pontos de aprendizagem são também selecionados como ponto de referência e $K = N$), o problema é unicamente determinado e, normalmente, com uma única solução $\hat{\mathbf{B}} = (\mathbf{D}_x)^{-1} \mathbf{\Delta}_y$. Claramente, menos interessante é o caso onde em (4.27) o número de equações é menor que o número de incógnitas (i.e., para $K > N$, corresponde à situação onde, depois selecionar os pontos de referência, apenas um pequeno número de pontos de aprendizagem é usado, por isso leva a um problema subdeterminado com, normalmente, infinitas soluções.

Dada a possibilidade de $\mathbf{B}$ ser exclusivamente resolvido ou estimado (Equação 4.29), para um ponto de teste $\mathbf{x} \in \mathbb{R}^D$ cujas distâncias de K pontos de referência $\{\mathbf{m}_k\}_{k=1}^K$ são coletadas no vetor $\mathbf{d}(\mathbf{x}, \mathcal{R}) = [d(\mathbf{x}, \mathbf{m}_1), \dots, d(\mathbf{x}, \mathbf{m}_K)]$, as distâncias correspondentes entre a saída desconhecida $\mathbf{y}$ e as saídas conhecidas $\{\mathbf{t}_k\}_{k=1}^K$ dos pontos de referências são estimadas por

$$\hat{\delta}(\mathbf{y}, \mathcal{T}) = \mathbf{d}(\mathbf{x}, \mathcal{R}) \hat{\mathbf{B}}. \quad (4.30)$$

O vetor $\hat{\delta}(\mathbf{y}, \mathcal{T})$ fornece uma estimativa da configuração geométrica entre $\mathbf{y}$ e o conjunto de pontos de referência $\mathcal{T}$, no espaço- $\mathcal{Y}$.

4.5.1.2 Estimativa da saída

O problema de estimar a saída $\mathbf{y}$, dada as saídas $\{\mathbf{t}_k\}_{k=1}^k$ para todos os pontos de referências e uma estimativa $\hat{\delta}(\mathbf{y}, \mathcal{T})$ das suas distâncias mútuas, pode ser entendido

como um problema de multilateração para estimar sua localização em $\mathcal{Y}$. O problema de localizar $\mathbf{y}$ é equivalente a resolver um conjunto sobredeterminado de equações não-lineares correspondentes a K hiper-esferas centradas em $\mathbf{t}_k$ e passando por todo $\mathbf{y}$, isto é, com um raio igual a $\hat{\delta}(\mathbf{y}, \mathbf{t}_k)$:

$$(\mathbf{y} - \mathbf{t}_k)^\top (\mathbf{y} - \mathbf{t}_k) = \hat{\delta}^2(\mathbf{y}, \mathbf{t}_k), \forall = 1, \dots, K. \quad (4.31)$$

O problema na Equação 4.31 pode ser formulado como um problema de otimização onde uma estimativa $\hat{\mathbf{y}}$ pode ser obtida seguindo a minimização:

$$\hat{\mathbf{y}} = \arg \min_{\mathbf{y}} \sum_{k=1}^K ((\mathbf{y} - \mathbf{t}_k)^\top (\mathbf{y} - \mathbf{t}_k) - \hat{\delta}^2(\mathbf{y}, \mathbf{t}_k))^2. \quad (4.32)$$

O objetivo tem um mínimo igual a 0 que pode ser alcançado se, e somente se, $\mathbf{y}$ é a solução da Equação 4.31. Se existe, essa solução é global e única.

Devido à incerteza introduzida pelas estimativas $\hat{\delta}(\mathbf{y}, \mathbf{t}_k)$, uma solução ótima para Equação 4.32 pode ainda ser alcançada usando o método gradiente descendente ou o algoritmo Levenberg-Marquardt.

4.5.2 Máquina de Aprendizagem Mínima para Classificação

Uma importante classe de problemas é classificação (ou reconhecimento de padrões), onde o objetivo é prever categorias, normalmente denotadas por saídas qualitativas (ou rótulos de classe). Novamente, tem-se disponível um conjunto de N pontos de entrada $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^N$, com $\mathbf{x}_i \in \mathbb{R}^D$, e o conjunto dos rótulos correspondentes $\mathcal{L} = \{l_i\}_{i=1}^N$, com $l_i \in \{C_1, \dots, C_S\}$, onde C_j denota a j -ésima classe; para $S = 2$, o problema é dito ser classificação binária, enquanto para $S > 2$ tem-se classificação multiclasse.

A Máquina de Aprendizagem Mínima pode ser estendida para problemas de classificação de uma maneira direta representando as S classes de forma vetorial através de um esquema binário de codificação (1-de- S). Nessa abordagem, uma variável qualitativa com S símbolos é representada por um vetor binário de dimensão S , dos quais somente uma componente é “ativa”. Matematicamente, o conjunto de saídas $\mathcal{Y} = \{\mathbf{y}_i\}_{i=1}^N$, com $\mathbf{y}_i \in \mathbb{R}^S$, em correspondência aos pontos de entrada $\mathcal{X}$, é definida de tal forma que a j -ésima componente de $\mathbf{y}_i$ é igual a α se $l_i = C_j$ e β caso contrário, onde α e β são inteiros com $\alpha > \beta$. Uma escolha usual é $\alpha = 1$ e $\beta = -1$.

Na etapa de classificação de uma observação de teste $\mathbf{x}$ com rótulo de classe desconhecido $l \in \{C_1, \dots, C_S\}$, a classe estimada $\hat{l}$ associada com o vetor de saída estimado $\hat{\mathbf{y}}$ pelo método básico da MLM é dado por $\hat{l} = C_{s^*}$, onde

$$s^* = \arg \max_{s=1, \dots, S} \{\hat{y}^{(s)}\}. \quad (4.33)$$

4.6 Máquina de Aprendizagem Mínima com Vizinhos Mais Próximos

A estratégia de encontrar a saída através do procedimento de otimização pode ser computacionalmente “pesada”. Dessa forma, Mesquita, Gomes e Souza Junior (2015) propuseram uma abordagem rápida para a etapa de estimação da saída. Uma forma alternativa para computar a estimativa para a saída $\mathbf{y}$ é baseada em V vizinhos mais próximos (*nearest neighbors*). Nessa abordagem, uma vez que as distâncias $\hat{\mathbf{d}}(\mathbf{y}, \mathcal{T})$ tenham sido estimadas, pode-se utilizar as classes dos V pontos de referências mais próximos no espaço de saída a $\mathbf{y}$ para estimar a classe do padrão de teste l . Dentre as possibilidades, a estratégia de voto majoritário consiste em atribuir a moda das classes dos V vizinhos mais próximos ao padrão de teste. Neste trabalho, a rede MLM com a abordagem dos vizinhos mais próximos será referenciada pela sigla MLM-NN (*MLM-Nearest Neighbors*).

Como discutido anteriormente, a MLM e a MLM-NN diferem no último passo do teste. No primeiro método, a estimativa de $\mathbf{y}$ pode ser conseguida por um algoritmo de gradiente descendente sobre a função custo na Equação 4.32. Por outro lado, a MLM-NN simplifica essa etapa através dos vizinhos mais próximos, como mostra o Algoritmo 2. O procedimento de treino da MLM e MLM-NN é mostrado no Algoritmo 1.

Algoritmo 1 Algoritmo de treinamento da MLM e MLM-NN

- 1: **procedimento** TREINO(Conjunto de treinamento $\mathcal{X}$, $\mathcal{Y}$ e K)
 - 2: Aleatoriamente selecione K pontos de referência, $\mathcal{R}$, de $\mathcal{X}$ e suas saídas correspondentes, $\mathcal{T}$, de $\mathcal{Y}$.
 - 3: Calcule $\mathbf{D}_x$: As matriz de distância entre $\mathcal{X}$ e $\mathcal{R}$.
 - 4: Calcule Δ_y : As matriz de distância entre $\mathcal{Y}$ e $\mathcal{T}$.
 - 5: Calcule $\hat{\mathbf{B}} = (\mathbf{D}_x^T \mathbf{D}_x)^{-1} \mathbf{D}_x^T \Delta_y$.
- devolve $\hat{\mathbf{B}}$, $\mathcal{R}$ e $\mathcal{T}$.
-

Algoritmo 2 Algoritmo de Teste da MLM-NN

- 1: **procedimento** TESTE_MLM-NN($\hat{\mathbf{B}}$, $\mathcal{R}$, $\mathcal{T}$ e $\mathbf{x}$)
 - 2: Calcule $\mathbf{d}(\mathbf{x}, \mathcal{R})$.
 - 3: Calcule $\hat{\mathbf{d}}(\mathbf{y}, \mathcal{T}) = \mathbf{d}(\mathbf{x}, \mathcal{R}) \hat{\mathbf{B}}$.
 - 4: $\hat{\mathbf{y}} = \arg \min \hat{\mathbf{d}}(\mathbf{y}, \mathcal{T})$.
- devolve $\hat{\mathbf{y}}$;
-

4.7 Conclusão

Este capítulo apresenta os classificadores que são avaliados nesta dissertação, a saber: classificador Bayesiano, SVM, LSSVM, MLP, MLM e MLM-NN. Estes algoritmos abordam diversos conceitos de técnicas adotadas em problemas de classificação e, por-

tanto, conhecimentos básicos sobre o seu funcionamento é importante na compreensão da abordagem proposta.

5 Materiais e Métodos

Esse capítulo apresenta a metodologia adotada no trabalho, juntamente com as contribuições proposta no mesmo. O capítulo inicia com a explicação das etapas empregadas na abordagem proposta, as quais consistem, de forma geral, na captura da imagem, extração de características e classificação. Em seguida, é elucidado como foi realizado o teste de navegação, bem como uma forma de aprimorar o processo de classificação. Por último, as métricas de avaliação são explanadas.

Durante a navegação, o robô utiliza o sistema de localização baseado em visão computacional para se localizar. Esse sistema foi avaliado com imagens de alta resolução com campo de visão convencional e, posteriormente, com imagens omnidirecionais. Os dois ambientes foram modelados em simuladores com intuito de se realizar um primeiro estudo com as imagens sintéticas. Posteriormente, a tarefa de localização e navegação foi avaliada com as imagens reais e, finalmente, os testes práticos foram executados.

A Figura 19 mostra o fluxo adotado na abordagem proposta, o qual inicia com a captura da imagem. A Seção 5.1 apresenta o robô com os dois tipos de câmera juntamente com o mapa topológico utilizado, enquanto o processo de aquisição das imagens é explicado na Seção 5.2. Uma vez obtida a imagem digital, a obtenção de característica da mesma é efetuada através dos extratores descritos nos capítulos anteriores. A Seção 5.3 explica como esses extratores de características foram avaliados. Em seguida, o vetor de características, resultante da etapa anterior, é utilizado como entrada para o classificador reconhecer o ambiente. O processo de classificação é detalhado na Seção 5.4. Além disso, informações do mapa topológico podem ser utilizadas para melhorar o processo de tomada de decisão como é descrito na Seção 5.6. Os métodos SIFT e SURF foram avaliados separadamente devido às características dos mesmos.

5.1 Robô e Mapa Topológico

Um robô móvel adaptado para locomoção em um ambiente doméstico foi utilizado nesse trabalho. Ele consiste de uma plataforma de metal equipado com motores, placas lógicas, baterias e um sistema de engrenagens interconectadas com dois pares de rodas. O robô tem o seu próprio software que comunica com um microcontrolador por telemetria. A Figura 20(a) mostra o robô com uma câmera de alta resolução (GoPro®) adotada na primeira etapa da metodologia, a qual avalia as imagens de visão convencional. Em relação às imagens omnidirecionais, foi utilizada uma lente que captura imagens em 360 graus, como pode ser observado na Figura 20(b).

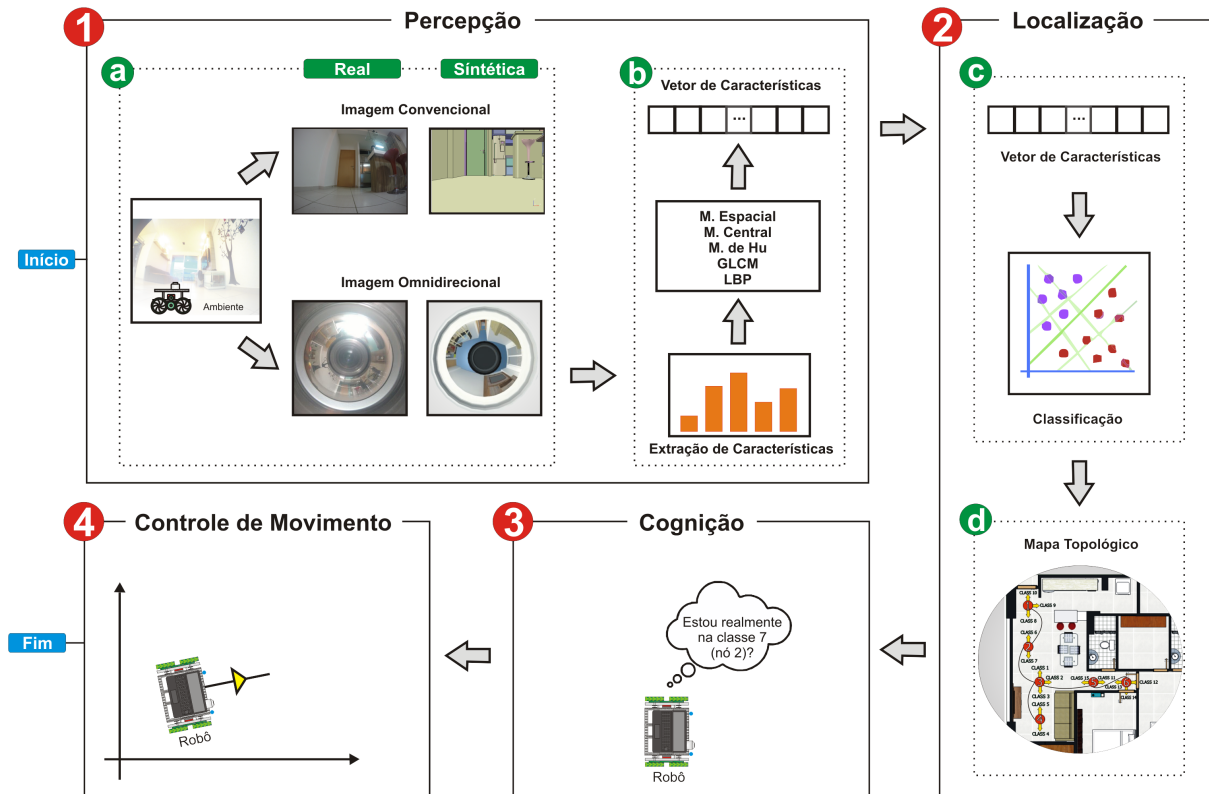
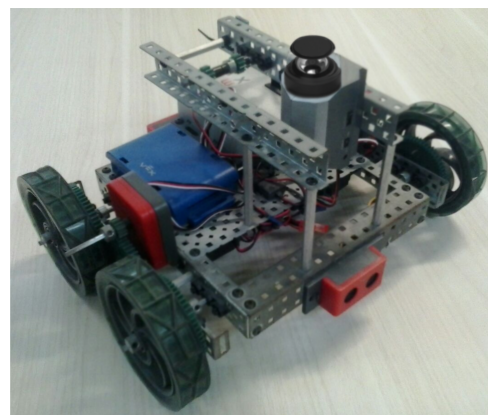
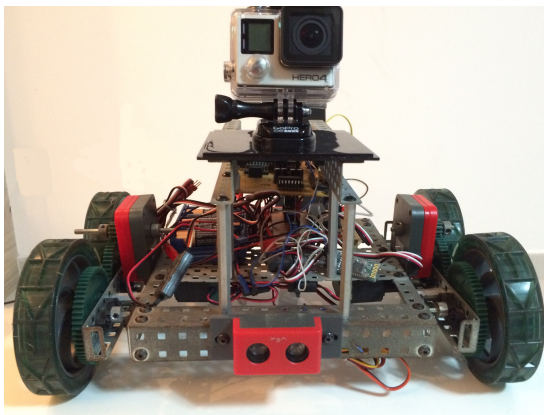


Figura 19 – Etapas do fluxo adotado da abordagem proposta neste trabalho para navegação de robôs móveis.



(a) Robô com a câmera de alta resolução (b) Robô com uma lente que captura imagens em 360 graus.

Figura 20 – Robô utilizado na navegação com as respectivas câmeras adotadas.

O robô navega em um ambiente interno, neste caso um apartamento. Esse ambiente foi escolhido devido à existência de características que ajudam o reconhecimento e a navegação do robô. No mapa topológico, os nós, os quais representam os locais no apartamento, são numerados de 1 a 6. A partir desses nós, foram tiradas fotos de pontos estratégicos das classes numeradas de 1 a 15, essas indicam a orientação dentro de um determinado local. O robô, por exemplo, pode estar localizado na entrada na casa, nó 1 e com a frente posicionada para sala (classe 8), cozinha (classe 9) ou ainda para porta (classe 10). O sistema de navegação e localização proposto nessa dissertação é capaz de identificar essas diversas orientações. Por fim, as arestas são os possíveis caminhos do robô durante a navegação. A Figura 21(a) apresenta o mapa topológico do ambiente de navegação através de uma vista superior, enquanto uma perspectiva diferente do mesmo é apresentado na Figura 21(b).

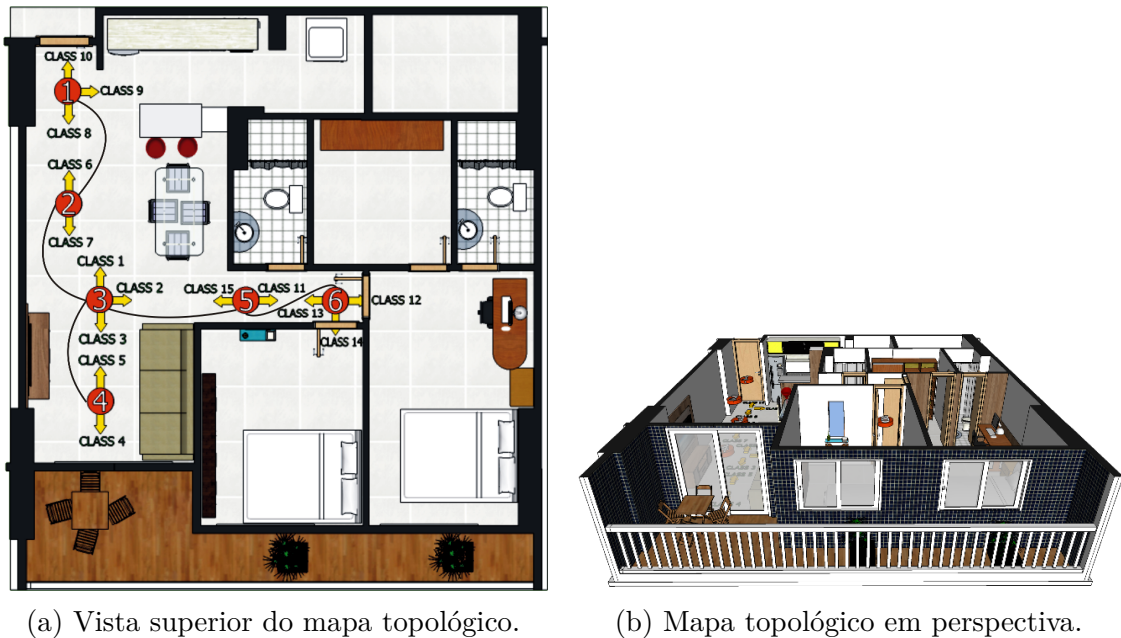


Figura 21 – Mapa topológico do ambiente de navegação através de duas perspectivas.

5.2 Aquisição de Imagens

As imagens sintéticas foram obtidas a partir de simuladores que permitiram a modelagem do ambiente com campo de visão convencional e omnidirecional. A primeira modelagem foi realizada através do simulador V-Rep (ROHMER S. P. N. SINGH, 2013). Primeiramente, foram capturadas 40 imagens de cada classe no ambiente sintético, mas com informações semelhantes do ponto de vista do mesmo nó no mapa. Da mesma forma, foram capturadas 40 imagens de cada classe no ambiente real através de uma câmera de alta resolução apresentada na Figura 20(a). Esse procedimento foi necessário de modo a assegurar um desempenho correto do sistema de navegação. Deste modo, são previstas situações em que o robô está perto do nó, mas não efetivamente nele.

O simulador Sketchup (GASPAR, 2013) juntamente com o V-Ray (LI-ZHEN, 2010) foi usado para modelar o ambiente com a câmera omnidirecional. Similarmente, foram capturadas 50 imagens por classe no ambiente sintético e, em seguida, a mesma quantidade de imagens foi obtida no ambiente real com a câmera omnidirecional exposta na Figura 20(b).

A quantidade de imagens por classe que compõem os quatros conjuntos de imagens descritos anteriormente juntamente com suas resoluções são apresentados na Tabela 4. Para um melhor entendimento, uma imagem sintética e real de cada câmera referente à classe 7 no nó 2 do mapa topológico é apresentada na Figura 22. Uma vez que o problema possui 15 classes, a base de dados com a câmera omnidirecional, por exemplo, contém 750 imagens no total.

Tabela 4 – Quantidade de imagens por classe e resolução dos quatro conjuntos de imagens obtidas.

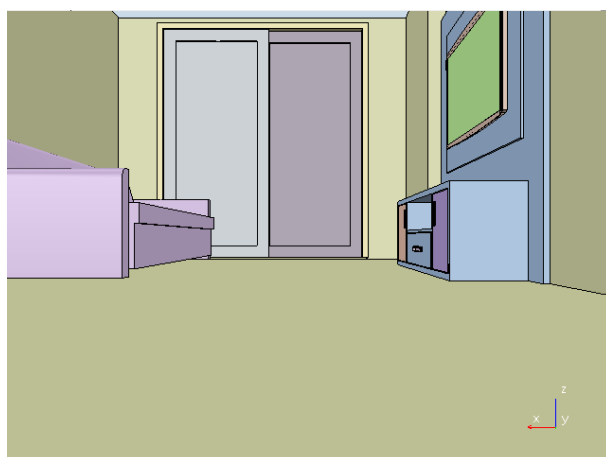
Tipo da Câmera	Quantidade por Classe	Resolução das Imagens (pixels)
Comum (Sintético)	40	640×480
Comum (Real)	40	4000×3000
Omnidirecional (Sintético)	50	800×600
Omnidirecional (Real)	50	4160×3120

5.3 Etapa de Extração

Dentre as técnicas de extração de atributos existentes na literatura, buscou-se diferentes categorias de métodos consolidados, com intuito de escolher a melhor configuração para tarefa de interesse. Diante disso, adotou-se técnicas de extração de características globais e locais baseadas em momentos (Momentos Espaciais, Momentos Centrais e Momentos Invariantes), textura (LBP e GLCM) e descritores de ponto-chave (SIFT e SURF).

Os extratores Momento Espaciais, Momento Centrais, Momentos Invariantes, GLCM e LBP foram aplicados em todas as imagens obtidas na etapa de aquisição. Avaliou-se os momentos de ordem 0, 1, 2 e 3. Para a GLCM, adotou-se a distância $D = 1$ e direção $\theta = 0$. No LBP, utilizou-se a versão do LBP com padrões uniformes com objetivo de diminuir a dimensionalidade do problema. O número de atributos gerados por cada técnica de extração é apresentada na Tabela 5. Posteriormente, as amostras criadas na etapa de extração de atributos foram utilizadas no processo de classificação.

No caso dos métodos SIFT e SURF, dez permutações aleatórias diferentes das imagens capturadas em ambientes virtuais e reais foram realizadas e uma imagem de cada classe foi escolhido para criar o conjunto de treinamento e os restantes para o conjunto de



(a) Comum (Sintético).



(b) Comum (Real).



(c) Omnidirecional (Sintético)



(d) Omnidirecional (Real).

Figura 22 – Imagens capturadas no local referente à classe 7 (nó 2) nos ambientes (a e c) sintético e (b e d) real com a câmera convencional e omnidirecional, respectivamente.

Tabela 5 – Número de atributos gerados por cada técnica de extração de características sobre os quatro conjuntos de imagens.

Extrator de Característica	Número de Atributos
Momentos Espaciais	10
Momentos Centrais	7
Momentos Invariantes	7
GLCM	14
LBP	59

teste. Utilizou-se a biblioteca VLFeat (VEDALDI; FULKERSON, 2008) para o método SIFT e a biblioteca de visão computacional do MATLAB para o método SURF.

5.4 Etapa de Classificação

De forma similar aos extratores de características, adotou-se diferentes tipos de classificadores com objetivo de encontrar a melhor configuração para a tarefa de navegação e localização de robôs móveis com visão computacional. Dentre as categorias de técnicas existentes, utilizou-se o classificado Bayesiano, o qual é baseado em probabilidade e estatística, os classificadores embasados pela Teoria de Aprendizado, SVM e LSSVM, a rede neural artificial MLP e os classificadores baseados em distância, MLM e MLM-NN.

Após a etapa de extração de características, são formados 20 conjuntos de dados provenientes das 5 técnica de extração em cada uma dos 4 conjunto de imagens. Todos as bases de dados foram igualmente pré-processadas. Os padrões foram divididos aleatoriamente em dois grupos, quatro quintos para o treinamento e o restante para teste, em cada um das dez iterações. O conjunto de treino foi normalizado (média zero e variância unitária) e os conjuntos de teste também foram normalizados utilizando as mesmas medidas de normalização que o conjunto de treinamento. A divisão das classes foi balanceada em ambos conjuntos de treino e teste.

Os hiperparâmetros para os classificadores SVM, LSSVM e rede MLP foram escolhidos através das técnicas de validação cruzada com 10 *folds* e da busca em grade (*grid search*) (HAYKIN; ENGEL, 2007). A técnica de validação cruzada é explicada na Seção 5.7.2. Utilizou-se os *kernels* lineares e RBF para os classificadores SVM e LSSVM. Ainda sobre esses classificadores, foi adotado o intervalo $[2^{-2}, 2^{11}]$ na busca de seus hiperparâmetros. A rede MLP foi treinada com o método Levenberg-Marquardt e um intervalo de neurônios na camada oculta de 1 a 50. Adotou-se a função de densidade de probabilidade Gaussiana (normal) no classificador Bayesiano. As MLM e MLM-NN foram treinadas com as métricas de distância Euclidiana, Manhattan e Mahalanobis, enquanto o número dos K pontos de referência foram variados com porcentagens dos dados de treinamento (10%, 20%, 30%, ..., 90%, 100%).

5.5 Rotas de Navegação Adotadas

Finalmente, após a avaliação de diversas configurações de extratores de características e classificadores com as imagens do ambiente real e sintético usando os dois tipos de câmeras, foi realizado o teste de navegação. Para isso, dez diferentes rotas foram avaliadas dez vezes no apartamento. Os comandos usados para mover o robô, assim como, os pontos iniciais e finais de localização de cada rota são apresentados na Tabela 6.

Tabela 6 – Comandos para mover o robô e os pontos iniciais e finais usados no teste de navegação: seguir em frente (SF), virar à esquerda (VE), virar à direita (VD) e girar 180 graus (G180).

Rota	Início (Classe)	Comandos	Fim (Classe)
1	8	SF, SF, SF	4
2	1	SF, SF, VD	9
3	5	SF e VD, SF, SF	12
4	5	SF, SF, SF e VD, VE	10
5	8	SF, SF e VE, SF, SF	12
6	13	SF, SF e VE, SF, G180, SF	1
7	8	SF, SF, VE, SF, SF	12
8	9	VD, SF, SF e VE, SF, G180, SF e VE	3
9	2	SF, SF e VD, VD, SF, SF e VE	3
10	12	G180, SF, SF e VE, SF, G180, SF, SF, SF	10

Com objetivo de facilitar o entendimento da tarefa de navegação, a rota 7 é exemplificada na Figura 23. Nessa rota, o robô inicia na classe 8 no nó 1 e recebe dois comandos para seguir em frente. Depois disso, ele vira à esquerda e recebe novamente dois comandos para seguir em frente e alcança a classe 12 no nó 6.

5.6 Suporte de Decisão para Localização (Proposta)

Mesmo com o bom desempenho dos classificadores, podem ocorrer predições que seriam melhor se descartadas ou tomada outra abordagem de decisão. A classificação com opção de rejeição, por exemplo, é uma área de estudo importante que tenta aumentar a confiabilidade do classificador em alguns sistemas de decisão. Nesta abordagem, um elemento que não é considerado suficientemente confiável é automaticamente rejeitado. O elemento é então dirigido para um especialista humano ou outro classificador (SOUSA et al., 2015; ROCHA NETO, 2011). No entanto, outros procedimentos podem ser adotados além da opção com rejeição dependendo da aplicação. Diante disso, esse trabalho propõe uma nova abordagem de classificação com opção de rejeição para localização de robôs móveis utilizando informação do mapa topológico para otimizar a navegação do robô, chamada de Suporte de Decisão para Localização (SDL). Tal otimização gerada pelo SDL,

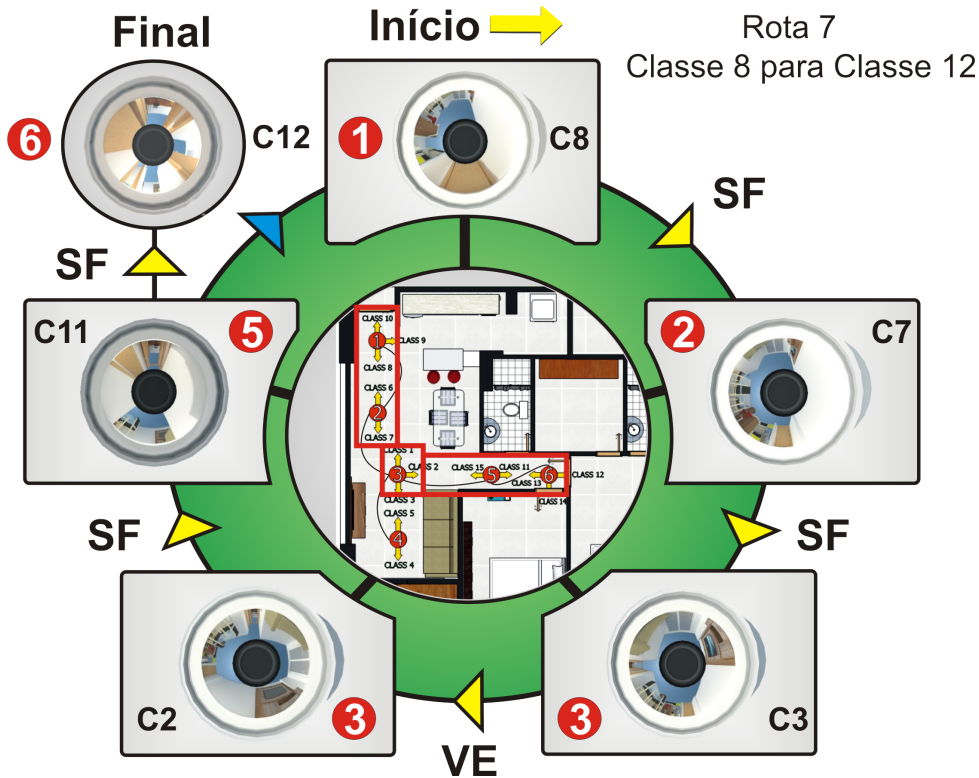


Figura 23 – Ilustração da rota 7 utilizada no teste de navegação. Nessa rota, o robô começa na classe 8 nó 1 e segue em frente duas vezes. Ele vira à esquerda e novamente segue em frente duas vezes. Finalmente, ele alcança a classe 12 no nó 6.

aprimora o reconhecimento da posição atual do robô dentro do mapa topológico pela exclusão de locais afastados do ponto de localização imediatamente anterior à posição atual.

Com intuito de facilitar a compreensão da proposta, imaginemos uma situação representada na Figura 24, na qual o robô está em um local correspondente a classe 1 (nó 3) e recebe comandos para mover para a classe 6 (nó 2). Ao chegar no destino, o classificador prediz que há uma alta probabilidade do local ser a classe 12 (nó 6) e uma segunda maior probabilidade de ser a classe 6, induzindo a um erro de localização. Se o robô utilizasse alguma informação extra, por exemplo, vizinhança entre os nós do mapa topológico, poderia evitar a primeira opção da classificação, já que não é possível chegar a este nó diretamente a partir da posição atual. Nesse contexto, pensou-se em uma abordagem para facilitar a tarefa de localização através de informações de distância no mapa topológico.

Mais formalmente, o mapa topológico é definido como um grafo $G = (V, E)$, onde V é o conjunto de nós (que contêm as classes, ou seja, localização e orientação no espaço) e E (caminhos entre as classes). Seja $\mathbf{x}$ um conjunto de atributos obtidos pelo processo de extração de uma imagem capturada. A resposta de um determinado classificador utilizando esse vetor de característica é uma sequência ordenada $C = (c_1, c_2, \dots, c_l)$, onde l é o número

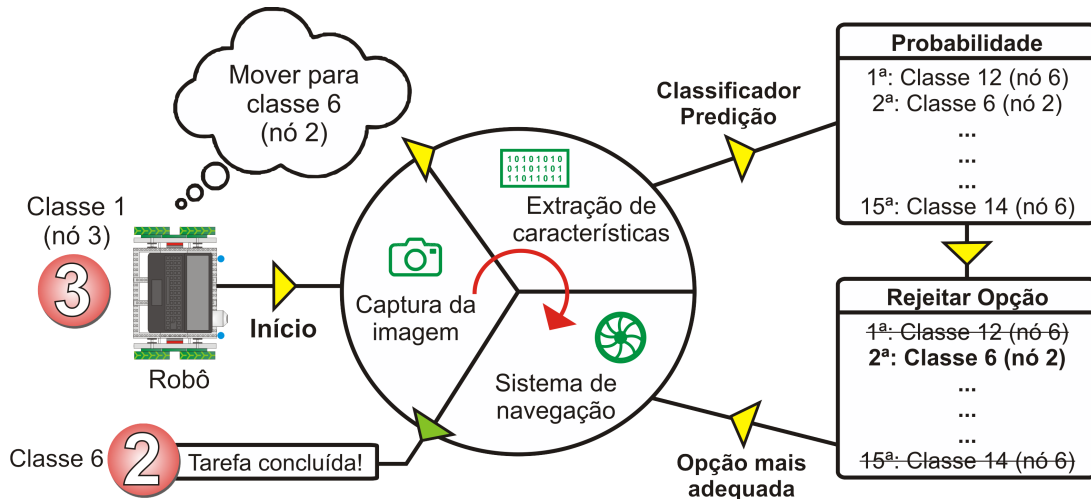


Figura 24 – Situação hipotética na qual o classificador prediz que há uma maior probabilidade da posição atual ser um lugar muito distante do real. O suporte de decisão descarta a primeira sugestão do classificador pelo critério de vizinhança no mapa e escolhe a próxima.

de classes. Nessa sequência C , c_1 e c_2 são as duas classes mais prováveis de $\mathbf{x}$. Quando o robô recebe um comando para mover para um nó n_1 e a primeira classe c_1 predita pelo classificador é uma do nó n_2 , onde n_1 e n_2 não são adjacentes no mapa topológico, então a segunda classe (c_2) é atribuída a $\mathbf{x}$. Caso contrário, a primeira classe é escolhida. Os resultados mostram que o SDL aumentou a acurácia de todos os classificadores avaliados neste trabalho, ou seja, a classificação é otimizada.

O SDL foi inicialmente idealizado para classificadores com saídas probabilísticas, no entanto, outros tipos de abordagens de classificação podem ser adaptados para utilizá-lo. A ideia consiste em ordenar as saídas dos classificadores em uma lista de tal forma que os primeiros valores absolutos da lista sejam mais próximos da saída esperada. Primeiramente, consideremos o caso dos classificadores que usam a representação 1-de- S para classificação multiclasse, tais como a MLM, MLM-NN e a rede MLP nessa dissertação. Nessa situação, as saídas são ordenadas de modo decrescente, uma vez que a componente “ativa” recebe o maior valor. No caso dos classificadores SVM e LSSVM que são avaliados com a abordagem um-contratodos na nossa abordagem, a estratégia é semelhante. Os valores absolutos das saídas, sem passar pela função sinal, são ordenadas de modo decrescente se durante o treinamento a classe em questão for diferenciada das demais pelo maior valor quando utilizada a função sinal. De maneira semelhante, é ordenada de modo crescente no caso da classe ser identificada pelo menor valor. Além disso, as acurácias do treinamento podem ser utilizadas como fator de ponderação para as saídas.

Esta abordagem proposta nesta dissertação está publicada em Marinho et al. (2016).

5.7 Avaliação de Desempenho

Esta seção apresenta as métricas adotadas na avaliação dos resultados obtidos nos testes da abordagem proposta neste trabalho.

5.7.1 Métricas de Avaliação

Com objetivo de analisar o desempenho dos classificadores na tarefa de interesse, seis métricas de avaliação são empregadas e definidas a seguir: acurácia, sensibilidade, especificidade, valor preditivo positivo, $F1$ -score e média harmônica. Essas medidas são obtidas a partir da matriz de confusão. Vale ressaltar que a classe representa a localização e orientação do robô no ambiente de acordo com o mapa topológico. O cálculo de cada categoria é dado da seguinte forma:

- Verdadeiro Positivo (VP): corresponde ao número de vezes que o robô é localizado corretamente pelo modelo de classificação.
- Falso Negativo (FN): corresponde ao número de vezes que a localização em análise é prevista erroneamente como outra localização pelo modelo de classificação.
- Falso Positivo (FP): corresponde ao número de vezes que outra localização é prevista erroneamente como a localização em análise pelo modelo de classificação
- Verdadeiro Negativo (VN): corresponde ao número de vezes que outra localização é prevista corretamente pelo modelo de classificação

A matriz de confusão que resume as localizações do robô previstas correta ou incorretamente por um modelo de classificação é apresentada Tabela 7. Com objetivo de facilitar a compreensão das métricas apresentadas, destacou-se a classe 1 em negrito, uma vez que o VP, FN, FP e VN são relativos às classes analisadas. Por fim, observa-se que de acordo com o número de classes, a ordem da matriz é 15.

Tabela 7 – Matriz de confusão para um problema de classificação multiclasse.

		Classe Prevista				
		1	2	3	...	15
Classe Real	1	VP	FN	FN	FN	FN
	2	FP	VN	FN	FN	FN
	3	FP	FN	VN	FN	FN
	⋮	FP	FN	FN	VN	FN
	15	FP	FN	FN	FN	VN

A acurácia (Ac) é a proporção de predições corretas, sem levar em consideração a diferença entre a localização analisada das demais. Essa medida é influenciada pelo

desbalanceamentos do conjunto de dados e pode facilmente induzir a uma conclusão errada sobre o desempenho do sistema.

$$Ac = \frac{\#total\ de\ acertos}{\#total\ de\ padrões\ no\ conjunto}. \quad (5.1)$$

A sensibilidade (Se) é a proporção de verdadeiros positivos: a capacidade do sistema em prever corretamente a localização para casos onde o robô está de fato.

$$Se = \frac{VP}{VP + FN}, \quad (5.2)$$

em que VP e FN representam o número de amostras de uma determinada classe corretamente e incorretamente classificados, respectivamente.

A especificidade (Es) é a proporção de verdadeiros negativos: a capacidade do sistema em prever corretamente as outras localizações para os casos que realmente estão nessas localizações.

$$Es = \frac{VN}{VN + FP}, \quad (5.3)$$

em que VN representa o número de amostras que não pertencem a uma determinada classe classificadas como não pertencentes à classe considerada, enquanto FP representa o número de amostras classificadas incorretamente como pertencendo a uma determinada classe.

O valor preditivo positivo (VPP) mede a proporção de verdadeiros positivos em relação a todas as predições positivas.

$$VPP = \frac{VP}{VP + FP}. \quad (5.4)$$

O VPP e a sensibilidade podem ser resumidas em outra métrica conhecida como a F -Score, que é a média harmônica ponderada desses valores. Na fórmula F -Score, o valor de β é o coeficiente de ponderação entre o VPP a sensibilidade.

$$F\text{-Score} = (1 + \beta^2) \cdot \frac{VPP \cdot Se}{(\beta^2 \cdot VPP) + Se}. \quad (5.5)$$

Nessa dissertação, o *balance-weighted F-Score*, conhecido como $F1$ -Score, é utilizado. Essa fórmula usa o valor de β igual a 1 e é calculada a seguir. Um valor de $F1$ -Score atinge o seu melhor valor em 1 e pior resultado em 0.

$$F1\text{-Score} = 2 \cdot \frac{VPP \cdot Se}{VPP + Se}. \quad (5.6)$$

Além dessas métricas, foi avaliada, também, a média harmônica (MH) entre a sensibilidade e a especificidade.

$$MH = 2 \cdot \frac{Se \cdot Es}{Se + Es}. \quad (5.7)$$

5.7.2 Validação Cruzada

A validação cruzada é um método estatístico para avaliação e comparação de algoritmos de aprendizado de máquina que divide os dados em dois conjuntos distintos: um para o treinamento e outro para testar o modelo (STONE, 1974).

O tipo mais simples de validação cruzada é o *hold out*. Nesse método, o conjunto das amostras é dividido em dois subconjuntos, onde um se torna o conjunto de treino e o outro é utilizado para a validação. Nessa dissertação, para cada uma dos conjuntos de dados provenientes da etapa de extração de características, particionou-se aleatoriamente os padrões em dois grupos, onde oitenta por cento dos dados foram utilizados na etapa de treinamento e o restante para classificação. Esse procedimento foi repetido dez vezes.

Outro método, chamado de *k-fold*, divide as amostras em K subconjuntos. Nesse caso, um dos subconjuntos torna-se o conjunto de teste e os demais formam o conjunto de treino. Em seguida são feitas k iterações de treino, utilizando $k - 1$ *folds* para o treinamento e o *fold* restante para o teste do modelo. A Figura 25 mostra um exemplo da validação cruzada com *k-fold*. Para cada iteração do *hold out*, utilizou-se a validação cruzada com 10 *folds* juntamente com a busca em grade (*grid search*) (HAYKIN; ENGEL, 2007). O número de *folds* adotado é uma quantidade que fornece uma boa estimativa e é comumente utilizada para validação de algoritmos de aprendizado de máquina (STONE, 1974),

O valor de K pode ser igual ao número de padrões rotulados. Em outras palavras, o conjunto de teste pode consistir em apenas um elemento, e a operação pode ser repetida para cada padrão. Tal variante é conhecida como *leave-one-out* e é usada, principalmente, se o número de amostras disponíveis é bastante limitado.

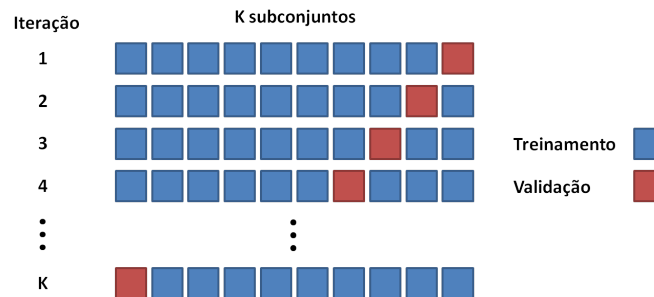


Figura 25 – Exemplo de validação cruzada com *k-fold*. Os conjuntos de treinamento são representados em cor azul e o de validação em vermelho.

5.8 Conclusão

O capítulo apresentou uma abordagem de navegação para robôs móveis em locais internos através de uma mapa topológico utilizando dois tipos de câmera, com campo de visão convencional e omnidirecional. Além disso, um método para otimizar a classificação na tarefa de interesse foi apresentado. Resumidamente, o trabalho segue as seguintes etapas: aquisição de imagens, extração de características, classificação e avaliação a partir das métricas discutidas.

No próximo capítulo serão apresentados os resultados obtidos com as variações desta abordagem com objetivo de identificar a mais adequada para navegação de robôs móveis.

6 Resultados e Discussões

Nesse capítulo são apresentados os resultados obtidos pela combinação dos extratores de características com os classificadores na tarefa de localização. Para essa análise foram utilizadas os quatro conjuntos de imagens provenientes do ambiente sintético e real com os dois tipos de câmeras. Além disso, o teste de navegação com o robô é validado juntamente com a abordagem proposta de Suporte de Decisão para Localização.

Os resultados são apresentados em dois grupos maiores: GoPro[®] e Omnidirecional. Os experimentos foram computados em um iMac com processador Core i5 2.5 GHz, 4GB RAM e sistema operacional Mac OS X Yosemite 10.10. Tanto os extratores e os classificadores foram simulados no Matlab[®] R2014b.

6.1 Resultados com a câmera GoPro[®]

O primeiro grupo de resultados é proveniente de imagens com campo de visão convencional. Primeiramente, os extratores de características são aplicados nas imagens sintéticas do ambiente simulado gerando as amostras de entrada para os classificadores. A Tabela 8 mostra os valores médios e os desvios padrões da especificidade (Es), sensibilidade (Se), vapor preditivo positivo (VPP), F1-Score, média harmônica (MH) e acurácia (Ac) obtidas pela combinação dos classificadores com extratores em imagens do ambiente sintético. O classificador com maior acurácia em cada extrator está destacado em verde e os menores em vermelho.

Ao analisar a Tabela 8, depreende-se que a maioria das combinações obtiveram bons resultados. Dentre os extratores, o LBP destaca-se com valores superiores à 99%, inclusive com os classificadores lineares, SVM e LSSVM Linear. Esse resultado é esperado devido às suas propriedades. O LBP é um extrator de textura invariante à translação e robusto às mudanças nos níveis de intensidade, por exemplo, por variações na iluminação. Além disso, a sua versão de padrões uniformes o torna invariante à rotação. A MLM-NN e MLM podem ser destacadas entre os classificadores que conseguiram os maiores valores independente do extrator adotado. A combinação que obteve a maior acurácia (99,98%), ratificada pelos valores de Média Harmônica (99,98%) e F1-Score (99,95%) foi o LBP com MLM e MLM-NN utilizando a distância Manhattan.

Tabela 8 – Especificidade (Es), sensibilidade (Se), valor preditivo positivo (VPP), F1-score, média harmônica (MH) and Acurácia (Ac) obtidos pelos extratores e classificadores no ambiente **sintético** usando as imagens com campo de visão **convencional**.

Extrator	Classificador	Configuração	Es (%)	Se (%)	VPP (%)	F1-Score (%)	MH (%)	Ac (%)
Momentos Espaciais	Bayes	Normal	99,98±0,02	99,68±0,23	99,70±0,21	99,68±0,23	99,82±0,13	99,68±0,23
	LSSVM	Linear	94,80±0,17	56,19±0,96	62,04±3,81	51,31±1,20	61,69±1,48	56,19±0,96
		RBF	99,95±0,03	99,40±0,37	99,42±0,35	99,39±0,37	99,67±0,20	99,40±0,37
	MLM	Manhattan	99,80±0,05	97,22±0,61	97,31±0,60	97,22±0,61	98,46±0,35	97,22±0,61
		Euclidiana	99,97±0,02	99,57±0,20	99,59±0,19	99,57±0,20	99,77±0,11	99,57±0,20
		Mahalanobis	99,99±0,01	99,89±0,10	99,90±0,09	99,89±0,10	99,94±0,05	99,89±0,10
	MLM-NN	Manhattan	99,80±0,05	97,30±0,69	97,39±0,66	97,30±0,69	98,51±0,39	97,30±0,69
		Euclidiana	99,97±0,02	99,57±0,20	99,59±0,19	99,57±0,20	99,77±0,11	99,57±0,20
		Mahalanobis	99,99±0,01	99,89±0,10	99,90±0,09	99,89±0,10	99,94±0,05	99,89±0,10
	MLP		99,94±0,03	99,15±0,46	99,18±0,45	99,15±0,46	99,53±0,25	99,15±0,46
SVM	Linear	97,09±0,37	60,39±5,05	56,20±4,29	55,30±4,96	59,77±5,42	60,39±5,05	
	RBF	99,92±0,04	98,83±0,61	98,95±0,53	98,83±0,61	99,34±0,35	98,83±0,61	
Momentos Centrais	Bayes	Normal	99,94±0,03	99,15±0,35	99,18±0,33	99,15±0,34	99,53±0,19	99,15±0,35
	LSSVM	Linear	93,18±0,34	49,31±1,37	44,49±3,49	41,56±1,41	52,66±1,24	49,31±1,37
		RBF	99,98±0,01	99,66±0,19	99,67±0,18	99,66±0,19	99,82±0,10	99,66±0,19
	MLM	Manhattan	99,71±0,06	96,13±0,83	96,22±0,82	96,11±0,83	97,83±0,47	96,13±0,83
		Euclidiana	99,99±0,01	99,81±0,18	99,82±0,17	99,81±0,18	99,90±0,10	99,81±0,18
		Mahalanobis	99,98±0,02	99,69±0,24	99,70±0,22	99,69±0,24	99,83±0,13	99,69±0,24
	MLM-NN	Manhattan	99,72±0,06	96,15±0,80	96,24±0,78	96,13±0,80	97,84±0,45	96,15±0,80
		Euclidiana	99,99±0,01	99,81±0,18	99,82±0,17	99,81±0,18	99,90±0,10	99,81±0,18
		Mahalanobis	99,98±0,02	99,67±0,24	99,68±0,23	99,67±0,24	99,82±0,13	99,67±0,24
	MLP		99,94±0,12	99,16±1,48	98,91±2,67	99,01±2,13	99,38±1,52	99,16±1,48
SVM	Linear	97,15±0,46	70,50±3,79	68,58±3,53	65,59±4,12	74,89±3,64	70,50±3,79	
	RBF	99,97±0,02	99,55±0,23	99,57±0,22	99,55±0,23	99,75±0,13	99,55±0,23	
Momentos Invariantes	Bayes	Normal	99,88±0,03	98,26±0,47	98,34±0,44	98,26±0,47	99,04±0,26	98,26±0,47
	LSSVM	Linear	90,24±0,39	36,62±1,14	38,59±4,47	27,69±1,36	37,53±1,45	36,62±1,14
		RBF	99,89±0,04	98,50±0,55	98,59±0,50	98,51±0,54	99,17±0,30	98,50±0,55
	MLM	Manhattan	99,36±0,08	91,78±0,96	91,87±0,98	91,67±0,98	95,22±0,60	91,78±0,96
		Euclidiana	99,93±0,03	99,07±0,44	99,08±0,43	99,06±0,44	99,49±0,24	99,07±0,44
		Mahalanobis	99,93±0,03	98,99±0,41	99,01±0,40	98,98±0,41	99,44±0,22	98,99±0,41
	MLM-NN	Manhattan	99,39±0,09	92,01±1,08	92,08±1,09	91,92±1,09	95,37±0,66	92,01±1,08
		Euclidiana	99,93±0,03	99,07±0,44	99,08±0,43	99,06±0,44	99,49±0,24	99,07±0,44
		Mahalanobis	99,93±0,03	98,99±0,40	99,01±0,38	98,98±0,40	99,45±0,22	98,99±0,40
	MLP		99,76±0,16	96,66±2,16	96,28±3,48	96,34±2,90	97,78±2,09	96,66±2,16
SVM	Linear	96,65±0,55	65,58±4,70	66,66±5,20	61,14±5,04	69,47±5,23	65,58±4,70	
	RBF	99,87±0,04	98,13±0,62	98,31±0,50	98,15±0,61	98,96±0,36	98,13±0,62	
GLCM	Bayes	Normal	99,65±0,07	95,49±0,79	95,67±0,74	95,47±0,80	97,44±0,47	95,49±0,79
	LSSVM	Linear	97,54±0,09	73,99±0,70	84,54±2,58	71,17±0,94	78,02±0,83	73,99±0,70
		RBF	99,96±0,02	99,45±0,30	99,46±0,29	99,45±0,30	99,70±0,17	99,45±0,30
	MLM	Manhattan	99,91±0,03	98,83±0,36	98,86±0,35	98,83±0,36	99,36±0,20	98,83±0,36
		Euclidiana	99,95±0,02	99,32±0,28	99,33±0,28	99,32±0,28	99,63±0,15	99,32±0,28
		Mahalanobis	99,97±0,02	99,65±0,27	99,66±0,26	99,65±0,27	99,81±0,15	99,65±0,27
	MLM-NN	Manhattan	99,91±0,03	98,86±0,39	98,89±0,38	98,86±0,39	99,38±0,21	98,86±0,39
		Euclidiana	99,95±0,02	99,33±0,28	99,34±0,27	99,33±0,28	99,63±0,15	99,33±0,28
		Mahalanobis	99,97±0,02	99,65±0,27	99,66±0,27	99,65±0,27	99,81±0,15	99,65±0,27
	MLP		99,91±0,03	98,82±0,45	98,85±0,44	98,81±0,45	99,35±0,25	98,82±0,45
SVM	Linear	99,23±0,29	89,85±3,96	92,01±4,66	88,73±4,88	92,00±4,23	89,85±3,96	
	RBF	99,93±0,04	99,02±0,54	99,09±0,46	99,02±0,54	99,46±0,31	99,02±0,54	
LBP	Bayes	Normal	99,95±0,02	99,27±0,30	99,30±0,29	99,27±0,30	99,60±0,17	99,27±0,30
	LSSVM	Linear	99,70±0,08	95,95±1,00	96,44±0,79	95,98±0,99	97,69±0,60	95,95±1,00
		RBF	99,97±0,02	99,63±0,28	99,65±0,26	99,63±0,28	99,80±0,15	99,63±0,28
	MLM	Manhattan	100,00±0,0	99,95±0,10	99,96±0,10	99,95±0,10	99,98±0,05	99,95±0,10
		Euclidiana	99,99±0,01	99,88±0,20	99,89±0,19	99,88±0,20	99,94±0,11	99,88±0,20
		Mahalanobis	99,99±0,01	99,86±0,19	99,87±0,19	99,86±0,19	99,93±0,11	99,86±0,19
	MLM-NN	Manhattan	100,00±0,0	99,95±0,10	99,95±0,10	99,95±0,10	99,97±0,06	99,95±0,10
		Euclidiana	99,99±0,01	99,88±0,20	99,89±0,19	99,88±0,20	99,94±0,11	99,88±0,20
		Mahalanobis	99,99±0,01	99,86±0,19	99,87±0,19	99,86±0,19	99,93±0,11	99,86±0,19
	MLP		99,96±0,02	99,53±0,32	99,55±0,31	99,53±0,32	99,74±0,18	99,53±0,32
SVM	Linear	99,93±0,03	99,08±0,35	99,12±0,34	99,08±0,34	99,50±0,19	99,08±0,35	
	RBF	99,96±0,03	99,49±0,35	99,52±0,32	99,50±0,35	99,72±0,19	99,49±0,35	
SIFT			99,02±0,14	87,62±1,60	89,77±1,27	86,89±1,86	91,59±1,42	87,62±1,60
SURF			98,69±0,56	84,67±5,67	87,38±4,10	83,70±6,50	89,23±4,93	84,67±5,67

A etapa posterior dos experimentos foi realizada com as imagens da câmera GoPro®. Do mesmo modo, o processo de extração é realizado para uma posterior classificação. A Tabela 9 exibe os resultados obtidos pela combinação dos classificadores e extratores, agora, com imagens do ambiente real.

De acordo com a Tabela 9, percebe-se que os resultados com as imagens dos ambientes real e sintético foram semelhante, como esperado. No entanto, algumas combinações de extratores e classificadores conseguiram resultados melhores. Isso ocorreu devido a ausência de alguns detalhes durante a modelagem do ambiente na simulação computacional. A acurácia da combinação GLCM com SVM Linear, por exemplo, passou de 89,85% para 98,08%. O SIFT e o SURF foram beneficiados também com os detalhes presentes nesse conjunto de imagens. Vale destacar a eficiência do classificador LBP que se manteve. Em contrapartida, os Momentos Invariantes não conseguiram resultados semelhante aos demais. Esse extrator costuma ter melhor desempenho na detecção de formas. Quanto aos classificadores, novamente, a MLM e MLM-NN permaneceram com os maiores valores de acurácia em todos os extratores. Os classificadores com maior acurácia foram o MLM e MLM-NN com a distância Mahalanobis que obtiveram 99,96% com os Momentos Espaciais e 99,83% com o LBP.

Outra análise importante é quanto ao tempo de processamento dos métodos. A próxima seção expõe os resultados dos métodos destacados em negrito na Tabela 9.

6.1.1 Análise do tempo de processamento (GoPro®)

A Tabela 10 mostra características que são relevantes em sistemas embarcados tais como: acurácia, tempo de predição de um padrão e de extração. O classificador com menor tempo de classificação em cada extrator está destacado em verde e os maiores em vermelho.

Analizando o custo computacional das etapas de treinamento e teste, verifica-se que os resultados são semelhantes entre os extratores. A rede neural MLP foi o classificador que consumiu mais tempo na etapa de treinamento, 11,71 segundos, seguido do SVM com 1,04 segundos e LSSVM com 0,44 segundos. Observa-se ainda que o LSSVM obteve menos tempo de treino em relação ao SVM. No geral, o descritor SIFT apresentou o pior tempo de treinamento (181,32s) e de teste juntamente com o SURF (38,11s e 3,43s). Em contrapartida, o classificador Bayesiano foi o método mais rápido para reconhecer uma amostra desconhecida, gastando em média 47,3µs com os Momentos Espaciais. Além disso, é possível ver que a MLM-NN consegue uma considerável redução no tempo de teste em relação a MLM. Com os Momentos Invariantes, um dos extratores com menor número de atributos, a MLM-NN foi 305 vezes mais rápida que a MLM.

Tabela 9 – Especificidade (Es), sensibilidade (Se), valor preditivo positivo (VPP), F1-score, média harmônica (MH) and Acurácia (Ac) obtidos pelos extratores e classificadores no ambiente **real** usando as imagens provenientes da câmera **GoPro[®]**.

Extrator	Classificador	Configuração	Es (%)	Se (%)	VPP (%)	F1-Score (%)	MH (%)	Ac (%)
Momentos Espaciais	Bayes	Normal	100,00±0,0	99,94±0,21	99,95±0,19	99,94±0,21	99,97±0,12	99,94±0,21
	LSSVM	Linear	94,67±0,64	55,17±3,15	54,25±5,35	47,95±4,22	58,72±3,89	55,17±3,15
		RBF	99,95±0,05	99,25±0,71	99,33±0,65	99,25±0,72	99,57±0,41	99,25±0,71
	MLM	Manhattan	99,68±0,14	95,67±1,81	96,29±1,38	95,62±1,95	97,39±1,30	95,67±1,81
		Euclidiana	99,92±0,07	98,92±0,90	99,04±0,77	98,92±0,90	99,38±0,52	98,92±0,90
	Mahalanobis	100,00±0,0	99,96±0,19	99,96±0,17	99,96±0,19	99,98±0,11	99,96±0,19	
	MLM-NN	Manhattan	99,68±0,14	95,67±1,81	96,28±1,39	95,61±1,95	97,39±1,30	95,67±1,81
		Euclidiana	99,92±0,07	98,92±0,90	99,04±0,77	98,92±0,90	99,38±0,52	98,92±0,90
	Mahalanobis	100,00±0,0	99,96±0,19	99,96±0,17	99,96±0,19	99,98±0,11	99,96±0,19	
	MLP		99,85±0,18	98,00±2,32	97,66±3,65	97,71±3,10	98,53±2,22	98,00±2,32
SVM	Linear	99,72±0,15	96,12±1,99	96,51±1,95	96,03±2,08	97,52±1,43	96,12±1,99	
	RBF	99,91±0,08	98,71±1,16	98,84±1,02	98,71±1,14	99,25±0,68	98,71±1,16	
Momentos Centrais	Bayes	Normal	100,00±0,00	99,94±0,21	99,95±0,19	99,94±0,21	99,97±0,12	99,94±0,21
	LSSVM	Linear	94,88±0,46	56,00±2,35	52,89±4,06	47,66±2,98	58,58±2,79	56,00±2,35
		RBF	99,99±0,02	99,88±0,31	99,89±0,27	99,87±0,31	99,93±0,17	99,88±0,31
	MLM	Manhattan	99,85±0,08	97,92±1,03	98,15±0,92	97,90±1,04	98,79±0,61	97,92±1,03
		Euclidiana	99,99±0,02	99,83±0,34	99,85±0,30	99,83±0,34	99,91±0,19	99,83±0,34
	Mahalanobis	100,00±0,0	99,96±0,19	99,96±0,17	99,96±0,19	99,98±0,11	99,96±0,19	
	MLM-NN	Manhattan	99,85±0,08	97,92±1,03	98,15±0,92	97,90±1,04	98,79±0,61	97,92±1,03
		Euclidiana	99,99±0,02	99,83±0,34	99,85±0,30	99,83±0,34	99,91±0,19	99,83±0,34
	Mahalanobis	100,00±0,0	99,96±0,19	99,96±0,17	99,96±0,19	99,98±0,11	99,96±0,19	
	MLP		99,95±0,11	99,33±1,52	99,13±2,59	99,19±2,13	99,46±1,54	99,33±1,52
SVM	Linear	99,30±0,22	90,67±2,64	94,10±2,21	90,38±2,76	93,38±1,99	90,67±2,64	
	RBF	99,98±0,04	99,67±0,57	99,71±0,49	99,67±0,56	99,81±0,32	99,67±0,57	
Momentos Invariantes	Bayes	Normal	99,83±0,11	97,71±1,40	98,05±1,11	97,72±1,39	98,68±0,83	97,71±1,40
	LSSVM	Linear	88,73±1,01	27,38±2,71	24,78±4,29	19,53±2,66	25,47±3,34	27,38±2,71
		RBF	99,75±0,15	96,54±1,90	97,15±1,39	96,53±1,92	97,95±1,21	96,54±1,90
	MLM	Manhattan	99,04±0,23	87,88±2,57	88,86±2,45	87,70±2,59	92,39±1,68	87,88±2,57
		Euclidiana	99,80±0,14	97,25±1,82	97,54±1,64	97,23±1,83	98,40±1,08	97,25±1,82
	Mahalanobis	99,82±0,11	97,50±1,48	97,80±1,31	97,46±1,51	98,52±0,90	97,50±1,48	
	MLM-NN	Manhattan	99,04±0,23	87,88±2,61	88,86±2,47	87,71±2,62	92,39±1,70	87,88±2,61
		Euclidiana	99,80±0,14	97,25±1,82	97,54±1,64	97,23±1,83	98,40±1,08	97,25±1,82
	Mahalanobis	99,82±0,11	97,50±1,48	97,80±1,31	97,46±1,51	98,52±0,90	97,50±1,48	
	MLP		99,53±0,35	93,83±4,41	93,40±6,51	93,19±5,68	95,56±4,38	93,83±4,41
SVM	Linear	95,98±0,58	59,50±4,05	68,52±4,53	58,65±4,01	66,12±3,74	59,50±4,05	
	RBF	99,62±0,18	94,92±2,31	95,98±1,63	94,96±2,28	96,96±1,46	94,92±2,31	
GLCM	Bayes	Normal	99,77±0,07	96,83±1,00	97,40±0,81	96,82±0,99	98,12±0,60	96,83±1,00
	LSSVM	Linear	98,08±0,22	77,75±2,13	79,54±2,50	74,06±2,67	80,77±2,28	77,75±2,13
		RBF	99,94±0,05	99,17±0,66	99,27±0,56	99,16±0,67	99,52±0,39	99,17±0,66
	MLM	Manhattan	99,94±0,05	99,17±0,66	99,25±0,61	99,16±0,67	99,52±0,38	99,17±0,66
		Euclidiana	99,90±0,07	98,62±0,91	98,81±0,77	98,62±0,91	99,20±0,53	98,62±0,91
	Mahalanobis	99,96±0,05	99,42±0,72	99,50±0,61	99,41±0,73	99,66±0,42	99,42±0,72	
	MLM-NN	Manhattan	99,94±0,05	99,17±0,66	99,25±0,61	99,16±0,67	99,52±0,38	99,17±0,66
		Euclidiana	99,90±0,07	98,62±0,91	98,81±0,77	98,62±0,91	99,20±0,53	98,62±0,91
	Mahalanobis	99,96±0,05	99,42±0,72	99,50±0,61	99,41±0,73	99,66±0,42	99,42±0,72	
	MLP		99,93±0,07	98,96±1,01	99,11±0,79	98,95±1,04	99,39±0,65	98,96±1,01
SVM	Linear	99,86±0,07	98,08±1,02	98,40±0,78	98,05±1,06	98,85±0,66	98,08±1,02	
	RBF	99,88±0,07	98,29±0,95	98,55±0,79	98,30±0,94	99,01±0,57	98,29±0,95	
LBP	Bayes	Normal	99,97±0,04	99,58±0,51	99,63±0,45	99,58±0,51	99,76±0,29	99,58±0,51
	LSSVM	Linear	99,97±0,04	99,58±0,57	99,64±0,49	99,58±0,57	99,76±0,33	99,58±0,57
		RBF	99,97±0,04	99,58±0,51	99,63±0,45	99,58±0,51	99,76±0,29	99,58±0,51
	MLM	Manhattan	99,98±0,04	99,67±0,50	99,70±0,44	99,67±0,50	99,81±0,28	99,67±0,50
		Euclidiana	99,98±0,04	99,67±0,50	99,70±0,44	99,67±0,50	99,81±0,28	99,67±0,50
	Mahalanobis	99,99±0,02	99,83±0,34	99,85±0,30	99,83±0,34	99,91±0,19	99,83±0,34	
	MLM-NN	Manhattan	99,98±0,04	99,67±0,50	99,70±0,44	99,67±0,50	99,81±0,28	99,67±0,50
		Euclidiana	99,98±0,04	99,67±0,50	99,70±0,44	99,67±0,50	99,81±0,28	99,67±0,50
	Mahalanobis	99,99±0,02	99,83±0,34	99,85±0,30	99,83±0,34	99,91±0,19	99,83±0,34	
	MLP		99,94±0,16	99,17±2,16	98,68±3,68	98,88±3,04	99,21±2,19	99,17±2,16
SVM	Linear	99,96±0,05	99,42±0,72	99,50±0,61	99,42±0,72	99,66±0,42	99,42±0,72	
	RBF	99,97±0,04	99,63±0,50	99,67±0,45	99,62±0,51	99,79±0,29	99,63±0,50	
SIFT			99,96±0,07	99,37±0,94	99,47±0,77	99,36±0,97	99,62±0,58	99,37±0,94
SURF			99,78±0,17	97,03±2,27	97,56±1,70	96,86±2,43	98,06±1,59	97,03±2,27

Tabela 10 – Acurácia (Ac), tempo de treino, tempo de teste e de extração com todos extratores combinados com as melhores configurações dos classificadores utilizando as imagens da GoPro®.

Momentos Espaciais				
Classificador	Ac (%)	Tempo de Treino (s)	Tempo de Classificação (μs)	Tempo de Extração (s)
MLM (Mahalanobis)	99,96±0,19	0,08±0,01	3,29e4±9,5e2	0,165±0,01
MLM-NN (Mahalanobis)	99,96±0,19	0,08±0,01	120,8±11,0	
Bayes	99,94±0,21	0,003±0,01	47,3±3,2	
LSSVM (RBF)	99,25±0,71	0,44±0,01	5,8e3±50	
MLM-NN (Euclidiana)	98,92±0,90	0,04±0,01	90,6±2,0	
SVM (RBF)	98,71±1,16	1,04±0,01	5,19e3 ±50	
MLP	98,00±2,32	3,33±0,6	305,4±4,6	
Momentos Centrais				
Classificador	Ac (%)	Tempo de Treino (s)	Tempo de Classificação (μs)	Tempo de Extração (s)
MLM (Mahalanobis)	99,96±0,19	0,07±0,0	3,25e4±200	0,169±0,01
MLM-NN (Mahalanobis)	99,96±0,19	0,08±0,0	110,5±3,4	
Bayes	99,94±0,21	0,003±0,0	48,5±22,4	
MLM-NN (Euclidiana)	99,83±0,34	0,04±0,0	94,4±3,4	
LSSVM (RBF)	99,88±0,31	0,43±0,0	5,75e3±30	
SVM (RBF)	99,67±0,57	0,49±0,0	5,33e3±60	
MLP	99,33±1,52	3,75±1,0	339,6±8,5	
Momentos Invariantes				
Classificador	Ac (%)	Tempo de Treino (s)	Tempo de Classificação (μs)	Tempo de Extração (s)
Bayes	97,71±1,40	0,003±0,0	50,3±23,7	0,174±0,01
MLM (Mahalanobis)	97,50±1,48	0,07±0,0	3,26e4±1,9e2	
MLM-NN (Mahalanobis)	97,50±1,48	0,08±0,0	106,8±9,1	
MLM-NN (Euclidiana)	97,25±1,82	0,04±0,0	89,6±4,0	
LSSVM (RBF)	96,54±1,90	0,41±0,0	5,83e3±1,2e2	
SVM (RBF)	94,92±2,31	0,77±0,0	5,35e3±1,2e2	
MLP	93,83±4,41	9,71±2,6	319,7±6,1	
GLCM				
Classificador	Ac (%)	Tempo de Treino (s)	Tempo de Classificação (μs)	Tempo de Extração (s)
MLM (Mahalanobis)	99,42±0,72	0,12±0,0	3,26e4±2,90e2	0,494±0,03
MLM-NN (Mahalanobis)	99,42±0,72	0,12±0,0	210,8±11,1	
MLM-NN (Manhattan)	99,17±0,66	0,04±0,0	92,8±2,6	
LSSVM (RBF)	99,17±0,66	0,44±0,0	6,24e3±1e2	
MLP	98,96±1,01	11,71±3,9	351,3±9,8	
Bayes	96,83±1,00	0,005±0,0	65,6±4,5	
SVM (RBF)	98,29±0,95	0,49±0,0	5,4e3±70	
LBP				
Classificador	Ac (%)	Tempo de Treino (s)	Tempo de Classificação (μs)	Tempo de Extração (s)
MLM (Mahalanobis)	99,83±0,34	0,51±0,1	3,99e4±6,18e3	0,311±0,02
MLM-NN (Mahalanobis)	99,83±0,34	0,47±0,0	942,0±13,9	
MLM-NN (Manhattan)	99,67±0,50	0,05±0,0	114,7±9,5	
Bayes	99,58±0,51	0,007±0,0	108,8±70,5	
SVM (RBF)	99,63±0,50	0,42±0,0	5,22e3±50	
LSSVM (RBF)	99,68±0,51	0,40±0,0	7,13e3±60	
MLP	99,17±2,16	5,63±0,8	403,2±186,5	
Descritor	Ac (%)	Tempo de Treino (s)	Tempo de Classificação (μs)	
SIFT	99,37±0,94	181,32±1,4	3,81e7±9e5	15,654±0,33
SURF	97,03±2,27	40,71±0,2	3,43e6±2,2e5	3,091±0,01

O tempo médio gasto no processo de extração também foi computado e é exibido na última coluna da Tabela 10. A partir dos resultados apresentados, depreende-se que os Momentos Espaciais conseguiram menor tempo de extração, 0,165s. Por outro lado, o SIFT obteve o pior tempo, 15,654s. O LBP pode ser destacado entre os extratores de textura. No entanto, o número de atributos gerado por esse método ocasionou um moderado aumento nos tempos de treinamento e teste dos classificadores. Outra observação pertinente é quanto a combinação entre as métricas de distância da MLM-NN e os extratores. Com os momentos, a distância Euclidiana reduziu o tempo de treinamento e teste da MLM-NN enquanto manteve a acurácia. Ao passo que com a GLCM e LBP, foi suficiente utilizar a distância Manhattan na MLM-NN para se obter resultados semelhantes.

6.1.2 Teste de Navegação (GoPro®)

Finalmente, após a análise isolada das imagens capturadas no ambiente simulado e real é feito um teste de navegação com as melhores configurações dos métodos.

A Figura 26 compara os resultados gerais do teste de navegação, com e sem o Suporte de Decisão para Localização, obtidos pela combinação entre os classificadores e extratores de características, além disso, a figura mostra os desempenhos do SIFT e do SURF. Uma rota é considerada correta se o sistema de classificação baseado em visão acerta todos os lugares (classes) durante o trajeto. Os resultados detalhados por rota com e sem a aplicação do SDL são apresentados nas Tabelas 11 e 12. Os valores dos hiperparâmetros para os classificadores são detalhados no Apêndice A.

Analisando os resultados do teste de navegação na Figura 26 juntamente com as Tabelas 11 e 12, verifica-se que os métodos tiveram desempenho semelhante aos experimentos isolados com as imagens da GoPro®. Os extratores Momentos Centrais, Espaciais e LBP obtiveram acurácia entre 98% e 100%. O último foi eficiente, inclusive, com os classificadores lineares. O SIFT pode ser destacado também pela sua acurácia, 98%, no entanto, esse descritor tornou os experimentos mais lentos devido seu elevado custo computacional. Em contrapartida, os Momentos Invariantes não alcançaram bom desempenho. Esse método é mais adequado na detecção de formas devido sua formulação.

Em relação aos classificadores, destacam-se a MLM, MLM-NN e o classificador Bayesiano que conseguiram os maiores valores de acurácia em todos os extratores, 98%-100%. De maneira semelhante aos testes isolados, a MLM-NN conseguiu o mesmo desempenho da MLM, enquanto reduziu significativamente o tempo para reconhecer uma amostra desconhecida. Ademais, observa-se que a influência das métricas de distância foram semelhantes aos experimentos anteriores. Quando combinadas com os extratores de textura, a distância Manhattan, além de possuir um custo computacional baixo, obteve acurácia semelhante as demais.

De acordo com os resultados detalhados por rota apresentados nas Tabelas 11 e 12, percebe-se que o ganho médio foi de 10,76% utilizando a nossa abordagem. Além disso, observa-se que em todas as combinações de extratores e classificadores o SDL aumenta ou mantém a acurácia dos classificadores. Destacam-se situações como a do SVM-Linear que estavam com 81% e 49% de acurácia nos Momentos Espaciais e Centrais, respectivamente e conseguiram valores acima de 90%. Dessa forma, a abordagem proposta mostra-se adequada e efetiva na tarefa de localização e navegação.

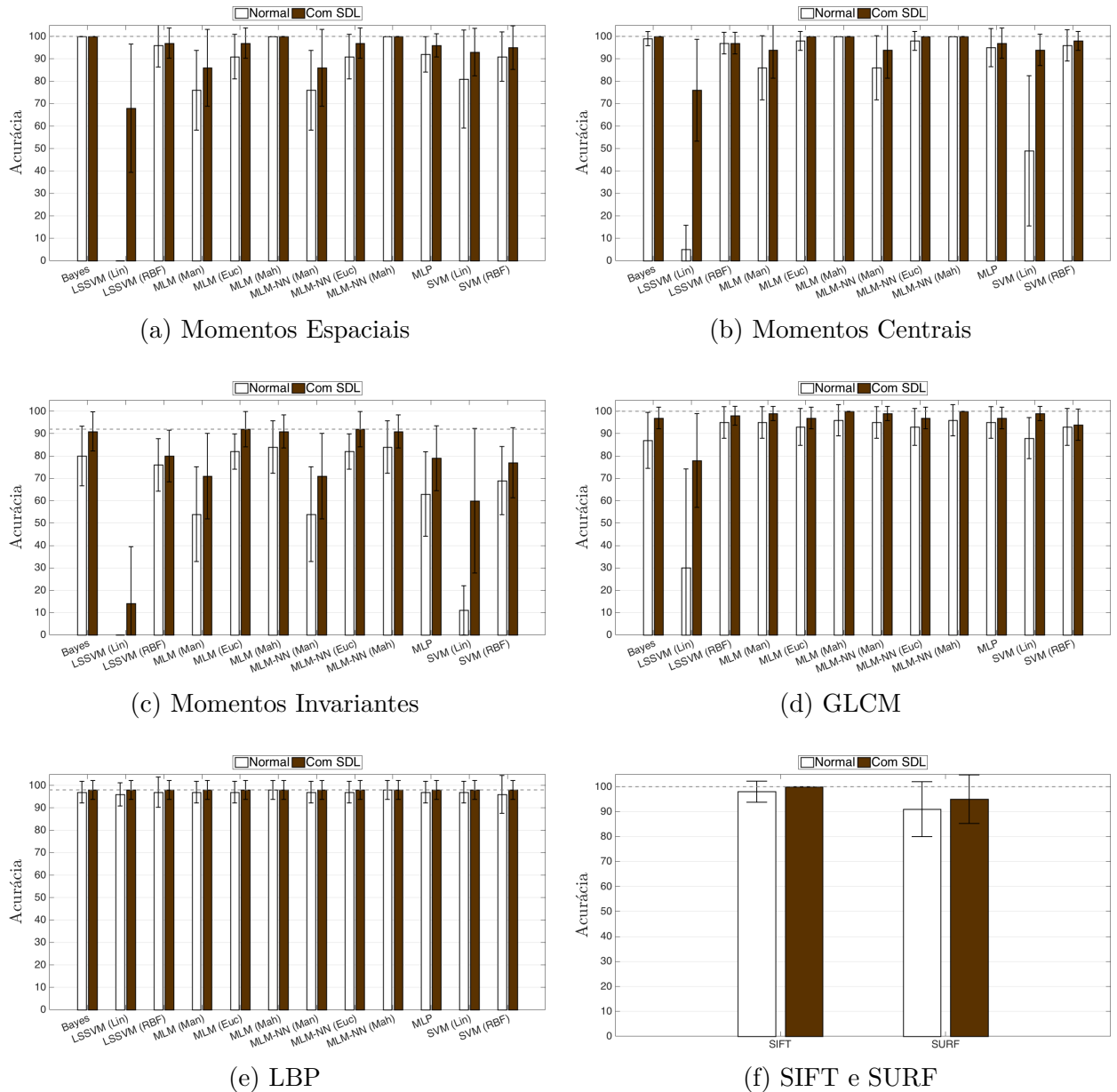


Figura 26 – Taxa de acerto média do teste de navegação, com e sem o Suporte de Decisão para Localização, obtidos pela combinação entre os classificadores e extratores utilizando a **GoPro**.

Conclui-se, portanto, que a combinação do classificador Bayesiano com os Momentos Espaciais é uma escolha adequada para a tarefa de navegação com a câmera de visão

convencional quando se leva em consideração o custo computacional e acurácia. Outra possibilidade é utilizar um classificador linear, SVM por exemplo, juntamente com o LBP e o SDL.

Tabela 11 – Taxa de acerto por rota das melhores configurações dos classificadores combinados com os extratores **sem o Suporte de Decisão para Localização** utilizando a **GoPro®**.

Extrator	Classificador	Configuração	Rota 1	Rota 2	Rota 3	Rota 4	Rota 5	Rota 6	Rota 7	Rota 8	Rota 9	Rota 10
Momentos Espaciais	Bayes	Normal	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	LSSVM	Linear	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0
		RBF	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	70,0±48,3	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	MLM	Manhattan	90,0±31,6	100,0±0,0	80,0±42,2	80,0±42,2	60,0±51,6	90,0±31,6	60,0±51,6	80,0±42,2	80,0±42,2	40,0±51,6
		Euclidiana	100,0±0,0	100,0±0,0	90,0±31,6	80,0±42,2	70,0±48,3	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	90,0±31,6
		Mahalanobis	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	MLM-NN	Manhattan	90,0±31,6	100,0±0,0	80,0±42,2	80,0±42,2	60,0±51,6	90,0±31,6	60,0±51,6	80,0±42,2	80,0±42,2	40,0±51,6
		Euclidiana	100,0±0,0	100,0±0,0	90,0±31,6	80,0±42,2	70,0±48,3	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	90,0±31,6
		Mahalanobis	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	MLP		100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	80,0±42,2	90,0±31,6	100,0±0,0	90,0±31,6	80,0±42,2	90,0±31,6
Momentos Centrais	LSSVM	Linear	30,0±48,3	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	20,0±42,2	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0
		RBF	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	90,0±31,6
	MLM	Manhattan	90,0±31,6	90,0±31,6	90,0±31,6	100,0±0,0	90,0±31,6	80,0±42,2	100,0±0,0	50,0±52,7	90,0±31,6	80,0±42,2
		Euclidiana	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0
		Mahalanobis	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	MLM-NN	Manhattan	90,0±31,6	90,0±31,6	90,0±31,6	100,0±0,0	90,0±31,6	80,0±42,2	100,0±0,0	50,0±52,7	90,0±31,6	80,0±42,2
		Euclidiana	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0
		Mahalanobis	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	MLP		100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	80,0±42,2	100,0±0,0	80,0±42,2
	SVM	Linear	100,0±0,0	90,0±31,6	10,0±31,6	80,0±42,2	40,0±51,6	70,0±48,3	30,0±48,3	40,0±51,6	20,0±42,2	10,0±31,6
Momentos Invariantes	LSSVM	Linear	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0
		RBF	70,0±48,3	100,0±0,0	70,0±48,3	90,0±31,6	70,0±48,3	80,0±42,2	80,0±42,2	70,0±48,3	60,0±51,6	70,0±48,3
	MLM	Manhattan	90,0±31,6	60,0±51,6	40,0±51,6	60,0±51,6	30,0±48,3	70,0±48,3	40,0±51,6	60,0±51,6	20,0±42,2	70,0±48,3
		Euclidiana	80,0±42,2	90,0±31,6	70,0±48,3	90,0±31,6	80,0±42,2	90,0±31,6	80,0±42,2	90,0±31,6	70,0±48,3	80,0±42,2
		Mahalanobis	100,0±0,0	100,0±0,0	70,0±48,3	100,0±0,0	80,0±42,2	80,0±42,2	80,0±42,2	80,0±42,2	70,0±48,3	80,0±42,2
	MLM-NN	Manhattan	90,0±31,6	60,0±51,6	40,0±51,6	60,0±51,6	30,0±48,3	70,0±48,3	40,0±51,6	60,0±51,6	20,0±42,2	70,0±48,3
		Euclidiana	80,0±42,2	90,0±31,6	70,0±48,3	90,0±31,6	80,0±42,2	90,0±31,6	80,0±42,2	90,0±31,6	70,0±48,3	80,0±42,2
		Mahalanobis	100,0±0,0	100,0±0,0	70,0±48,3	100,0±0,0	80,0±42,2	80,0±42,2	80,0±42,2	80,0±42,2	70,0±48,3	80,0±42,2
	MLP		90,0±31,6	90,0±31,6	70,0±48,3	50,0±52,7	60,0±51,6	80,0±42,2	40,0±51,6	60,0±51,6	40,0±51,6	50,0±52,7
	SVM	Linear	20,0±42,2	30,0±48,3	10,0±31,6	10,0±31,6	20,0±42,2	0,0±0,0	20,0±42,2	0,0±0,0	0,0±0,0	0,0±0,0
GLCM	LSSVM	Linear	100,0±0,0	70,0±48,3	80,0±42,2	80,0±42,2	80,0±42,2	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	70,0±48,3
		RBF	90,0±31,6	100,0±0,0	0,0±0,0	90,0±31,6	0,0±0,0	20,0±42,2	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0
	MLM	Manhattan	100,0±0,0	100,0±0,0	80,0±42,2	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6
		Euclidiana	100,0±0,0	80,0±42,2	80,0±42,2	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6
		Mahalanobis	100,0±0,0	100,0±0,0	80,0±42,2	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0
	MLM-NN	Manhattan	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	80,0±42,2	100,0±0,0
		Euclidiana	100,0±0,0	80,0±42,2	80,0±42,2	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6
		Mahalanobis	100,0±0,0	100,0±0,0	80,0±42,2	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0
	MLP		100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	80,0±42,2	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6
	SVM	Linear	100,0±0,0	90,0±31,6	80,0±42,2	100,0±0,0	90,0±31,6	100,0±0,0	80,0±42,2	80,0±42,2	100,0±0,0	80,0±42,2
LBP	LSSVM	Linear	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
		RBF	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	80,0±42,2	90,0±31,6	100,0±0,0	100,0±0,0
	MLM	Manhattan	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
		Euclidiana	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
		Mahalanobis	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
	MLM-NN	Manhattan	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
		Euclidiana	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
		Mahalanobis	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
	MLP		100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
	SVM	Linear	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
SIFT			100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0
SURF			100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	70,0±48,3	90,0±31,6	90,0±31,6	80,0±42,2	100,0±0,0	80,0±42,2

Tabela 12 – Taxa de acerto por rota das melhores configurações dos classificadores combinados com os extratores **com o Suporte de Decisão para Localização** utilizando a **GoPro®**.

Extrator	Classificador	Configuração	Rota 1	Rota 2	Rota 3	Rota 4	Rota 5	Rota 6	Rota 7	Rota 8	Rota 9	Rota 10
Momentos Espaciais	Bayes	Normal	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	LSSVM	Linear	90,0±31,6	100,0±0,0	80,0±42,1	100,0±0,0	70,0±48,3	50,0±52,7	90,0±31,6	40,0±51,6	20,0±42,1	40,0±51,6
		RBF	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	80,0±42,1	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	MLM	Manhattan	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	70,0±48,3	100,0±0,0	70,0±48,3	90,0±31,6	90,0±31,6	50,0±52,7
		Euclidiana	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	80,0±42,1	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
		Mahalanobis	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	MLM-NN	Manhattan	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	70,0±48,3	100,0±0,0	70,0±48,3	90,0±31,6	90,0±31,6	50,0±52,7
		Euclidiana	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	80,0±42,1	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
		Mahalanobis	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	MLP		100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6
SVM	Linear	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	80,0±42,1	70,0±48,3	
	RBF	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	70,0±48,3	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6	
Momentos Centrais	Bayes	Normal	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	LSSVM	Linear	100,0±0,0	90,0±31,6	60,0±51,6	100,0±0,0	90,0±31,6	40,0±51,6	100,0±0,0	70,0±48,3	50,0±52,7	60,0±51,6
		RBF	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	90,0±31,6
	MLM	Manhattan	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	60,0±51,6	100,0±0,0	90,0±31,6
		Euclidiana	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
		Mahalanobis	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	MLM-NN	Manhattan	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	60,0±51,6	100,0±0,0	90,0±31,6
		Euclidiana	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
		Mahalanobis	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	MLP		100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	80,0±42,1
SVM	Linear	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	90,0±31,6	90,0±31,6	80,0±42,1	
	RBF	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	90,0±31,6	
Momentos Invariantes	Bayes	Normal	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	80,0±42,1	100,0±0,0	90,0±31,6	80,0±42,1	100,0±0,0	80,0±42,1
	LSSVM	Linear	80,0±42,1	30,0±48,3	0,0±0,0	10,0±31,6	0,0±0,0	0,0±0,0	20,0±42,1	0,0±0,0	0,0±0,0	0,0±0,0
		RBF	70,0±48,3	100,0±0,0	80,0±42,1	100,0±0,0	70,0±48,3	80,0±42,1	80,0±42,1	70,0±48,3	80,0±42,1	70,0±48,3
	MLM	Manhattan	90,0±31,6	80,0±42,1	60,0±51,6	80,0±42,1	50,0±52,7	90,0±31,6	50,0±52,7	80,0±42,1	40,0±51,6	90,0±31,6
		Euclidiana	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	80,0±42,1	100,0±0,0	90,0±31,6	90,0±31,6	80,0±42,1	90,0±31,6
		Mahalanobis	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	80,0±42,1	90,0±31,6	90,0±31,6	90,0±31,6	80,0±42,1	90,0±31,6
	MLM-NN	Manhattan	90,0±31,6	80,0±42,1	60,0±51,6	80,0±42,1	50,0±52,7	90,0±31,6	50,0±52,7	80,0±42,1	40,0±51,6	90,0±31,6
		Euclidiana	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	80,0±42,1	100,0±0,0	90,0±31,6	90,0±31,6	80,0±42,1	90,0±31,6
		Mahalanobis	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	80,0±42,1	90,0±31,6	90,0±31,6	90,0±31,6	80,0±42,1	90,0±31,6
	MLP		100,0±0,0	100,0±0,0	80,0±42,1	60,0±51,6	80,0±42,1	90,0±31,6	60,0±51,6	80,0±42,1	70,0±48,3	70,0±48,3
SVM	Linear	100,0±0,0	70,0±48,3	60,0±51,6	70,0±48,3	100,0±0,0	20,0±42,1	90,0±31,6	30,0±48,3	10,0±31,6	50,0±52,7	
	RBF	60,0±51,6	90,0±31,6	90,0±31,6	100,0±0,0	70,0±48,3	90,0±31,6	80,0±42,1	50,0±52,7	70,0±48,3	70,0±48,3	
GLCM	Bayes	Normal	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6
	LSSVM	Linear	100,0±0,0	100,0±0,0	50,0±52,7	100,0±0,0	70,0±48,3	100,0±0,0	70,0±48,3	60,0±51,6	50,0±52,7	80,0±42,1
		RBF	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	MLM	Manhattan	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0
		Euclidiana	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6
		Mahalanobis	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	MLM-NN	Manhattan	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0
		Euclidiana	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6
		Mahalanobis	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	MLP		100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6
SVM	Linear	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	
	RBF	100,0±0,0	80,0±42,1	90,0±31,6	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	
LBP	Bayes	Normal	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
	LSSVM	Linear	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
		RBF	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0
	MLM	Manhattan	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
		Euclidiana	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
		Mahalanobis	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
	MLM-NN	Manhattan	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
		Euclidiana	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
		Mahalanobis	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
	MLP		100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0
SVM	Linear	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0	
	RBF	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0	
SIFT			100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
SURF			100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	70,0±48,3	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6

6.1.3 Custo de armazenamento (GoPro®)

A Tabela 13 mostra o custo de armazenamento, em kilobyte (KB), dos modelos dos classificadores por extrator de características utilizados no teste de navegação com a GoPro®.

De acordo com os resultados apresentados na Tabela 13, verifica-se que os classificadores que obtiveram o menor espaço de armazenamento em média foram a rede MLP (4,96 KB), SVM-Linear (89,18 KB), classificador Bayesiano (94,71 KB) e SVM-RBF (106,82 KB). Quanto aos extratores, os momentos consumiram menos espaço de armazenamento nos classificadores.

Tabela 13 – Espaço de armazenamento dos modelos embarcados para cada classificador com a **GoPro®**.

	ME (KB)	MC (KB)	MI (KB)	GLCM (KB)	LBP (KB)	MÉDIA (KB)
Bayes (Normal)	15,16	8,84	8,84	23,60	417,12	94,71
LSSVM (Linear)	875,95	664,31	664,31	1087,59	4332,75	1524,98
LSSVM (RBF)	877,71	666,07	666,07	1089,35	4334,51	1526,74
MLM (Manhattan)	2935,26	2921,20	2921,20	2949,32	3164,95	2978,39
MLM (Euclidiana)	2935,26	2921,20	2921,20	2949,32	3164,95	2978,39
MLM (Mahalanobis)	2938,15	2923,69	2923,69	2952,75	3194,25	2986,51
MLM-NN (Manhattan)	2935,26	2921,20	2921,20	2949,32	3164,95	2978,39
MLM-NN (Euclidiana)	2935,26	2921,20	2921,20	2949,32	3164,95	2978,39
MLM-NN (Mahalanobis)	2938,15	2923,69	2923,69	2952,75	3194,25	2986,51
MLP	2,66	4,23	5,12	6,3	6,49	4,96
SVM (Linear)	63,14	70,59	119,25	66,77	126,13	89,18
SVM (RBF)	53,09	101,11	74,60	68,77	236,55	106,82
MÉDIA	1625,42	1587,28	1589,20	1670,43	2375,15	
SIFT : 330332,60				SURF: 3641,64		

6.2 Resultados com a câmera omnidirecional

O segundo grupo de resultados é proveniente de imagens omnidirecionais. De maneira semelhante, os extratores são utilizado para gerar as entradas para os classificadores. A Tabela 14 exibe os valores médios e os desvios padrões da especificidade (Es), sensibilidade (Se), vapor preditivo positivo (VPP), F1-Score, média harmônica (MH) e acurácia (Ac) obtidas pela combinação dos classificadores com extratores em imagens omnidirecionais do ambiente sintético.

Com base nos resultados apresentados na Tabela 14, constata-se que o LBP foi o melhor extrator. De maneira semelhante, o outro extrator de textura, GLCM, combinado com os classificadores MLM e MLM-NN conseguiram bom desempenho chegando à 99% em todas as métricas. Por outro lado, os momentos obtiveram resultados um pouco inferior. Os Momentos Invariantes, por exemplo, só atingiram 81% de acurácia. Evidencia também o bom resultado do SIFT com valores acima de 98%, entretanto, como será mostrado adiante, essa técnica possui um elevado custo computacional. No geral, os classificadores não-lineares tiveram boa performance, destacando-se o classificador Bayesiano, MLM e MLM-NN com acurácia acima de 99,8%.

A etapa subsequente dos experimentos foi realizada com as imagens reais da câmera omnidirecional. Novamente, o processo de extração é realizado para uma posterior classificação. A Tabela 15 exibe os resultados obtidos pela combinação dos classificadores e extratores com imagens do ambiente real.

Observando a Tabela 15, verifica-se que os resultados das imagens reais foram relativamente menores que as do ambiente sintético. Esse comportamento é esperado devido os detalhes mais representativos do simulador, além disso, o ambiente controlado não sofre de interferências luminosas. Novamente, a combinação que obteve maior acurácia foi o LBP com MLM e MLM-NN (96,6%) usando a distância Mahalanobis. Destaca-se também os Momentos Espaciais e Centrais com acurácia acima de 94%. Por outro lado, os descritos de ponto-chave SIFT e SURF tiveram uma performance ruim, 60,95% e 47,35% de acurácia, respectivamente. Observa-se, portanto, que os detalhes fornecidos pela câmera omnidirecional não são tão representativos quanto os da imagem de campo de visão convencional. Os objetos nas imagens omnidirecionais são menores, deformados e, consecutivamente, possuem menos informações.

A próxima seção mostra os resultados dos métodos destacados em negrito na Tabela 15.

Tabela 14 – Especificidade (Es), sensibilidade (Se), valor preditivo positivo (VPP), F1-score, média harmônica (MH) and Acurácia (Ac) obtidos pelos extratores e classificadores no ambiente **sintético** usando as imagens da câmera **omnidirecional**.

Extrator	Classificador	Configuração	Es (%)	Se (%)	VPP (%)	F1-Score (%)	MH (%)	Ac (%)
Momentos Espaciais	Bayes	Normal	99,88±0,05	98,40±0,72	98,58±0,63	98,38±0,73	99,08±0,43	98,40±0,72
	LSSVM	Linear	89,40±0,72	34,67±1,78	34,01±4,73	27,14±2,35	35,94±2,68	34,67±1,78
		RBF	98,91±0,22	86,53±2,31	89,73±1,87	85,56±2,48	90,56±1,77	86,53±2,31
	MLM	Manhattan	98,93±0,15	86,67±1,75	87,56±1,96	86,63±1,78	91,77±1,38	86,67±1,75
		Euclidiana	99,23±0,12	90,07±1,39	90,92±1,43	90,05±1,37	94,02±0,86	90,07±1,39
	Mahalanobis	99,69±0,12	95,80±1,63	96,19±1,57	95,78±1,63	97,53±0,99	95,80±1,63	
	MLM-NN	Manhattan	98,95±0,17	86,87±1,99	87,73±2,18	86,85±2,03	91,91±1,53	86,87±1,99
		Euclidiana	99,23±0,11	90,07±1,35	90,89±1,41	90,04±1,34	94,01±0,85	90,07±1,35
		Mahalanobis	99,70±0,11	95,93±1,46	96,31±1,40	95,91±1,45	97,62±0,87	95,93±1,46
	MLP		98,98±0,27	87,20±2,99	86,63±4,79	86,33±4,13	91,11±3,22	87,20±2,99
SVM	Linear	95,53±1,01	59,40±5,99	55,01±6,65	51,71±7,38	62,12±7,43	59,40±5,99	
	RBF	99,33±0,22	91,40±2,58	92,97±1,91	91,17±2,62	94,60±1,70	91,40±2,58	
Momentos Centrais	Bayes	Normal	99,73±0,09	96,27±1,14	96,55±1,03	96,15±1,21	97,74±0,75	96,27±1,14
	LSSVM	Linear	87,32±1,15	30,20±2,44	29,70±4,02	23,64±3,28	31,49±4,10	30,20±2,44
		RBF	99,44±0,18	92,60±2,23	93,66±1,82	92,43±2,31	95,38±1,49	92,60±2,23
	MLM	Manhattan	99,39±0,21	92,07±2,50	92,51±2,50	91,95±2,62	95,21±1,65	92,07±2,50
		Euclidiana	99,63±0,12	94,93±1,64	95,25±1,67	94,78±1,65	96,81±1,07	94,93±1,64
	Mahalanobis	99,35±0,20	91,53±2,44	92,19±2,31	91,41±2,46	94,79±1,57	91,53±2,44	
	MLM-NN	Manhattan	99,39±0,19	92,07±2,32	92,51±2,30	92,00±2,39	95,25±1,50	92,07±2,32
		Euclidiana	99,63±0,12	94,93±1,64	95,25±1,67	94,78±1,65	96,81±1,07	94,93±1,64
		Mahalanobis	99,35±0,20	91,53±2,44	92,19±2,31	91,41±2,46	94,79±1,57	91,53±2,44
	MLP		99,49±0,16	93,27±2,05	93,29±3,40	92,83±2,75	95,50±2,04	93,27±2,05
SVM	Linear	96,78±0,45	67,20±3,23	68,03±4,13	63,51±3,88	72,72±3,23	67,20±3,23	
	RBF	99,42±0,23	92,33±2,83	93,75±2,17	92,31±2,92	95,22±2,02	92,33±2,83	
Momentos Invariantes	Bayes	Normal	95,78±0,30	57,20±2,01	50,96±2,67	51,45±1,34	58,54±1,36	57,20±2,01
	LSSVM	Linear	88,15±0,37	19,93±0,66	8,98±4,90	9,33±1,09	12,12±1,14	19,93±0,66
		RBF	96,59±0,33	65,20±2,20	57,47±1,47	58,44±2,25	66,65±1,78	65,20±2,20
	MLM	Manhattan	97,86±0,17	76,07±1,49	76,94±1,53	75,81±1,56	83,94±1,27	76,07±1,49
		Euclidiana	97,74±0,35	75,00±3,05	75,97±2,99	74,76±2,90	82,78±2,39	75,00±3,05
	Mahalanobis	98,46±0,24	81,80±2,31	82,52±2,42	81,50±2,45	88,17±1,73	81,80±2,31	
	MLM-NN	Manhattan	97,87±0,12	76,13±1,03	77,20±0,99	75,95±0,92	84,00±0,72	76,13±1,03
		Euclidiana	97,77±0,35	75,27±3,09	76,19±3,00	75,01±2,86	83,00±2,31	75,27±3,09
		Mahalanobis	98,46±0,24	81,73±2,31	82,43±2,45	81,44±2,47	88,12±1,74	81,73±2,31
	MLP		97,58±0,86	74,07±7,50	74,19±9,60	72,27±8,58	80,14±7,25	74,07±7,50
SVM	Linear	93,66±0,83	49,60±3,43	43,32±4,29	41,72±4,13	51,74±4,31	49,60±3,43	
	RBF	96,95±0,31	68,47±2,18	69,52±5,36	63,47±1,95	72,31±1,43	68,47±2,18	
GLCM	Bayes	Normal	98,49±0,25	82,07±2,48	85,98±2,49	80,54±2,61	86,82±1,86	82,07±2,48
	LSSVM	Linear	98,15±0,34	78,73±3,05	82,13±2,93	75,93±2,85	82,66±2,41	78,73±3,05
		RBF	99,85±0,05	97,87±0,69	98,04±0,64	97,86±0,69	98,79±0,39	97,87±0,69
	MLM	Manhattan	99,69±0,14	95,80±1,83	96,17±1,75	95,78±1,87	97,58±1,11	95,80±1,83
		Euclidiana	99,87±0,10	98,27±1,41	98,38±1,32	98,24±1,47	98,98±0,91	98,27±1,41
	Mahalanobis	99,96±0,03	99,47±0,42	99,52±0,38	99,47±0,42	99,70±0,24	99,47±0,42	
	MLM-NN	Manhattan	99,69±0,14	95,80±1,83	96,17±1,75	95,78±1,87	97,58±1,11	95,80±1,83
		Euclidiana	99,87±0,10	98,27±1,41	98,38±1,32	98,24±1,47	98,98±0,91	98,27±1,41
		Mahalanobis	99,96±0,03	99,47±0,42	99,52±0,38	99,47±0,42	99,70±0,24	99,47±0,42
	MLP		99,71±0,27	96,20±3,49	94,79±5,94	95,31±4,83	96,88±3,45	96,20±3,49
SVM	Linear	99,77±0,12	96,80±1,53	97,21±1,29	96,80±1,51	98,14±0,92	96,80±1,53	
	RBF	99,79±0,10	97,13±1,34	97,40±1,19	97,11±1,35	98,35±0,79	97,13±1,34	
LBP	Bayes	Normal	99,99±0,02	99,87±0,28	99,88±0,26	99,87±0,28	99,93±0,16	99,87±0,28
	LSSVM	Linear	99,77±0,10	96,93±1,34	97,25±1,23	96,91±1,36	98,23±0,79	96,93±1,34
		RBF	99,96±0,04	99,47±0,61	99,52±0,55	99,46±0,62	99,70±0,35	99,47±0,61
	MLM	Manhattan	99,97±0,03	99,53±0,45	99,59±0,39	99,53±0,45	99,73±0,26	99,53±0,45
		Euclidiana	99,98±0,03	99,67±0,35	99,70±0,32	99,67±0,35	99,81±0,20	99,67±0,35
	Mahalanobis	100,00±0,00	100,00±0,00	100,00±0,00	100,00±0,00	100,00±0,00	100,00±0,00	
	MLM-NN	Manhattan	99,97±0,03	99,53±0,45	99,59±0,39	99,53±0,45	99,73±0,26	99,53±0,45
		Euclidiana	99,98±0,03	99,67±0,35	99,70±0,32	99,67±0,35	99,81±0,20	99,67±0,35
		Mahalanobis	100,00±0,00	100,00±0,00	100,00±0,00	100,00±0,00	100,00±0,00	100,00±0,00
	MLP		99,99±0,03	99,80±0,45	99,83±0,38	99,80±0,45	99,88±0,26	99,80±0,45
SVM	Linear	99,95±0,06	99,33±0,89	99,42±0,79	99,33±0,90	99,61±0,52	99,33±0,89	
	RBF	99,97±0,03	99,60±0,47	99,65±0,40	99,60±0,46	99,78±0,26	99,60±0,47	
SIFT			99,92±0,06	98,83±0,89	98,95±0,77	98,81±0,92	99,32±0,54	98,83±0,89
SURF			99,28±0,16	90,69±1,97	91,73±1,60	90,35±2,20	94,02±1,58	90,69±1,97

Tabela 15 – Especificidade (Es), sensibilidade (Se), valor preditivo positivo (VPP), F1-score, média harmônica (MH) and Acurácia (Ac) obtidos pelos extratores e classificadores no ambiente **real** usando as imagens da câmara **omnidirecional**.

Extrator	Classificador	Configuração	Es (%)	Se (%)	VPP (%)	F1-Score (%)	MH (%)	Ac (%)
Momentos Espaciais	Bayes	Normal	99,74±0,13	96,47±1,69	96,77±1,59	96,41±1,72	97,93±1,03	96,47±1,69
	LSSVM	Linear	90,07±0,59	36,60±1,68	31,37±4,13	26,18±2,27	35,82±2,53	36,60±1,68
		RBF	99,35±0,21	91,60±2,50	92,88±1,98	91,40±2,53	94,79±1,62	91,60±2,50
	MLM	Manhattan	98,66±0,27	83,93±2,78	84,76±2,61	83,67±2,92	89,84±2,00	83,93±2,78
		Euclidiana	99,31±0,17	91,07±2,02	91,81±2,02	91,01±2,02	94,65±1,25	91,07±2,02
	Mahalanobis	99,57±0,10	94,27±1,23	94,75±1,24	94,25±1,23	96,64±0,74	94,27±1,23	
		Manhattan	98,66±0,27	83,87±2,75	84,69±2,58	83,60±2,90	89,80±1,99	83,87±2,75
	Euclidiana	99,31±0,17	91,07±2,02	91,81±2,02	91,01±2,02	94,65±1,25	91,07±2,02	
		Mahalanobis	99,57±0,10	94,27±1,23	94,75±1,24	94,25±1,23	96,64±0,74	94,27±1,23
	MLP		98,51±0,39	82,47±3,84	78,30±7,17	79,44±5,48	85,97±4,38	82,47±3,84
Momentos Centrais	Bayes	Normal	99,34±0,17	91,40±2,07	92,24±1,81	91,39±2,05	94,90±1,29	91,40±2,07
	LSSVM	Linear	89,61±0,66	34,53±1,72	26,42±3,32	23,71±1,53	32,16±1,96	34,53±1,72
		RBF	99,49±0,19	93,20±2,37	94,30±1,70	93,16±2,39	95,87±1,58	93,20±2,37
	MLM	Manhattan	98,87±0,25	86,07±2,69	87,04±2,65	85,83±2,83	91,24±2,00	86,07±2,69
		Euclidiana	99,61±0,09	94,80±1,12	95,23±1,21	94,77±1,15	96,96±0,69	94,80±1,12
	Mahalanobis	99,57±0,15	94,27±1,94	94,77±1,88	94,17±1,98	96,59±1,20	94,27±1,94	
		Manhattan	98,88±0,26	86,27±2,73	87,21±2,69	86,02±2,86	91,36±2,01	86,27±2,73
	Euclidiana	99,61±0,09	94,80±1,12	95,23±1,21	94,77±1,15	96,96±0,69	94,80±1,12	
		Mahalanobis	99,57±0,15	94,27±1,94	94,77±1,88	94,17±1,98	96,59±1,20	94,27±1,94
	MLP		99,05±0,60	88,40±6,43	85,50±10,43	86,25±8,79	90,70±6,80	88,40±6,43
Momentos Invariantes	Bayes	Normal	99,00±0,20	87,47±2,31	88,62±2,36	87,11±2,31	91,94±1,53	87,47±2,31
	LSSVM	Linear	91,91±0,56	41,20±2,13	40,31±6,42	31,40±2,36	41,35±2,57	41,20±2,13
		RBF	98,84±0,18	85,67±1,99	89,12±1,06	85,69±1,87	90,91±1,49	85,67±1,99
	MLM	Manhattan	97,67±0,45	74,80±3,69	75,95±4,00	74,60±3,88	83,61±2,82	74,80±3,69
		Euclidiana	99,19±0,24	89,73±2,69	90,47±2,62	89,54±2,87	93,76±1,89	89,73±2,69
	Mahalanobis	99,25±0,22	90,40±2,56	91,12±2,27	90,36±2,58	94,34±1,60	90,40±2,56	
		Manhattan	97,68±0,44	74,87±3,61	76,05±3,88	74,67±3,79	83,66±2,76	74,87±3,61
	Euclidiana	99,19±0,24	89,73±2,69	90,47±2,62	89,54±2,87	93,76±1,89	89,73±2,69	
		Mahalanobis	99,25±0,22	90,40±2,56	91,12±2,27	90,36±2,58	94,34±1,60	90,40±2,56
	MLP		98,61±0,43	83,60±4,53	81,91±7,06	81,86±5,91	88,12±4,46	83,60±4,53
GLCM	Bayes	Normal	97,28±0,34	71,47±2,63	73,99±4,90	69,50±3,18	78,71±2,58	71,47±2,63
	LSSVM	Linear	93,96±0,56	50,87±2,63	51,51±3,75	44,35±3,29	55,16±3,22	50,87±2,63
		RBF	98,87±0,26	86,07±2,87	88,44±2,46	85,96±2,95	91,21±2,07	86,07±2,87
	MLM	Manhattan	98,31±0,26	80,47±2,52	81,69±2,38	80,29±2,53	87,62±1,74	80,47±2,52
		Euclidiana	98,78±0,22	85,13±2,31	86,68±2,26	85,13±2,34	90,82±1,54	85,13±2,31
	Mahalanobis	99,36±0,21	91,73±2,61	92,46±2,29	91,75±2,57	95,12±1,61	91,73±2,61	
		Manhattan	98,33±0,26	80,60±2,54	81,84±2,44	80,42±2,59	87,70±1,78	80,60±2,54
	Euclidiana	98,78±0,22	85,13±2,31	86,68±2,26	85,13±2,34	90,82±1,54	85,13±2,31	
		Mahalanobis	99,36±0,21	91,73±2,61	92,46±2,29	91,75±2,57	95,12±1,61	91,73±2,61
	MLP		98,52±0,69	82,93±6,26	81,92±9,21	81,56±7,91	88,02±5,99	82,93±6,26
LBP	Bayes	Normal	99,27±0,14	90,60±1,68	91,23±1,49	90,49±1,72	94,29±1,14	90,60±1,68
	LSSVM	Linear	97,67±0,35	74,33±3,05	78,29±2,83	71,89±3,34	79,63±2,70	74,33±3,05
		RBF	99,65±0,12	95,27±1,55	95,81±1,26	95,24±1,58	97,24±0,98	95,27±1,55
	MLM	Manhattan	99,57±0,06	94,20±0,77	94,66±0,85	94,19±0,76	96,63±0,45	94,20±0,77
		Euclidiana	99,63±0,08	95,07±1,00	95,44±0,91	95,03±1,01	97,15±0,61	95,07±1,00
	Mahalanobis	99,75±0,10	96,67±1,37	96,86±1,33	96,64±1,38	98,09±0,79	96,67±1,37	
		Manhattan	99,57±0,06	94,20±0,77	94,66±0,85	94,19±0,76	96,63±0,45	94,20±0,77
	Euclidiana	99,63±0,08	95,07±1,00	95,44±0,91	95,03±1,01	97,15±0,61	95,07±1,00	
		Mahalanobis	99,75±0,10	96,67±1,37	96,86±1,33	96,64±1,38	98,09±0,79	96,67±1,37
	MLP		99,54±0,22	93,87±2,79	93,39±4,45	93,35±3,74	95,80±2,83	93,87±2,79
SIFT	Bayes	Normal	99,27±0,14	90,60±1,68	91,23±1,49	90,49±1,72	94,29±1,14	90,60±1,68
	LSSVM	Linear	97,67±0,35	74,33±3,05	78,29±2,83	71,89±3,34	79,63±2,70	74,33±3,05
		RBF	99,65±0,12	95,27±1,55	95,81±1,26	95,24±1,58	97,24±0,98	95,27±1,55
	MLM	Manhattan	99,57±0,06	94,20±0,77	94,66±0,85	94,19±0,76	96,63±0,45	94,20±0,77
		Euclidiana	99,63±0,08	95,07±1,00	95,44±0,91	95,03±1,01	97,15±0,61	95,07±1,00
	Mahalanobis	99,75±0,10	96,67±1,37	96,86±1,33	96,64±1,38	98,09±0,79	96,67±1,37	
		Manhattan	99,57±0,06	94,20±0,77	94,66±0,85	94,19±0,76	96,63±0,45	94,20±0,77
	Euclidiana	99,63±0,08	95,07±1,00	95,44±0,91	95,03±1,01	97,15±0,61	95,07±1,00	
		Mahalanobis	99,75±0,10	96,67±1,37	96,86±1,33	96,64±1,38	98,09±0,79	96,67±1,37
	MLP		99,54±0,22	93,87±2,79	93,39±4,45	93,35±3,74	95,80±2,83	93,87±2,79
SURF	Bayes	Normal	99,27±0,14	90,60±1,68	91,23±1,49	90,49±1,72	94,29±1,14	90,60±1,68
	LSSVM	Linear	97,67±0,35	74,33±3,05	78,29±2,83	71,89±3,34	79,63±2,70	74,33±3,05
		RBF	99,65±0,12	95,27±1,55	95,81±1,26	95,24±1,58	97,24±0,98	95,27±1,55
	MLM	Manhattan	99,57±0,06	94,20±0,77	94,66±0,85	94,19±0,76	96,63±0,45	94,20±0,77
		Euclidiana	99,63±0,08	95,07±1,00	95,44±0,91	95,03±1,01	97,15±0,61	95,07±1,00
	Mahalanobis	99,75±0,10	96,67±1,37	96,86±1,33	96,64±1,38	98,09±0,79	96,67±1,37	
		Manhattan	99,57±0,06	94,20±0,77	94,66±0,85	94,19±0,76	96,63±0,45	94,20±0,77
	Euclidiana	99,63±0,08	95,07±1,00	95,44±0,91	95,03±1,01	97,15±0,61	95,07±1,00	
		Mahalanobis	99,75±0,10	96,67±1,37	96,86±1,33	96,64±1,38	98,09±0,79	96,67±1,37
	MLP		99,54±0,22	93,87±2,79	93,39±4,45	93,35±3,74	95,80±2,83	93,87±2,79
SURF	Bayes	Normal	99,27±0,14	90,60±1,68	91,23±1,49	90,49±1,72	94,29±1,14	90,60±1,68
	LSSVM	Linear	97,67±0,35	74,33±3,05	78,29±2,83	71,89±3,34	79,63±2,70	74,33±3,05
		RBF	99,65±0,12	95,27±1,55	95,81±1,26	95,24±1,58	97,24±0,98	95,27±1,55
	MLM	Manhattan	99,57±0,06	94,20±0,77	94,66±0,85	94,19±0,76	96,63±0,45	94,20±0,77
		Euclidiana	99,63±0,08	95,07±1,00	95,44±0,91	95,03±1,01	97,15±0,61	95,07±1,00
	Mahalanobis	99,75±0,10	96,67±1,37	96,86±1,33	96,64±1,38	98,09±0,79	96,67±1,37	
		Manhattan	99,57±0,06	94,20±0,77	94,66±0,85	94,19±0,76	96,63±0,45	94,20±0,77
	Euclidiana	99,63±0,08	95,07±1,00	95,44±0,91	95,03±1,01	97,15±0,61	95,07±1,00	
		Mahalanobis	99,75±0,10	96,67±1,37	96,86±1,33	96,64±1,38	98,09±0,79	96,67±1,37
	MLP		99,54±0,22	93,87±2,79	93,39±4,45	93,35±3,74	95,80±2,83	93,87±2,79
SURF	Bayes	Normal	99,27±0,14	90,60±1,68	91,23±1,49	90,49±1,72	94,29±1,14	90,60±1,68
	LSSVM	Linear	97,67±0,35	74,33±3,05	78,29±2,83	71,89±3,34	79,63±2,70	74,33±3,05
		RBF	99,65±0,12	95,27±1,55	95,81±1,26	95,24±1,58	97,24±0,98	95,27±1,55
	MLM	Manhattan	99,57±0,06	94,20±0,77	94,66±0,85	94,19±0,76	96,63±0,45	94,20±0,77
		Euclidiana	99,63±0,08	95,07±1,00	95,44±0,91	95,03±1,01	97,15±0,61	95,07±1,00
	Mahalanobis	99,75±0,10	96,67±1,37	96,86±1,33	96,64±1,38	98,09±0,79	96,67±1,37	
		Manhattan	99,57±0,06	94,20±0,77	94,66±0,85	94,19±0,76	96,63±0,45	94,20±0,77
	Euclidiana	99,63±0,08	95,07±1,00	95,44±0,91	95,03±1,01	97,15±0,61	95,07±1,00	
		Mahalanobis	99,75±0,10	96,67±1,37	96,86±1,33	96,64±1,38	98,09±0,79	96,67±1,37
	MLP		99,54±0,22	93,87±2,79	93,39±4,45	93,35±3,74	95,80±2,83	93,87±2,79
SURF	Bayes	Normal	99,27±0,14	90,60±1,68	91,23±1,49	90,49±1,72	94,29±1,14	90,60±1,68
	LSSVM	Linear	97,67±0,35	74,33±3,05	78,29±2,83	71,89±3,34	79,63±2,70	74,33±3,05
		RBF	99,65±0,12	95,27±1,55	95,81±1,26	95,24±1,58	97,24±0,98	95,27±1,55
	MLM	Manhattan	99,57±0,06	94,20±0,77	94,66±0,85	94,19±0,76	96,63±0,45	94,20±0,77
		Euclidiana	99,63±0,08	95,07±1,00	95,44±0,91	95,03±1,01	97,15±0,61	95,07±1,00
	Mahalanobis	99,75±0,10	96,67±1,37	96,86±1,33	96,64±1,38	98,09±0,79	96,67±1,37	
		Manhattan	99,57±0,06	94,20±0,77	94,66±0,85	94,19±0,76	96,63±0,45	94,20±0,77
	Euclidiana	99,63±0,08	95,07±1,00	95,44±0,91	95,03±1,01	97,15±0,61	95,07±1,00	
		Mahalanobis	99,75±0,10	96,67±1,37	96,86±1,33	96,64±1,38	98,09±0,79	96,67±1,37
	MLP		99,54±0,22	93,87±2,79	93,39±4,45	93,35±3,74	95,80±2,83	93,87±2,79
SURF	Bayes	Normal	99,27±0,14	90,60±1,68	91,23±1,49	90,49±1,72	94,29±1,14	90,60±1,68
	LSSVM	Linear	97,67±0,35	74,33±3,05	78,29±2,83	71,89±3,34	79,63±2,70	74,33±3,05
		RBF	99,65±0,12	95,27±1,55	95,81±1,26	95,24±1,58	97,24±0,98	95,27±1,55
	MLM	Manhattan	99,57±0,06	94,20±0,77	94,66±0,85	94,19±0,76	96,63±0,45	94,20±0,77
		Euclidiana	99,63±0,08	95,07±1,00	95,44±0,91	95,03±1,01	97,15±0,61	95,07±1,00
	Mahalanobis	99,75±0,10	96,67±1,37	96,86±1,33	96,64±1,38	98,09±0,79	96,67±1,37	
		Manhattan	99,57±0,06	94,20±0,77	94,66±0,85	94,19±0,76	96,63±0,45	94,20±0,77
	Euclidiana	99,63±0,08	95,07±1,00	95,44±0,91	95,03±1,01	97,15±0,61	95,07±1,00	
		Mahalanobis	99,75±0,10	96,67±1,37	96,86±1,33	96,64±1,38	98,09±0,79	96,67±1,37
	MLP		99,54±0,22	93,87±2,79	93,39±4,45	93,35±3,74	95,80±2,83	93,87±2,79
SURF	Bayes	Normal	99,27±0,14	90,60±1,68	91,23±1,49	90,49±1,72	94,29±1,14	90,60±1,68
	LSSVM	Linear	97,67±0,35	74,33±3,05	78,29±2,83	71,89±3,34	79,63±2,70	74,33±3,05
		RBF	99,65±0,12	95,27±1,55	95,81±1,26	95,24±1,58	97,24±0,98	95,27±1,55
	MLM	Manhattan	99,57±0,06	94,20±0,77	94,66±0,85	94,19±0,76	96,63±0,45	94,20±0,77
		Euclidiana	99,63±0,08	95,07±1,00	95,44±0,91	95,03±1,01	97,15±0,61	95,07±1,00
	Mahalanobis	99,75±0,10	96,67±1,37	96,86±1,33	96,64±1,38	98,		

6.2.1 Análise do tempo de processamento (Omnidirecional)

A Tabela 16 exibe a acurácia, tempo de predição de um padrão e de extração, além do tempo de treino; propriedades importantes na tarefa de navegação e localização analisada nessa dissertação.

De maneira semelhante aos resultados apresentados com as imagens da GoPro®, o custo computacional durante o treinamento e teste dos classificadores é análogo entre os extratores. A rede MLP, novamente, obteve o maior tempo de treinamento, seguido do SVM e LSSVM. Verifica-se também um aumento no tempo de treinamento com as imagens omnidirecionais em relação as da GoPro®. Esse comportamento é resultante das características obtidas pelos extratores sobre as imagens omnidirecionais que dificultou a convergência de alguns classificadores na etapa de treino. De modo geral, o SIFT e SURF continuaram com o pior tempo de treino e teste. O classificador Bayesiano e o SVM destacam-se na etapa de teste com os menores tempos.

Quanto aos extratores, verifica-se que o mais rápido é o Momento Espacial com 0,178 segundos. Destaca-se também o LBP por ter conseguido os maiores valores de acurácia com os classificadores, ao mesmo tempo que obteve um tempo de extração pequeno. Por outro lado, o SIFT e SURF foram os mais lentos. Por fim, observa-se que, de maneira semelhante aos resultados obtidos com a câmera GoPro®, a distância Manhattan na MLM reduz o tempo de treinamento e teste enquanto fornece um valor equivalente de acurácia em relação a distância Mahalanobis.

6.2.2 Teste de Navegação (Omnidirecional)

De maneira similar aos experimentos da câmera GoPro®, após um estudo preliminar com as imagens capturadas no ambiente sintético e real é realizado o teste de navegação com as melhores configurações dos métodos.

A Figura 27 mostra os resultados gerais do teste de navegação utilizando a câmera omnidirecional, com e sem o Suporte de Decisão para Localização, obtidos pela combinação entre os classificadores e extratores, ademais, exibe o desempenho do SIFT e SURF. Os resultados detalhados são mostrados nas Tabelas 17 e 18. Os valores dos hiperparâmetros para os classificadores são detalhados no Apêndice A.

De acordo com os resultados apresentados na Figura 27 e nas Tabelas 17 e 18, constata-se que a melhor combinação foi o extrator LBP com os classificadores MLM e MLM-NN usando a distância Mahalanobis com 86% de acurácia. O classificador Bayesiano com os Momentos Espaciais podem, também, ser destacados com taxa de acerto de 82%. Em contrapartida, os classificadores lineares ficaram com as piores acurácias, ou seja, nenhum extrator conseguiu tornar o problema linearmente separável com as imagens provenientes da câmera omnidirecional. Fica evidente que os resultados obtidos pela

Tabela 16 – Acurácia (Ac), tempo de treino, tempo de teste e de extração com todos extratores combinados com as melhores configurações dos classificadores utilizando as imagens da câmera omnidirecional.

Momentos Espaciais				
Classificador	Ac (%)	Tempo de Treino (s)	Tempo de Classificação (μ s)	Tempo de Extração (s)
Bayes	96,47 $\pm$ 1,69	0,004 $\pm$ 0,0	59,5 $\pm$ 70,6	0,178 $\pm$ 0,04
MLM (Mahalanobis)	94,27 $\pm$ 1,23	0,175 $\pm$ 0,1	3,43e4 $\pm$ 1,41e2	
MLM-NN (Mahalanobis)	94,27 $\pm$ 1,23	0,145 $\pm$ 0,0	165,9 $\pm$ 15,1	
LSSVM (RBF)	91,60 $\pm$ 2,50	0,620 $\pm$ 0,0	3,43e2 $\pm$ 7,9e2	
SVM (RBF)	90,13 $\pm$ 2,79	5,264 $\pm$ 0,4	8,28e1 $\pm$ 4e1	
MLP	82,47 $\pm$ 3,84	26,493 $\pm$ 6,7	290,7 $\pm$ 9,3	
Momentos Centrais				
Classificador	Ac (%)	Tempo de Treino (s)	Tempo de Classificação (μ s)	Tempo de Extração (s)
MLM (Euclidiana)	94,80 $\pm$ 1,12	0,072 $\pm$ 0,0	6,39e4 $\pm$ 2,57e3	0,183 $\pm$ 0,04
MLM-NN (Euclidiana)	94,80 $\pm$ 1,12	0,070 $\pm$ 0,0	119,7 $\pm$ 5,0	
LSSVM (RBF)	93,20 $\pm$ 2,37	0,674 $\pm$ 0,0	3,20e2 $\pm$ 5,4e1	
SVM (RBF)	92,07 $\pm$ 2,75	0,835 $\pm$ 0,0	9,72e1 $\pm$ 1,2e1	
Bayes	91,40 $\pm$ 2,07	0,004 $\pm$ 0,0	75,4 $\pm$ 124,2	
MLP	88,40 $\pm$ 6,43	25,440 $\pm$ 5,0	378,7 $\pm$ 25,9	
Momentos Invariantes				
Classificador	Ac (%)	Tempo de Treino (s)	Tempo de Classificação (μ s)	Tempo de Extração (s)
MLM (Mahalanobis)	90,40 $\pm$ 2,56	0,136 $\pm$ 0,0	3,53e4 $\pm$ 7,58e2	0,188 $\pm$ 0,04
MLM-NN (Mahalanobis)	90,40 $\pm$ 2,56	0,133 $\pm$ 0,0	160,3 $\pm$ 15,2	
MLM-NN (Euclidiana)	89,73 $\pm$ 2,69	0,074 $\pm$ 0,0	126,4 $\pm$ 5,3	
Bayes	87,47 $\pm$ 2,31	0,003 $\pm$ 0,0	36,4 $\pm$ 0,7	
LSSVM (RBF)	85,67 $\pm$ 2,31	0,594 $\pm$ 0,0	3,74e2 $\pm$ 9,4e1	
SVM (RBF)	85,13 $\pm$ 2,74	4,080 $\pm$ 0,4	8,71e1 $\pm$ 4,2e1	
MLP	83,60 $\pm$ 4,53	13,637 $\pm$ 2,5	291,6 $\pm$ 10,3	
GLCM				
Classificador	Ac (%)	Tempo de Treino (s)	Tempo de Classificação (μ s)	Tempo de Extração (s)
MLM (Mahalanobis)	91,73 $\pm$ 2,61	0,244 $\pm$ 0,1	3,45e4 $\pm$ 1,75e2	0,535 $\pm$ 0,07
MLM-NN (Mahalanobis)	91,73 $\pm$ 2,61	0,209 $\pm$ 0,0	274,7 $\pm$ 5,5	
LSSVM (RBF)	86,07 $\pm$ 2,87	0,599 $\pm$ 0,0	2,97e2 $\pm$ 1,12e1	
SVM (RBF)	85,47 $\pm$ 1,69	16,543 $\pm$ 1,7	8,77e1 $\pm$ 3,4e1	
MLP	82,93 $\pm$ 6,26	32,539 $\pm$ 11,8	303,8 $\pm$ 8,3	
Bayes	71,47 $\pm$ 2,63	0,004 $\pm$ 0,0	158,2 $\pm$ 336,2	
LBP				
Classificador	Ac (%)	Tempo de Treino (s)	Tempo de Classificação (μ s)	Tempo de Extração (s)
MLM (Mahalanobis)	96,67 $\pm$ 1,37	0,835 $\pm$ 0,1	3,53e4 $\pm$ 5,8e2	0,337 $\pm$ 0,03
MLM-NN (Mahalanobis)	96,67 $\pm$ 1,37	0,760 $\pm$ 0,0	1,21e3 $\pm$ 23,2	
LSSVM (RBF)	95,27 $\pm$ 1,55	0,610 $\pm$ 0,0	3,32e2 $\pm$ 8,7e1	
SVM (RBF)	94,53 $\pm$ 1,29	0,797 $\pm$ 0,0	9,83e1 $\pm$ 3,4e1	
MLM-NN (Manhattan)	94,20 $\pm$ 0,77	0,095 $\pm$ 0,0	256,9 $\pm$ 307,5	
MLP	93,87 $\pm$ 2,79	44,683 $\pm$ 30,0	916,9 $\pm$ 1038,6	
Bayes	90,6 $\pm$ 1,7	0,049 $\pm$ 0,1	291,0 $\pm$ 640,5	
Descritor	Ac (%)	Tempo de Treino (s)	Tempo de Classificação (μ s)	Tempo de Extração (s)
SIFT	60,95 $\pm$ 0,58	189,134 $\pm$ 3,0	3,66e7 $\pm$ 3,04e6	12,353 $\pm$ 0,03
SURF	47,35 $\pm$ 3,35	30,242 $\pm$ 0,4	2,06e6 $\pm$ 7e4	4,919 $\pm$ 0,02

câmera GoPro[®] foram superiores aos da câmera omnidirecional.

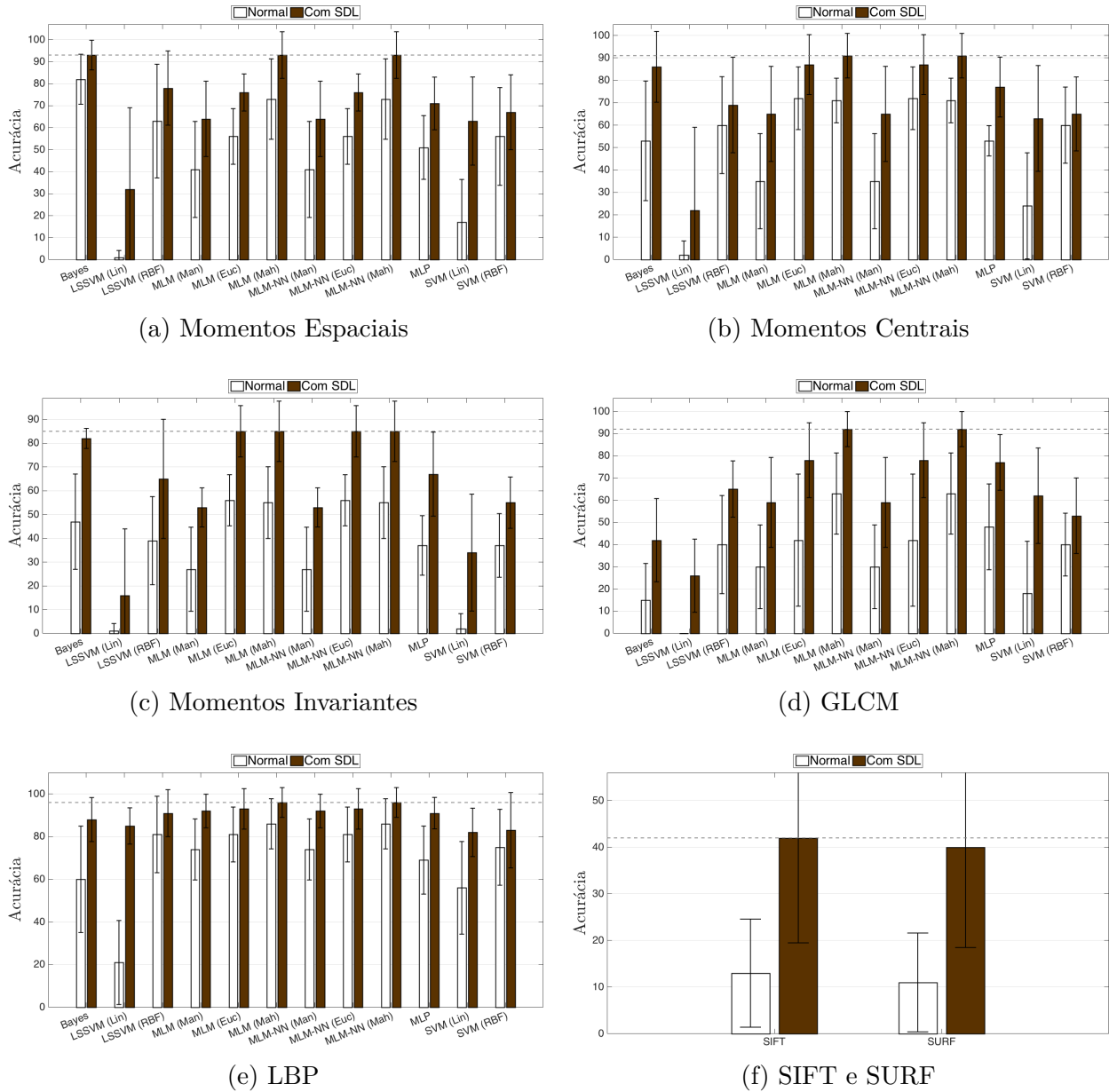


Figura 27 – Taxa de acerto média do teste de navegação, com e sem o Suporte de Decisão para Localização, obtidos pela combinação entre os classificadores e extratores utilizando a **câmera omnidirecional**.

Analisando os resultados detalhados por rota sem e com o SDL apresentados nas Tabelas 17 e 18, verifica-se que o ganho médio foi de 46,37% com o SDL. No geral, a abordagem proposta consegue realizar uma significativa melhora na acurácia. A MLM e MLM-NN com LBP, por exemplo, consegue 96% de acurácia com o SDL, enquanto que na situação anterior a acurácia máxima era 86%. Portanto, constata-se que a abordagem proposta é adequada e efetiva, também, com a câmera omnidirecional.

Tabela 17 – Taxa de acerto por rota das melhores configurações dos classificadores combinados com os extratores **sem o Suporte de Decisão para Localização** utilizando a **câmera omnidirecional**.

Extrator	Classificador	Configuração	Rota 1	Rota 2	Rota 3	Rota 4	Rota 5	Rota 6	Rota 7	Rota 8	Rota 9	Rota 10
Momentos Espaciais	Bayes	Normal	70,0±48,3	100,0±0,0	90,0±31,6	100,0±0,0	70,0±48,3	80,0±42,2	80,0±42,2	70,0±48,3	80,0±42,2	80,0±42,2
	LSSVM	Linear	0,0±0,0	0,0±0,0	10,0±31,6	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0
		RBF	40,0±51,6	50,0±52,7	80,0±42,2	20,0±42,2	50,0±52,7	90,0±31,6	90,0±31,6	50,0±52,7	100,0±0,0	60,0±51,6
	MLM	Manhattan	30,0±48,3	50,0±52,7	90,0±31,6	20,0±42,2	50,0±52,7	30,0±48,3	20,0±42,2	50,0±52,7	50,0±52,7	20,0±42,2
		Euclidiana	60,0±51,6	60,0±51,6	70,0±48,3	60,0±51,6	50,0±52,7	50,0±52,7	40,0±51,6	50,0±52,7	80,0±42,2	40,0±51,6
		Mahalanobis	60,0±51,6	100,0±0,0	100,0±0,0	60,0±51,6	90,0±31,6	70,0±48,3	80,0±42,2	60,0±51,6	50,0±52,7	60,0±51,6
	MLM-NN	Manhattan	30,0±48,3	50,0±52,7	90,0±31,6	20,0±42,2	50,0±52,7	30,0±48,3	20,0±42,2	50,0±52,7	50,0±52,7	20,0±42,2
		Euclidiana	60,0±51,6	60,0±51,6	70,0±48,3	60,0±51,6	50,0±52,7	50,0±52,7	40,0±51,6	50,0±52,7	80,0±42,2	40,0±51,6
		Mahalanobis	60,0±51,6	100,0±0,0	100,0±0,0	60,0±51,6	90,0±31,6	70,0±48,3	80,0±42,2	60,0±51,6	50,0±52,7	60,0±51,6
	MLP		60,0±51,6	50,0±52,7	60,0±51,6	50,0±52,7	80,0±42,2	60,0±51,6	40,0±51,6	40,0±51,6	40,0±51,6	30,0±48,3
	SVM	Linear	0,0±0,0	20,0±42,2	60,0±51,6	20,0±42,2	10,0±31,6	10,0±31,6	0,0±0,0	10,0±31,6	40,0±51,6	0,0±0,0
		RBF	40,0±51,6	50,0±52,7	70,0±48,3	30,0±48,3	40,0±51,6	70,0±48,3	60,0±51,6	70,0±48,3	100,0±0,0	30,0±48,3
Momentos Centrais	Bayes	Normal	40,0±51,6	90,0±31,6	40,0±51,6	50,0±52,7	50,0±52,7	70,0±48,3	40,0±51,6	0,0±0,0	90,0±31,6	60,0±51,6
	LSSVM	Linear	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	20,0±42,2	0,0±0,0
		RBF	60,0±51,6	40,0±51,6	90,0±31,6	50,0±52,7	50,0±52,7	70,0±48,3	60,0±51,6	50,0±52,7	100,0±0,0	30,0±48,3
	MLM	Manhattan	20,0±42,2	50,0±52,7	60,0±51,6	10,0±31,6	40,0±51,6	50,0±52,7	20,0±42,2	0,0±0,0	60,0±51,6	40,0±51,6
		Euclidiana	70,0±48,3	60,0±51,6	90,0±31,6	70,0±48,3	60,0±51,6	60,0±51,6	80,0±42,2	60,0±51,6	100,0±0,0	70,0±48,3
		Mahalanobis	70,0±48,3	70,0±48,3	80,0±42,2	70,0±48,3	60,0±51,6	60,0±51,6	70,0±48,3	60,0±51,6	80,0±42,2	90,0±31,6
	MLM-NN	Manhattan	20,0±42,2	50,0±52,7	60,0±51,6	10,0±31,6	40,0±51,6	50,0±52,7	20,0±42,2	0,0±0,0	60,0±51,6	40,0±51,6
		Euclidiana	70,0±48,3	60,0±51,6	90,0±31,6	70,0±48,3	60,0±51,6	60,0±51,6	80,0±42,2	60,0±51,6	100,0±0,0	70,0±48,3
		Mahalanobis	70,0±48,3	70,0±48,3	80,0±42,2	70,0±48,3	60,0±51,6	60,0±51,6	70,0±48,3	60,0±51,6	80,0±42,2	90,0±31,6
	MLP		60,0±51,6	50,0±52,7	50,0±52,7	60,0±51,6	50,0±52,7	50,0±52,7	40,0±51,6	60,0±51,6	60,0±51,6	50,0±52,7
	SVM	Linear	0,0±0,0	40,0±51,6	80,0±42,2	20,0±42,2	0,0±0,0	30,0±48,3	20,0±42,2	10,0±31,6	30,0±48,3	10,0±31,6
		RBF	50,0±52,7	50,0±52,7	80,0±42,2	60,0±51,6	70,0±48,3	70,0±48,3	50,0±52,7	40,0±51,6	90,0±31,6	40,0±51,6
Momentos Invariantes	Bayes	Normal	50,0±52,7	40,0±51,6	60,0±51,6	80,0±42,2	50,0±52,7	60,0±51,6	10,0±31,6	50,0±52,7	50,0±52,7	20,0±42,2
	LSSVM	Linear	0,0±0,0	0,0±0,0	10,0±31,6	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0
		RBF	30,0±48,3	50,0±52,7	50,0±52,7	30,0±48,3	10,0±31,6	40,0±51,6	30,0±48,3	40,0±51,6	80,0±42,2	30,0±48,3
	MLM	Manhattan	10,0±31,6	50,0±52,7	30,0±48,3	40,0±51,6	40,0±51,6	50,0±52,7	20,0±42,2	20,0±42,2	0,0±0,0	10,0±31,6
		Euclidiana	70,0±48,3	70,0±48,3	60,0±51,6	60,0±51,6	60,0±51,6	50,0±52,7	50,0±52,7	40,0±51,6	60,0±51,6	40,0±51,6
		Mahalanobis	50,0±52,7	60,0±51,6	80,0±42,2	50,0±52,7	50,0±52,7	80,0±42,2	60,0±51,6	40,0±51,6	40,0±51,6	40,0±51,6
	MLM-NN	Manhattan	10,0±31,6	50,0±52,7	30,0±48,3	40,0±51,6	40,0±51,6	50,0±52,7	20,0±42,2	20,0±42,2	0,0±0,0	10,0±31,6
		Euclidiana	70,0±48,3	70,0±48,3	60,0±51,6	60,0±51,6	60,0±51,6	50,0±52,7	50,0±52,7	40,0±51,6	60,0±51,6	40,0±51,6
		Mahalanobis	50,0±52,7	60,0±51,6	80,0±42,2	50,0±52,7	50,0±52,7	80,0±42,2	60,0±51,6	40,0±51,6	40,0±51,6	40,0±51,6
	MLP		60,0±51,6	40,0±51,6	50,0±52,7	20,0±42,2	30,0±48,3	40,0±51,6	20,0±42,2	40,0±51,6	30,0±48,3	40,0±51,6
	SVM	Linear	0,0±0,0	0,0±0,0	20,0±42,2	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0
		RBF	30,0±48,3	50,0±52,7	50,0±52,7	30,0±48,3	20,0±42,2	40,0±51,6	20,0±42,2	40,0±51,6	60,0±51,6	30,0±48,3
GLCM	Bayes	Normal	20,0±42,2	50,0±52,7	20,0±42,2	20,0±42,2	10,0±31,6	30,0±48,3	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0
	LSSVM	Linear	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0	0,0±0,0
		RBF	30,0±48,3	40,0±51,6	60,0±51,6	60,0±51,6	20,0±42,2	30,0±48,3	70,0±48,3	0,0±0,0	60,0±51,6	30,0±48,3
	MLM	Manhattan	30,0±48,3	40,0±51,6	60,0±51,6	50,0±52,7	20,0±42,2	50,0±52,7	10,0±31,6	10,0±31,6	10,0±31,6	20,0±42,2
		Euclidiana	10,0±31,6	80,0±42,2	80,0±42,2	80,0±42,2	20,0±42,2	60,0±51,6	30,0±48,3	10,0±31,6	30,0±48,3	20,0±42,2
		Mahalanobis	40,0±51,6	80,0±42,2	100,0±0,0	60,0±51,6	60,0±51,6	70,0±48,3	40,0±51,6	60,0±51,6	70,0±48,3	50,0±52,7
	MLM-NN	Manhattan	30,0±48,3	40,0±51,6	60,0±51,6	50,0±52,7	20,0±42,2	50,0±52,7	10,0±31,6	10,0±31,6	10,0±31,6	20,0±42,2
		Euclidiana	10,0±31,6	80,0±42,2	80,0±42,2	80,0±42,2	20,0±42,2	60,0±51,6	30,0±48,3	10,0±31,6	30,0±48,3	20,0±42,2
		Mahalanobis	40,0±51,6	80,0±42,2	100,0±0,0	60,0±51,6	60,0±51,6	70,0±48,3	40,0±51,6	60,0±51,6	70,0±48,3	50,0±52,7
	MLP		40,0±51,6	60,0±51,6	90,0±31,6	60,0±51,6	30,0±48,3	60,0±51,6	40,0±51,6	40,0±51,6	30,0±48,3	30,0±48,3
	SVM	Linear	0,0±0,0	40,0±51,6	70,0±48,3	30,0±48,3	0,0±0,0	20,0±42,2	0,0±0,0	0,0±0,0	20,0±42,2	0,0±0,0
		RBF	40,0±51,6	60,0±51,6	50,0±52,7	50,0±52,7	30,0±48,3	10,0±31,6	50,0±52,7	30,0±48,3	40,0±51,6	40,0±51,6
LBP	Bayes	Normal	10,0±31,6	60,0±51,6	80,0±42,2	60,0±51,6	90,0±31,6	60,0±51,6	50,0±52,7	80,0±42,2	80,0±42,2	30,0±48,3
	LSSVM	Linear	0,0±0,0	40,0±51,6	30,0±48,3	30,0±48,3	10,0±31,6	0,0±0,0	20,0±42,2	20,0±42,2	60,0±51,6	0,0±0,0
		RBF	70,0±48,3	90,0±31,6	90,0±31,6	60,0±51,6	100,0±0,0	70,0±48,3	80,0±42,2	100,0±0,0	100,0±0,0	50,0±52,7
	MLM	Manhattan	50,0±52,7	80,0±42,2	90,0±31,6	60,0±51,6	90,0±31,6	70,0±48,3	60,0±51,6	80,0±42,2	90,0±31,6	70,0±48,3
		Euclidiana	50,0±52,7	80,0±42,2	100,0±0,0	80,0±42,2	90,0±31,6	80,0±42,2	80,0±42,2	80,0±42,2	90,0±31,6	80,0±42,2
		Mahalanobis	60,0±51,6	90,0±31,6	100,0±0,0	80,0±42,2	90,0±31,6	90,0±31,6	80,0±42,2	80,0±42,2	100,0±0,0	90,0±31,6
	MLM-NN	Manhattan	50,0±52,7	80,0±42,2	90,0±31,6	60,0±51,6	90,0±31,6	70,0±48,3	60,0±51,6	80,0±42,2	90,0±31,6	70,0±48,3
		Euclidiana	50,0±52,7	80,0±42,2	100,0±0,0	80,0±42,2	90,0±31,6	80,0±42,2	80,0±42,2	80,0±42,2	90,0±31,6	80,0±42,2
		Mahalanobis	60,0±51,6	90,0±31,6	100,0±0,0	80,0±42,2	90,0±31,6	90,0±31,6	80,0±42,2	80,0±42,2	100,0±0,0	90,0±31,6
	MLP		50,0±52,7	80,0±42,2	90,0±31,6	60,0±51,6	90,0±31,6	50,0±52,7	50,0±52,7	80,0±42,2	70,0±48,3	70,0±48,3
	SVM	Linear	50,0±52,7	80,0±42,2	70,0±48,3	60,0±51,6	60,0±51,6	20,0±42,2	40,0±51,6	60,0±51,6	90,0±31,6	30,0±48,3
		RBF	70,0±48,3	90,0±31,6	100,0±0,0	70,0±48,3	70,0±48,3	70,0±48,3	70,0±48,3	70,0±48,3	100,0±0,0	40,0±51,6
SIFT			30,0±48,3	20,0±42,2	20,0±42,2	30,0±48,3	10,0±31,6	10,0±31,6	10,0±31,6	0,0±0,0	0,0±0,0	0,0±0,0
SURF			27,6±45,9	18,9±41,1	18,4±40,6	28,9±47,2	8,8±30,4	7,4±29,0	7,6±29,2	0,0±0,0	0,0±0,0	0,0±0,0

Tabela 18 – Taxa de acerto por rota das melhores configurações dos classificadores combinados com os extratores **com o Suporte de Decisão para Localização** utilizando a **câmera omnidirecional**.

Extrator	Classificador	Configuração	Rota 1	Rota 2	Rota 3	Rota 4	Rota 5	Rota 6	Rota 7	Rota 8	Rota 9	Rota 10
Momentos Espaciais	Bayes	Normal	80,0±42,2	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	90,0±31,6	90,0±31,6	90,0±31,6	100,0±0,0
	LSSVM	Linear	0,0±0,0	20,0±42,2	90,0±31,6	20,0±42,2	0,0±0,0	80,0±42,2	0,0±0,0	0,0±0,0	80,0±42,2	30,0±48,3
		RBF	60,0±51,6	60,0±51,6	90,0±31,6	50,0±52,7	70,0±48,3	90,0±31,6	90,0±31,6	80,0±42,2	100,0±0,0	90,0±31,6
	MLM	Manhattan	70,0±48,3	70,0±48,3	100,0±0,0	40,0±51,6	70,0±48,3	70,0±48,3	50,0±52,7	50,0±52,7	70,0±48,3	50,0±52,7
		Euclidiana	60,0±51,6	80,0±42,2	90,0±31,6	80,0±42,2	70,0±48,3	70,0±48,3	80,0±42,2	80,0±42,2	80,0±42,2	70,0±48,3
		Mahalanobis	70,0±48,3	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0	80,0±42,2	100,0±0,0
	MLM-NN	Manhattan	70,0±48,3	70,0±48,3	100,0±0,0	40,0±51,6	70,0±48,3	70,0±48,3	50,0±52,7	50,0±52,7	70,0±48,3	50,0±52,7
		Euclidiana	60,0±51,6	80,0±42,2	90,0±31,6	80,0±42,2	70,0±48,3	70,0±48,3	80,0±42,2	80,0±42,2	80,0±42,2	70,0±48,3
		Mahalanobis	70,0±48,3	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0	80,0±42,2	100,0±0,0
	MLP		60,0±51,6	70,0±48,3	90,0±31,6	70,0±48,3	90,0±31,6	80,0±42,2	60,0±51,6	70,0±48,3	60,0±51,6	60,0±51,6
SVM	Linear	30,0±48,3	60,0±51,6	80,0±42,2	50,0±52,7	40,0±51,6	80,0±42,2	80,0±42,2	50,0±52,7	90,0±31,6	70,0±48,3	
	RBF	50,0±52,7	60,0±51,6	70,0±48,3	40,0±51,6	60,0±51,6	80,0±42,2	80,0±42,2	70,0±48,3	100,0±0,0	60,0±51,6	
Momentos Centrais	Bayes	Normal	100,0±0,0	90,0±31,6	100,0±0,0	90,0±31,6	90,0±31,6	90,0±31,6	70,0±48,3	50,0±52,7	100,0±0,0	80,0±42,2
	LSSVM	Linear	0,0±0,0	0,0±0,0	100,0±0,0	0,0±0,0	0,0±0,0	60,0±51,6	0,0±0,0	0,0±0,0	60,0±51,6	0,0±0,0
		RBF	70,0±48,3	40,0±51,6	100,0±0,0	80,0±42,2	70,0±48,3	70,0±48,3	70,0±48,3	50,0±52,7	100,0±0,0	40,0±51,6
	MLM	Manhattan	70,0±48,3	60,0±51,6	70,0±48,3	40,0±51,6	90,0±31,6	80,0±42,2	40,0±51,6	40,0±51,6	100,0±0,0	60,0±51,6
		Euclidiana	80,0±42,2	70,0±48,3	100,0±0,0	90,0±31,6	60,0±51,6	90,0±31,6	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
		Mahalanobis	100,0±0,0	70,0±48,3	90,0±31,6	90,0±31,6	100,0±0,0	90,0±31,6	100,0±0,0	80,0±42,2	100,0±0,0	90,0±31,6
	MLM-NN	Manhattan	70,0±48,3	60,0±51,6	70,0±48,3	40,0±51,6	90,0±31,6	80,0±42,2	40,0±51,6	40,0±51,6	100,0±0,0	60,0±51,6
		Euclidiana	80,0±42,2	70,0±48,3	100,0±0,0	90,0±31,6	60,0±51,6	90,0±31,6	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0
		Mahalanobis	100,0±0,0	70,0±48,3	90,0±31,6	90,0±31,6	100,0±0,0	90,0±31,6	100,0±0,0	80,0±42,2	100,0±0,0	90,0±31,6
	MLP		90,0±31,6	80,0±42,2	80,0±42,2	90,0±31,6	60,0±51,6	50,0±52,7	70,0±48,3	80,0±42,2	80,0±42,2	90,0±31,6
SVM	Linear	50,0±52,7	90,0±31,6	100,0±0,0	60,0±51,6	40,0±51,6	70,0±48,3	60,0±51,6	20,0±42,2	80,0±42,2	60,0±51,6	
	RBF	60,0±51,6	50,0±52,7	90,0±31,6	70,0±48,3	70,0±48,3	70,0±48,3	60,0±51,6	40,0±51,6	90,0±31,6	50,0±52,7	
Momentos Invariantes	Bayes	Normal	80,0±42,2	80,0±42,2	80,0±42,2	90,0±31,6	90,0±31,6	80,0±42,2	80,0±42,2	80,0±42,2	80,0±42,2	80,0±42,2
	LSSVM	Linear	30,0±48,3	0,0±0,0	90,0±31,6	0,0±0,0	10,0±31,6	0,0±0,0	20,0±42,2	0,0±0,0	0,0±0,0	10,0±31,6
		RBF	50,0±52,7	60,0±51,6	100,0±0,0	80,0±42,2	20,0±42,2	80,0±42,2	50,0±52,7	60,0±51,6	100,0±0,0	50,0±52,7
	MLM	Manhattan	60,0±51,6	60,0±51,6	60,0±51,6	50,0±52,7	60,0±51,6	60,0±51,6	40,0±51,6	50,0±52,7	40,0±51,6	50,0±52,7
		Euclidiana	90,0±31,6	80,0±42,2	80,0±42,2	80,0±42,2	100,0±0,0	90,0±31,6	100,0±0,0	70,0±48,3	70,0±48,3	90,0±31,6
		Mahalanobis	100,0±0,0	80,0±42,2	100,0±0,0	80,0±42,2	100,0±0,0	90,0±31,6	80,0±42,2	80,0±42,2	80,0±42,2	60,0±51,6
	MLM-NN	Manhattan	60,0±51,6	60,0±51,6	60,0±51,6	50,0±52,7	60,0±51,6	60,0±51,6	40,0±51,6	50,0±52,7	40,0±51,6	50,0±52,7
		Euclidiana	90,0±31,6	80,0±42,2	80,0±42,2	80,0±42,2	100,0±0,0	90,0±31,6	100,0±0,0	70,0±48,3	70,0±48,3	90,0±31,6
		Mahalanobis	100,0±0,0	80,0±42,2	100,0±0,0	80,0±42,2	100,0±0,0	90,0±31,6	80,0±42,2	80,0±42,2	80,0±42,2	60,0±51,6
	MLP		80,0±42,2	50,0±52,7	90,0±31,6	40,0±51,6	90,0±31,6	60,0±51,6	60,0±51,6	70,0±48,3	50,0±52,7	80,0±42,2
SVM	Linear	40,0±51,6	30,0±48,3	90,0±31,6	0,0±0,0	30,0±48,3	40,0±51,6	50,0±52,7	30,0±48,3	20,0±42,2	10,0±31,6	
	RBF	60,0±51,6	60,0±51,6	70,0±48,3	50,0±52,7	40,0±51,6	60,0±51,6	40,0±51,6	50,0±52,7	70,0±48,3	50,0±52,7	
GLCM	Bayes	Normal	20,0±42,2	60,0±51,6	50,0±52,7	50,0±52,7	30,0±48,3	80,0±42,2	30,0±48,3	40,0±51,6	40,0±51,6	20,0±42,2
	LSSVM	Linear	10,0±31,6	30,0±48,3	60,0±51,6	20,0±42,2	20,0±42,2	20,0±42,2	10,0±31,6	20,0±42,2	50,0±52,7	20,0±42,2
		RBF	60,0±51,6	50,0±52,7	90,0±31,6	70,0±48,3	50,0±52,7	70,0±48,3	70,0±48,3	70,0±48,3	70,0±48,3	50,0±52,7
	MLM	Manhattan	60,0±51,6	60,0±51,6	90,0±31,6	80,0±42,2	70,0±48,3	70,0±48,3	40,0±51,6	20,0±42,2	50,0±52,7	50,0±52,7
		Euclidiana	70,0±48,3	90,0±31,6	100,0±0,0	100,0±0,0	70,0±48,3	70,0±48,3	50,0±52,7	90,0±31,6	60,0±51,6	80,0±42,2
		Mahalanobis	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	80,0±42,2	100,0±0,0	100,0±0,0	80,0±42,2
	MLM-NN	Manhattan	60,0±51,6	60,0±51,6	90,0±31,6	80,0±42,2	70,0±48,3	70,0±48,3	40,0±51,6	20,0±42,2	50,0±52,7	50,0±52,7
		Euclidiana	70,0±48,3	90,0±31,6	100,0±0,0	100,0±0,0	70,0±48,3	70,0±48,3	50,0±52,7	90,0±31,6	60,0±51,6	80,0±42,2
		Mahalanobis	90,0±31,6	90,0±31,6	100,0±0,0	100,0±0,0	90,0±31,6	90,0±31,6	80,0±42,2	100,0±0,0	100,0±0,0	80,0±42,2
	MLP		80,0±42,2	80,0±42,2	100,0±0,0	80,0±42,2	60,0±51,6	80,0±42,2	70,0±48,3	90,0±31,6	70,0±48,3	60,0±51,6
SVM	Linear	60,0±51,6	60,0±51,6	100,0±0,0	70,0±48,3	40,0±51,6	80,0±42,2	70,0±48,3	20,0±42,2	60,0±51,6	60,0±51,6	
	RBF	60,0±51,6	60,0±51,6	70,0±48,3	60,0±51,6	50,0±52,7	10,0±31,6	60,0±51,6	60,0±51,6	60,0±51,6	40,0±51,6	
LBP	Bayes	Normal	70,0±48,3	80,0±42,2	90,0±31,6	90,0±31,6	100,0±0,0	90,0±31,6	80,0±42,2	80,0±42,2	100,0±0,0	100,0±0,0
	LSSVM	Linear	90,0±31,6	100,0±0,0	90,0±31,6	80,0±42,2	80,0±42,2	80,0±42,2	70,0±48,3	90,0±31,6	90,0±31,6	80,0±42,2
		RBF	80,0±42,2	100,0±0,0	100,0±0,0	70,0±48,3	100,0±0,0	90,0±31,6	80,0±42,2	100,0±0,0	100,0±0,0	90,0±31,6
	MLM	Manhattan	90,0±31,6	80,0±42,2	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	90,0±31,6	80,0±42,2	100,0±0,0	90,0±31,6
		Euclidiana	80,0±42,2	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	80,0±42,2	80,0±42,2	100,0±0,0	100,0±0,0
		Mahalanobis	80,0±42,2	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	MLM-NN	Manhattan	90,0±31,6	80,0±42,2	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	90,0±31,6	80,0±42,2	100,0±0,0	90,0±31,6
		Euclidiana	80,0±42,2	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0	80,0±42,2	80,0±42,2	100,0±0,0	100,0±0,0
		Mahalanobis	80,0±42,2	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	90,0±31,6	100,0±0,0	100,0±0,0	100,0±0,0	100,0±0,0
	MLP		90,0±31,6	90,0±31,6	100,0±0,0	90,0±31,6	100,0±0,0	80,0±42,2	90,0±31,6	80,0±42,2	90,0±31,6	100,0±0,0
SVM	Linear	80,0±42,2	100,0±0,0	90,0±31,6	90,0±31,6	70,0±48,3	70,0±48,3	70,0±48,3	90,0±31,6	90,0±31,6	70,0±48,3	
	RBF	80,0±42,2	100,0±0,0	100,0±0,0	80,0±42,2	80,0±42,2	90,0±31,6	80,0±42,2	80,0±42,2	100,0±0,0	40,0±51,6	
SIFT			80,0±42,2	40,0±51,6	40,0±51,6	70,0±48,3	10,0±31,6	40,0±51,6	50,0±52,7	50,0±52,7	30,0±48,3	10,0±31,6
SURF			77,7±39,9	38,9±50,6	37,3±48,9	67,1±45,4	7,6±29,3	37,5±49,1	47,5±50,2	48,2±50,9	27,7±46,0	8,7±30,3

6.2.3 Custo de armazenamento (Omnidirecional)

A Tabela 19 mostra o custo de armazenamento, em kilobyte (KB), dos modelos dos classificadores por extrator de características utilizados no teste de navegação com a câmera omnidirecional.

De acordo com os resultados apresentados na Tabela 19, observa-se que os classificadores que obtiveram o menor espaço de armazenamento em média foram a rede MLP (8,79 KB), classificador Bayesiano (94,71 KB), SVM-RBF (136,11 KB) e SVM-Linear (188,88 KB). Quanto aos extratores, similar aos resultados da GoPro[®], os momentos consumiram menos espaço de armazenamento nos classificadores.

Tabela 19 – Tamanho em kilobyte dos modelos embarcados para cada classificador com a câmera omnidirecional.

	ME (KB)	MC (KB)	MI (KB)	GLCM (KB)	LBP (KB)	MÉDIA (KB)
Bayes (Normal)	15,16	8,84	8,84	23,60	417,12	94,71
LSSVM (Linear)	1089,09	824,71	824,71	1353,46	5407,21	1899,84
LSSVM (RBF)	1090,85	826,47	826,47	1355,22	5408,97	1901,60
MLM (Manhattan)	4547,76	4530,18	4530,18	4565,34	4834,87	4601,67
MLM (Euclidiana)	4547,76	4530,18	4530,18	4565,34	4834,87	4601,67
MLM (Mahalanobis)	4550,65	4532,67	4532,67	4568,76	4864,17	4609,78
MLM-NN (Manhattan)	4547,76	4530,18	4530,18	4565,34	4834,87	4601,67
MLM-NN (Euclidiana)	4547,76	4530,18	4530,18	4565,34	4834,87	4601,67
MLM-NN (Mahalanobis)	4550,65	4532,67	4532,67	4568,76	4864,17	4609,78
MLP	6,73	11,41	6,02	7,43	12,35	8,79
SVM (Linear)	166,49	147,53	173,20	172,56	284,64	188,88
SVM (RBF)	81,44	99,58	71,17	100,03	328,34	136,11
MÉDIA	2478,51	2425,38	2424,71	2534,27	3410,54	
	SIFT : 283335,60			SURF: 741,14		

6.3 Análise geral dos resultados

Esse suma, esse capítulo apresentou os resultados dos testes obtidos pela combinação dos extratores de características com os classificadores na tarefa de localização. O estudo foi realizado com quatro conjuntos de imagens provenientes do ambiente sintético e real com as câmeras GoPro[®] e omnidirecional. Por fim, os experimentos são validados através dos testes de navegação com os dois tipos de câmera. Os testes práticos foram, ainda, analisados com a abordagem proposta de Suporte de Decisão para Localização.

A melhor acurácia obtida com a câmera de campo de visão convencional foi 99,96% através da combinação da MLM-NN com os Momentos Espaciais, cujo tempo de classificação foi 128,8µs e de extração 0,165s, sendo este o melhor tempo de extração obtido.

Quanto ao tempo de classificação, o melhor resultado obtido foi 47,3 μ s com classificador Bayesiano, cuja acurácia foi 99,94%.

Nos experimentos com as imagens da câmera omnidirecional, a combinação com maior acurácia foi 96,67% através da MLM-NN com LBP, cujo tempo de classificação foi 1210 μ s e de extração 0,337s. Por outro lado, o menor tempo de extração foi de 0,178s com os ME e de classificação 36,4 μ s com classificador Bayesiano, cuja acurácia foi 87,5%.

Quanto aos ganhos obtidos pelo SDL, verifica-se que a acurácia média do teste de navegação com a primeira câmera foi 83% sem o SDL e 91,9% com o SDL, e passou de 36,3% para 52,1% com SDL na câmera omnidirecional. Utilizando a GoPro[®] sem o SDL para MLM-NN com ME e classificador Bayesiano com ME, a taxa de acerto dos testes de navegação foi 100% de acurácia, e o SDL apenas confirmou o trajeto obtido. Entretanto, utilizando a câmera omnidirecional, os valores obtidos na navegação aumentaram 11,6% na taxa de acerto para MLM-NN com LBP, passando de 86% para 96%. De forma similar, a acurácia do classificador Bayesiano com ME passou de 82% para 93%.

Em relação ao espaço de armazenamento, o classificador que obteve em média o menor valor com a GoPro[®] foi a rede MLP com 4,96 KB, cuja acurácia máxima no teste de navegação foi 97%. Por outro lado, o classificador Bayesiano consumiu em média 94,71 KB e obteve 100% de acurácia. Quanto à câmera omnidirecional, observa-se novamente que a rede MLP consumiu menos espaço de armazenamento, 8,79 KB, enquanto conseguiu 69% de acurácia no teste de navegação. Em contrapartida, a MLM-NN, classificador com maior acurácia no teste de navegação com a câmera omnidirecional, 86%, consumiu em média 4609,78 (KB).

Portanto, conclui-se que a combinação do classificador Bayesiano com os ME utilizando a câmera de visão convencional é a melhor escolha para a tarefa de navegação considerando a acurácia, custo computacional e de armazenamento, dentre as configurações avaliadas. Finalmente, conclui-se que o SDL aumenta a eficácia e confiabilidade na navegação em robótica móvel, podendo ser utilizado para este fim em aplicações reais.

7 Conclusão

Esta dissertação apresentou um estudo e avaliação de alternativas para o uso de dois tipos de câmera com extratores de características e classificadores em um robô móvel, mais especificamente na tarefa de localização em um ambiente interno.

Inicialmente, foram estudados os aspectos relacionados aos sistemas de reconhecimento de locais internos. Conforme visto, diversas técnicas foram desenvolvidas para as etapas de extração de características e classificação de padrões. Dentre elas, foram descritas e analisadas os extratores de textura LBP e GLC, os Momentos Espaciais, Centrais e de Hu, além dos descritores de ponto-chave, SIFT e SURF. Ademais, foram discutidas algumas abordagens existentes para o projeto do classificador de padrões, tais como o classificador Bayesiano, a rede MLP, classificadores embasados pela Teoria de Aprendizado Estatístico, SVM e LSSVM, e classificadores baseados em distância, MLM e MLM-NN.

Dentre as possíveis abordagens para representação do problema de localização, utilizou-se um mapa topológico preexistente onde cada nó representa um local no ambiente e, juntamente com sua orientação, forma-se uma classe. Portanto, durante a navegação, o robô captura a imagem, extrai suas características e identifica sua localização através de um dos classificadores. Com objetivo de melhorar a tomada de decisão, foi proposta uma abordagem que usa as informações de vizinhança do mapa topológico.

O ambiente foi simulado computacionalmente com as câmeras de visão convencional e omnidirecional para formação dos conjuntos de imagens sintéticas. De maneira semelhante, foram capturadas imagens com a câmera GoPro® e omnidirecional para criação dos bancos de imagens reais. O estudo, a princípio, foi executado com esses quatro conjuntos de imagens. Os métodos foram explorados quanto às medidas de avaliação, custo computacional e de armazenamento. Além disso, foram avaliadas três métricas de distância na MLM e MLM-NN. No SVM e LSSVM, foi investigado o uso de dois *kernels*. Finalmente, a abordagem proposta foi validada por meio dos testes de navegação com os dois tipos de câmera.

A primeira etapa dos estudos experimentais com a câmera convencional mostrou, de modo geral, que os resultados com as imagens dos ambientes sintético e real foram semelhantes. Esse comportamento evidencia que os métodos adequaram-se bem ao problema. No entanto, algumas combinações de extratores e classificadores, como a GLCM e o SVM Linear, conseguiram resultados melhores com as imagens do ambiente real. Isso ocorreu devido a ausência de alguns detalhes durante a modelagem do ambiente na simulação computacional. A melhor acurácia obtida foi 99,96% através da combinação da MLM-NN com os Momentos Espaciais, cujo tempo de classificação foi 128,8µs e de extração 0,165s, sendo este o melhor tempo de extração obtido. Em relação ao tempo de classificação, o

melhor resultado obtido foi 47,3 μ s com classificador Bayesiano, cuja acurácia foi 99,94%.

Após a investigação *offline*, realizou-se o teste de navegação com a câmera GoPro®. Novamente, a boa adequação dos métodos ao problema pode ser confirmada pela semelhança entre os resultados. A eficiência dos Momentos Espaciais, Centrais e LBP, novamente, é constatada pelas excelentes taxas de acerto entre 98% e 100%. O último extrator conseguiu boa performance, também, com os classificadores lineares. Outra observação relevante é quanto a influência das métricas de distância na MLM e MLM-NN que foi semelhante aos resultados *offline*. A distância Manhattan, além de possuir um custo computacional baixo, obtém acurácia semelhante as demais quando usada com os extratores de textura. Em relação aos ganhos obtidos pelo SDL, observa-se que a acurácia média do teste de navegação com a câmera GoPro® foi 83% sem o SDL e 91,9% com o SDL. Utilizando a mesma câmera sem o SDL para MLM-NN com ME e classificador Bayesiano com ME, a taxa de acerto dos testes de navegação foi 100% de acurácia, e o SDL apenas confirmou o trajeto obtido.

A segunda etapa dos experimentos foi realizada com a câmera omnidirecional. Ao comparar os resultados com os anteriores, observa-se que os da câmera omnidirecional foram inferiores aos da câmera de visão convencional. A combinação que obteve maior acurácia foi 96,67% através da MLM-NN com LBP, cujo tempo de classificação foi 1210 μ s e de extração 0,337s. Por outro lado, o menor tempo de extração foi de 0,178s com os ME e de classificação 36,4 μ s com classificador Bayesiano, cuja acurácia foi 87,5%.

Finalmente, foram realizados os testes de navegação com a câmera omnidirecional. Conforme os resultados, constata-se que a melhor combinação foi a do extrator LBP com o classificadores MLM e MLM-NN usando a distância Mahalanobis com 86% de acurácia. O classificador Bayesiano com os Momentos Espaciais é outra combinação importante com 82% de acurácia. De maneira semelhante aos resultados *offline*, os classificadores lineares não conseguiram bom desempenho. Ou seja, as imagens omnidirecionais não tiveram uma contribuição tão relevante quanto as de campo de visão convencional. Em relação aos ganhos obtidos pelo SDL, observa-se que a acurácia média do teste de navegação com a câmera omnidirecional foi 36,3% sem o SDL e 52,1% com o SDL. Dentre as melhorias na taxa de acerto, destaca-se o crescimento de 11,6% de acurácia da MLM e MLM-NN com LBP, passando de 86% para 96%. De forma similar, a acurácia do classificador Bayesiano com ME passou de 82% para 93%.

Quanto ao espaço de armazenamento, o classificador que obteve em média o menor valor com a câmera GoPro® foi a rede MLP com 4,96 KB, cuja acurácia máxima no teste de navegação foi 97%. Por outro lado, o classificado Bayesiano consumiu em média 94,71 KB e obteve 100% de acurácia. Quanto à câmera omnidirecional, observa-se novamente que a rede MLP consumiu menos espaço de armazenamento, 8,79 KB, enquanto conseguiu 69% de acurácia no teste de navegação. Em contrapartida, a MLM-NN, classificador com

maior acurácia no teste de navegação, consumiu em média 4609,78 (KB).

Portanto, constata-se que a câmera de visão convencional juntamente com classificador Bayesiano e os ME é a melhor escolha para a tarefa de navegação considerando a acurácia, custo computacional e de armazenamento, dentre as configurações analisadas. Por fim, conclui-se que o SDL aumenta a confiabilidade e eficiência na navegação em robótica móvel, o que permite seu uso em aplicações reais com este fim.

Além dos excelentes resultados obtidos pelo Suporte de Decisão para Localização, outra contribuição relevante dessa dissertação é a criação de quatro bancos de imagens. Eles podem ser usados em trabalhos futuros com intuito de analisar ou propor outras técnicas de localização de robôs móveis.

7.1 Proposta de Trabalhos Futuros

O trabalho apresentado nesta dissertação pode ser estendido em várias direções. Como trabalhos futuros, sugerem-se:

- Realizar um estudo com novos ambientes;
- Avaliar técnicas de aprendizagem incremental;
- Aplicar técnicas de seleção de características nas amostras geradas pelos extratores;
- Utilizar outras métricas de distância na MLM e MLM-NN;
- Investigar a influência da resolução das imagens;
- Analisar outros métodos, tal como a Floresta de Caminhos Ótimos (*Optimum-Path Forest*, OPF) (NUNES et al., 2014; PAPA et al., 2012);
- Aplicar paralelismo nos métodos analisados.

Referências

- AGINSKY, V.; TARR, M. J. How are different properties of a scene encoded in visual memory? *Visual Cognition*, Taylor & Francis, v. 7, n. 1-3, p. 147–162, 2000. Citado na página 39.
- ALARIFI, A.; AL-SALMAN, A. S.; ALSALEH, M.; ALNAFESSAH, A.; ALHADHRAMI, S.; AL-AMMAR, M. A.; AL-KHALIFA, H. S. Ultra wideband indoor positioning technologies: Analysis and recent advances. *Sensors*, v. 16, n. 5, p. 707, 2016. Citado 4 vezes nas páginas 13, 28, 29 e 30.
- ARBEITER, G.; BORMANN, R.; FISCHER, J.; HÄGELE, M.; VERL, A. Towards geometric mapping for semi-autonomous mobile robots. In: *Spatial Cognition VIII - International Conference, Spatial Cognition 2012, Kloster Seeon, Germany, August 31 - September 3, 2012. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 114–127. Citado na página 40.
- BAY, H.; ESS, A.; TUYTELAARS, T.; GOOL, L. V. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, Elsevier Science Inc., New York, NY, USA, v. 110, n. 3, p. 346–359, jun. 2008. ISSN 1077-3142. Citado 3 vezes nas páginas 13, 62 e 63.
- BESSA, J. A.; BARROSO, D. A.; NETO, A. R. da R.; ALEXANDRIA, A. R. de. Global location of mobile robots using artificial neural networks in omnidirectional images. *IEEE Latin America Transactions*, v. 13, n. 10, p. 3405–3414, Oct 2015. ISSN 1548-0992. Citado na página 31.
- BIEDERMAN, I. On the semantics of a glance at a scene. in: M. Kubovy, J. Pomerantz (eds.), *perceptual organization*. Erlbaum, Hillsdale, NJ, p. 213–253, 1981. Citado 3 vezes nas páginas 13, 35 e 36.
- BIEDERMAN, I.; MEZZANOTTE, R. J.; RABINOWITZ, J. C. Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive psychology*, Elsevier, v. 14, n. 2, p. 143–177, 1982. Citado na página 38.
- BISHOP, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006. ISBN 0387310738. Citado 2 vezes nas páginas 65 e 66.
- BONIN-FONT, F.; ORTIZ, A.; OLIVER, G. Visual navigation for mobile robots: A survey. *Journal of Intelligent and Robotic Systems*, v. 53, n. 3, p. 263, 2008. ISSN 1573-0409. Citado na página 31.
- BRAGA, A. de P. *Redes neurais artificiais: teoria e aplicações*. Rio de Janeiro, RJ, Brasil: LTC Editora, 2007. ISBN 9788521615644. Citado na página 74.
- CALDERONI, L.; FERRARA, M.; FRANCO, A.; MAIO, D. Indoor localization in a hospital environment using random forest classifiers. *Expert Systems with Applications*, v. 42, n. 1, p. 125 – 134, 2015. ISSN 0957-4174. Citado na página 30.

- CAMPOS, R. S.; LOVISOLO, L.; CAMPOS, M. L. R. de. Wi-fi multi-floor indoor positioning considering architectural aspects and controlled computational complexity. *Expert Systems with Applications*, v. 41, n. 14, p. 6211 – 6223, 2014. ISSN 0957-4174. Citado na página 30.
- CARON, L.; FILLIAT, D.; GEPPERTH, A. Neural network fusion of color, depth and location for object instance recognition on a mobile robot. In: *Computer Vision - ECCV 2014 Workshops - Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part III*. Cham: Springer International Publishing, 2014. p. 791–805. Citado na página 28.
- CASTELLANOS, J. A.; TARDOS, J. D. *Mobile robot localization and map building: A multisensor fusion approach*. Norwell, MA, USA: Springer Science & Business Media, 2012. Citado na página 28.
- CHARALAMPOUS, K.; KOSTAVELIS, L.; GASTERATOS, A. Thorough robot navigation based on {SVM} local planning. *Robotics and Autonomous Systems*, v. 70, p. 166 – 180, 2015. ISSN 0921-8890. Citado na página 28.
- CHENG, H.; CHEN, H.; LIU, Y. Topological indoor localization and navigation for autonomous mobile robot. *Automation Science and Engineering, IEEE Transactions on*, v. 12, n. 2, p. 729–738, April 2015. ISSN 1545-5955. Citado na página 40.
- CHIN, W. H.; LOO, C. K.; SEERA, M.; KUBOTA, N.; TODA, Y. Multi-channel bayesian adaptive resonance associate memory for on-line topological map building. *Applied Soft Computing*, v. 38, p. 269 – 280, 2016. ISSN 1568-4946. Citado na página 40.
- CHOWDHURY, T. J.; ELKIN, C.; DEVABHAKTUNI, V.; RAWAT, D. B.; OLUOCH, J. Advances on localization techniques for wireless sensor networks: A survey. *Computer Networks*, p. –, 2016. ISSN 1389-1286. Citado na página 30.
- CHRISTOU, C.; BÜLTHOFF, H. H. View-direction specificity in scene recognition after active and passive learning. CiteSeer, 1997. Citado na página 40.
- CRAMMER, K.; SINGER, Y.; CRISTIANINI, N.; SHAW-TAYLOR, J.; WILLIAMSON, B. On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research*, v. 2, p. 2001, 2001. Citado na página 67.
- CUMMINS, M.; NEWMAN, P. Appearance-only slam at large scale with fab-map 2.0. *The International Journal of Robotics Research*, SAGE Publications, v. 30, n. 9, p. 1100–1123, 2011. Citado na página 31.
- DAWOOD, F.; LOO, C. K. Robot behaviour learning using topological gaussian adaptive resonance hidden markov model. *Neural Computing and Applications*, p. 1–14, 2015. ISSN 1433-3058. Citado na página 40.
- DEAK, G.; CURRAN, K.; CONDELL, J. A survey of active and passive indoor localisation systems. *Computer Communications*, v. 35, n. 16, p. 1939 – 1954, 2012. ISSN 0140-3664. Citado 2 vezes nas páginas 28 e 30.
- DEBSKI, A.; GRAJEWSKI, W.; ZABOROWSKI, W.; TUREK, W. Open-source localization device for indoor mobile robots. *Procedia Computer Science*, v. 76, p. 139 – 146, 2015. ISSN 1877-0509. 2015 {IEEE} International Symposium on Robotics and Intelligent Sensors (IEEE IRIS2015). Citado na página 30.

- DUAN, K.-B.; KEERTHI, S. S. Which is the best multiclass svm method? an empirical study. In: OZA, N. C.; POLIKAR, R.; KITTLER, J.; ROLI, F. (Ed.). *International Workshop on Multiple Classifier Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. p. 278–285. ISBN 978-3-540-31578-0. Citado na página 67.
- DUDA, R.; HART, P.; STORK, D. *Pattern classification*. New Jersey, EUA: Wiley, 2001. (Pattern Classification and Scene Analysis: Pattern Classification). ISBN 9780471056690. Citado na página 66.
- ENRICO, P.; GROEN, F.; ARAI, T.; DILLMANN, R.; STENZ, A. *Intelligent autonomous systems 6*. Amsterdam: IOS Press, 2000. Citado na página 40.
- EPSTEIN, R.; HARRIS, A.; STANLEY, D.; KANWISHER, N. The parahippocampal place area: Recognition, navigation, or encoding? *Neuron*, Elsevier, v. 23, n. 1, p. 115–125, 1999. Citado na página 39.
- EPSTEIN, R.; KANWISHER, N. A cortical representation of the local visual environment. *Nature*, Nature Publishing Group, v. 392, n. 6676, p. 598–601, 1998. Citado na página 39.
- EPSTEIN, R. A.; HIGGINS, J. S.; THOMPSON-SCHILL, S. L. Learning places from views: variation in scene processing as a function of experience and navigational ability. *Journal of Cognitive Neuroscience*, MIT Press, v. 17, n. 1, p. 73–83, 2005. Citado na página 40.
- EVANS, C. *Notes on the OpenSURF Library*. Bristol, United Kingdom, 2009. Citado 2 vezes nas páginas 13 e 63.
- FANG, S.-H.; WANG, C.-H.; HUANG, T.-Y.; YANG, C.-H.; CHEN, Y.-S. An enhanced zigbee indoor positioning system with an ensemble approach. *IEEE Communications Letters*, IEEE, v. 16, n. 4, p. 564–567, 2012. Citado na página 30.
- FLUSSER, J.; ZITOVA, B.; SUK, T. *Moments and Moment Invariants in Pattern Recognition*. Chichester, England: Wiley Publishing, 2009. ISBN 0470699876, 9780470699874. Citado 3 vezes nas páginas 48, 49 e 50.
- FÖRSTNER, W. A framework for low level feature extraction. In: EKLUNDH, J.-O. (Ed.). *European Conference on Computer Vision*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994. p. 383–394. ISBN 978-3-540-48400-4. Citado na página 59.
- FRIEDMAN, A. Framing pictures: the role of knowledge in automatized encoding and memory for gist. *Journal of experimental psychology: General*, American Psychological Association, v. 108, n. 3, p. 316, 1979. Citado na página 38.
- FUENTES, J.; RUIZ, J.; RENDÓN, J. M. Visual simultaneous localization and mapping: a survey. *Artificial Intelligence Review*, v. 43, n. 1, p. 55–81, 2015. ISSN 1573-7462. Citado na página 31.
- GALVEZ-LÓPEZ, D.; TARDOS, J. D. Bags of binary words for fast place recognition in image sequences. *IEEE Transactions on Robotics*, v. 28, n. 5, p. 1188–1197, Oct 2012. ISSN 1552-3098. Citado na página 31.

GARCIA-FIDALGO, E.; ORTIZ, A. Vision-based topological mapping and localization methods. *Robot. Auton. Syst.*, North-Holland Publishing Co., Amsterdam, The Netherlands, The Netherlands, v. 64, n. C, p. 1–20, fev. 2015. ISSN 0921-8890. Citado 2 vezes nas páginas 28 e 31.

GASPAR, J. *Sketchup Pro 2013 Step by Step*. São Paulo, Brazil: GetPro Books, 2013. Citado na página 86.

GERSTMAYR-HILLEN, L.; RÖBEN, F.; KRZYKAWSKI, M.; KREFT, S.; VENJAKOB, D.; MÖLLER, R. Dense topological maps and partial pose estimation for visual control of an autonomous cleaning robot. *Robotics and Autonomous Systems*, v. 61, n. 5, p. 497 – 516, 2013. ISSN 0921-8890. Citado na página 40.

GONZALEZ, R. C.; WOODS, R. *Digital Image Processing*. 3a. ed. New Jersey, USA: Pearson Prentice Hall, 2010. ISBN 0-13-168728-x. Citado 4 vezes nas páginas 13, 37, 48 e 56.

HAFEZ, M. B.; LOO, C. K. Topological q-learning with internally guided exploration for mobile robot navigation. *Neural Computing and Applications*, v. 26, n. 8, p. 1939–1954, 2015. ISSN 1433-3058. Citado na página 40.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3, n. 6, p. 610–621, Nov 1973. ISSN 0018-9472. Citado 2 vezes nas páginas 54 e 57.

HARTMANN, B. *Human Worker activity recognition in industrial environments*. Karlsruhe, Germany: KIT Scientific Publishing, 2011. Citado na página 30.

HAYKIN, S.; ENGEL, P. *Redes neurais: princípios e prática*. Brasil: Bookman Editora, 2007. Citado 3 vezes nas páginas 75, 88 e 94.

HENDERSON, J. M.; FERREIRA, F. Scene perception for psycholinguists. Psychology Press, 2004. Citado na página 38.

HENDERSON, J. M.; HOLLINGWORTH, A. High-level scene perception. *Annual review of psychology*, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 50, n. 1, p. 243–271, 1999. Citado na página 35.

HU, M.-K. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, v. 8, n. 2, p. 179–187, February 1962. Citado na página 51.

JIMENEZ, A. R.; SECO, F.; PRIETO, C.; GUEVARA, J. A comparison of pedestrian dead-reckoning algorithms using a low-cost mems imu. In: *Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on*. Budapest, Hungary: IEEE, 2009. p. 37–42. Citado na página 30.

JOHNS, E.; YANG, G.-Z. Global localization in a dense continuous topological map. In: *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. Shanghai, China: IEEE, 2011. p. 1032–1037. Citado na página 31.

KASAT, N. R.; THEPADE, S. D. Novel content based image classification method using {LBG} vector quantization method with bayes and lazy family data mining classifiers. *Procedia Computer Science*, v. 79, p. 483 – 489, 2016. ISSN 1877-0509. Proceedings of

- International Conference on Communication, Computing and Virtualization (ICCCV) 2016. Citado na página 28.
- KIM, H.-S.; CHOI, J.-S. Advanced indoor localization using ultrasonic sensor and digital compass. In: *Control, Automation and Systems, 2008. ICCAS 2008. International Conference on*. Seoul, Korea: IEEE, 2008. p. 223–226. Citado na página 30.
- KO, N. Y.; KUC, T.-Y. Fusing range measurements from ultrasonic beacons and a laser range finder for localization of a mobile robot. *Sensors*, v. 15, n. 5, p. 11050, 2015. ISSN 1424-8220. Citado na página 27.
- KOTHARI, N.; KANNAN, B.; GLASGOW, E. D.; DIAS, M. B. Robust indoor localization on a commercial smart phone. *Procedia Computer Science*, v. 10, p. 1114 – 1120, 2012. ISSN 1877-0509. Citado na página 30.
- KWIECIEŃ, A.; MAĆKOWSKI, M.; KOJDER, M.; MANCZYK, M. Reliability of bluetooth smart technology for indoor localization system. In: GAJ, P.; KWIECIEŃ, A.; STERA, P. (Ed.). *Computer Networks: 22nd International Conference, CN 2015, Brunów, Poland, June 16-19, 2015. Proceedings*. Cham: Springer International Publishing, 2015. p. 444–454. ISBN 978-3-319-19419-6. Citado na página 30.
- LI, H. Low-cost 3d bluetooth indoor positioning with least square. *Wireless Personal Communications*, v. 78, n. 2, p. 1331–1344, 2014. ISSN 1572-834X. Citado na página 30.
- LI-ZHEN, X. Vray render 3ds max scene rendering techniques. *Coal Technology*, v. 9, p. 098, 2010. Citado na página 86.
- LIM, H.; KUNG, L.-C.; HOU, J. C.; LUO, H. Zero-configuration indoor localization over ieee 802.11 wireless infrastructure. *Wireless Networks*, v. 16, n. 2, p. 405–420, 2010. ISSN 1572-8196. Citado na página 30.
- LINDEBERG, T. Feature detection with automatic scale selection. *International Journal of Computer Vision*, v. 30, n. 2, p. 79–116, 1998. ISSN 1573-1405. Citado na página 60.
- LIU, M.; SIEGWART, R. Dp-fact: Towards topological mapping and scene recognition with color for omnidirectional camera. In: *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. Saint Paul, Minnesota, USA: IEEE, 2012. p. 3503–3508. ISSN 1050-4729. Citado na página 31.
- LOWE, D. G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, Kluwer Academic Publishers, Hingham, MA, USA, v. 60, n. 2, p. 91–110, nov. 2004. ISSN 0920-5691. Citado 3 vezes nas páginas 13, 60 e 61.
- MA, S.; JIA, L.; LI, X.; WANG, L.; ZHOU, H.; SUN, X. *Life System Modeling and Simulation: International Conference on Life System Modeling and Simulation, LSMS 2014, and International Conference on Intelligent Computing for Sustainable Energy and Environment, ICSEE 2014, Shanghai, China, September 2014, Proceedings*. China: Springer, 2014. Citado na página 27.
- MAINETTI, L.; PATRONO, L.; SERGI, I. A survey on indoor positioning systems. In: *Software, Telecommunications and Computer Networks (SoftCOM), 2014 22nd International Conference on*. Split, Croatia: IEEE, 2014. p. 111–120. Citado na página 30.

MARINHO, L. B.; ALMEIDA, J. S.; SOUZA, J. W. M.; ALBUQUERQUE, V. H. C.; FILHO, P. P. R. A novel mobile robot localization approach based on topological maps using classification with reject option in omnidirectional images. *Expert Systems with Applications*, Elsevier, 2016. Citado na página 91.

MAUTZ, R.; TILCH, S. Survey of optical indoor positioning systems. In: *Indoor Positioning and Indoor Navigation (IPIN), 2011 International Conference on*. Sapporo, Japan: IEEE, 2011. p. 1–7. Citado 2 vezes nas páginas 28 e 30.

MERCER, J. Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, The Royal Society, v. 209, n. 441-458, p. 415–446, 1909. ISSN 0264-3952. Citado na página 71.

MESQUITA, D. P. P.; GOMES, J. P. P.; SOUZA JUNIOR, A. H. Ensemble of minimal learning machines for pattern classification. In: *Advances in Computational Intelligence - 13th International Work-Conference on Artificial Neural Networks, IWANN 2015, Palma de Mallorca, Spain, June 10-12, 2015. Proceedings, Part II*. Palma de Mallorca, Spain: Springer, 2015. p. 142–152. Citado na página 80.

MIAH, M. S.; GUEAIEB, W. A fuzzy logic approach for indoor mobile robot navigation using ukf and customized rfid communication. In: KAMEL, M.; KARRAY, F.; GUEAIEB, W.; KHAMIS, A. (Ed.). *Autonomous and Intelligent Systems: Second International Conference, AIS 2011, Burnaby, BC, Canada, June 22-24, 2011. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. p. 21–30. ISBN 978-3-642-21538-4. Citado na página 30.

MIKOLAJCZYK, K.; SCHMID, C. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, v. 60, n. 1, p. 63–86, 2004. ISSN 1573-1405. Citado na página 60.

MONICA, S.; FERRARI, G. A swarm intelligence approach to 3d distance-based indoor uwb localization. In: MORA, A. M.; SQUILLERO, G. (Ed.). *Applications of Evolutionary Computation: 18th European Conference, EvoApplications 2015, Copenhagen, Denmark, April 8-10, 2015, Proceedings*. Cham: Springer International Publishing, 2015. p. 91–102. ISBN 978-3-319-16549-3. Citado na página 30.

MORIOKA, H.; YI, S.; HASEGAWA, O. Vision-based mobile robot's slam and navigation in crowded environments. In: *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Alberta, Canada: IEEE, 2011. p. 3998–4005. ISSN 2153-0858. Citado na página 31.

MUKUNDAN, R.; RAMAKRISHNAN, K. R. *Moment Functions in Image Analysis: Theory and Applications*. USA: World Scientific Publishing Company, 1988. Citado 2 vezes nas páginas 48 e 49.

NAGLA, K.; UDDIN, M.; SINGH, D. Improved occupancy grid mapping in specular environment. *Robotics and Autonomous Systems*, v. 60, n. 10, p. 1245 – 1252, 2012. ISSN 0921-8890. Citado na página 40.

NAVON, D. Forest before trees: The precedence of global features in visual perception. *Cognitive psychology*, Elsevier, v. 9, n. 3, p. 353–383, 1977. Citado na página 39.

- NAYAR, S. K. Catadioptric omnidirectional camera. In: *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. Puerto Rico, USA: IEEE, 1997. p. 482–488. ISSN 1063-6919. Citado na página 43.
- NEUBECK, A.; GOOL, L. V. Efficient non-maximum suppression. In: *18th International Conference on Pattern Recognition (ICPR'06)*. Hong Kong: IEEE, 2006. v. 3, p. 850–855. ISSN 1051-4651. Citado na página 63.
- NOWICKI, M.; SKRZYPCZYŃSKI, P. Indoor navigation with a smartphone fusing inertial and wifi data via factor graph optimization. In: SIGG, S.; NURMI, P.; SALIM, F. (Ed.). *Mobile Computing, Applications, and Services: 7th International Conference, MobiCASE 2015, Berlin, Germany, November 12–13, 2015, Revised Selected Papers*. Cham: Berlin, Germany, 2015. p. 280–298. Citado na página 30.
- NUNES, T. M.; COELHO, A. L.; LIMA, C. A.; PAPA, J. P.; ALBUQUERQUE, V. H. C. de. {EEG} signal classification for epilepsy diagnosis via optimum path forest – a systematic assessment. *Neurocomputing*, v. 136, p. 103 – 123, 2014. ISSN 0925-2312. Citado na página 119.
- OJALA, T.; PIETIKAINEN, M.; HARWOOD, D. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In: *Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision and Image Processing., Proceedings of the 12th IAPR International Conference on*. Jerusalem, Israel: IEEE, 1994. v. 1, p. 582–585 vol.1. Citado na página 52.
- OJALA, T.; PIETIKAINEN, M.; MAENPAA, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 7, p. 971–987, Jul 2002. ISSN 0162-8828. Citado na página 54.
- OLIVA, A.; SCHYNS, P. G. Coarse blobs or fine edges? evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive psychology*, v. 34, p. 72–107, 1997. Citado na página 38.
- OLIVA, A.; SCHYNS, P. G. Diagnostic colors mediate scene recognition. *Cognitive psychology*, Elsevier, v. 41, n. 2, p. 176–210, 2000. Citado 2 vezes nas páginas 38 e 39.
- PAPA, J. P.; FALCÃO, A. X.; ALBUQUERQUE, V. H. C. de; TAVARES, J. M. R. Efficient supervised optimum-path forest classification for large datasets. *Pattern Recognition*, v. 45, n. 1, p. 512 – 520, 2012. ISSN 0031-3203. Citado na página 119.
- PETROU, M.; SEVILLA, P. *Image processing: dealing with texture*. New jersey, USA: John Wiley & Sons Inc., 2006. ISBN 9780470026281. Citado na página 54.
- PIETIKÄINEN, M.; HADID, A.; ZHAO, G.; AHONEN, T. *Computer Vision Using Local Binary Patterns*. London, United Kingdom: Springer, 2011. (Computational Imaging and Vision, v. 40). ISBN 978-0-85729-747-1. Citado 2 vezes nas páginas 53 e 54.
- POTTER, M. C. Short-term conceptual memory for pictures. *Journal of experimental psychology: human learning and memory*, American Psychological Association, v. 2, n. 5, p. 509, 1976. Citado 2 vezes nas páginas 38 e 39.

RADY, S. Novel applications of intelligent systems. In: HADJISKI, M.; KASABOV, N.; FILEV, D.; JOTSOV, V. (Ed.). *Novel Applications of Intelligent Systems*. Cham: Springer International Publishing, 2016. p. 231–250. ISBN 978-3-319-14194-7. Citado na página 40.

RANGANATHAN, A.; DELLAERT, F. Online probabilistic topological mapping. *The International Journal of Robotics Research*, SAGE Publications, p. 0278364910393287, 2011. Citado na página 31.

REYNA, E. Arias-de. A cooperative localization algorithm for uwb indoor sensor networks. *Wireless Personal Communications*, v. 72, n. 1, p. 85–99, 2013. ISSN 1572-834X. Citado na página 30.

RITUERTO, A.; MURILLO, A.; GUERRERO, J. Semantic labeling for indoor topological mapping using a wearable catadioptric system. *Robotics and Autonomous Systems*, v. 62, n. 5, p. 685 – 695, 2014. ISSN 0921-8890. Special Issue Semantic Perception, Mapping and Exploration. Citado na página 31.

ROCHA NETO, A. *SINPATCO II: Novas Estratégias de Aprendizado de Máquinas para Classificação de Patologias da Coluna Vertebral*. Tese (Doutorado) — Departamento de Engenharia de Teleinformática, Universidade Federal do Ceará, 2011. Citado 5 vezes nas páginas 14, 67, 71, 72 e 89.

ROHMER S. P. N. SINGH, M. F. E. V-rep: a versatile and scalable robot simulation framework. In: *Proc. of The International Conference on Intelligent Robots and Systems (IROS)*. Chicago, USA: IEEE, 2013. Citado na página 85.

ROJAS, R. *Neural Networks: A Systematic Introduction*. New York, NY, USA: Springer-Verlag New York, Inc., 1996. ISBN 3-540-60505-3. Citado na página 75.

RUBIO, F.; MARTÍNEZ-GÓMEZ, J.; FLORES, M. J.; PUERTA, J. M. Comparison between bayesian network classifiers and {SVMs} for semantic localization. *Expert Systems with Applications*, v. 64, p. 434 – 443, 2016. ISSN 0957-4174. Citado na página 30.

SAEDAN, M.; LIM, C. W.; ANG, M. H. Appearance-based slam with map loop closing using an omnidirectional camera. In: *2007 IEEE/ASME international conference on advanced intelligent mechatronics*. Zurich, Switzerland: IEEE, 2007. p. 1–6. ISSN 2159-6247. Citado na página 31.

SCHINDLER, G.; METZGER, C.; STARNER, T. A wearable interface for topological mapping and localization in indoor environments. In: *Location- and Context-Awareness, Second International Workshop, LoCA 2006, Dublin, Ireland, May 10-11, 2006, Proceedings*. Dublin, Ireland: Springer, 2006. p. 64–73. Citado na página 30.

SCHYNS, P. G.; OLIVA, A. From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological science*, SAGE Publications, v. 5, n. 4, p. 195–200, 1994. Citado na página 38.

SIMONS, D. J.; LEVIN, D. T. Change blindness. *Trends in cognitive sciences*, Elsevier, v. 1, n. 7, p. 261–267, 1997. Citado na página 39.

- SIOUTIS, M.; TAN, Y. User indoor location system with passive infrared motion sensors and space subdivision. In: STREITZ, N.; MARKOPOULOS, P. (Ed.). *Distributed, Ambient, and Pervasive Interactions: Second International Conference, DAPI 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014. Proceedings*. Cham: Springer International Publishing, 2014. p. 486–497. Citado na página 30.
- SON, J.; KIM, S.; SOHN, K. A multi-vision sensor-based fast localization system with image matching for challenging outdoor environments. *Expert Systems with Applications*, v. 42, n. 22, p. 8830 – 8839, 2015. ISSN 0957-4174. Citado na página 30.
- SONG, X.; LI, X.; TANG, W.; ZHANG, W. A fusion strategy for reliable vehicle positioning utilizing {RFID} and in-vehicle sensors. *Information Fusion*, v. 31, p. 76 – 86, 2016. ISSN 1566-2535. Citado na página 27.
- SONG, Z.; JIANG, G.; HUANG, C. Theoretical and mathematical foundations of computer science: Second international conference, ictmf 2011, singapore, may 5-6, 2011. selected papers. In: ZHOU, Q. (Ed.). *Theoretical and Mathematical Foundations of Computer Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. p. 198–206. ISBN 978-3-642-24999-0. Citado 2 vezes nas páginas 27 e 30.
- SOUSA, R. G.; NETO, A. R. R.; CARDOSO, J. S.; BARRETO, G. A. Robust classification with reject option using the self-organizing map. *Neural Computing and Applications*, v. 26, n. 7, p. 1603–1619, 2015. ISSN 1433-3058. Citado 2 vezes nas páginas 29 e 89.
- SOUZA JÚNIOR, A. H. de; CORONA, F.; BARRETO, G. A.; MICHE, Y.; LENDASSE, A. Minimal learning machine: A novel supervised distance-based approach for regression and classification. *Neurocomputing*, v. 164, p. 34 – 44, 2015. ISSN 0925-2312. Citado 3 vezes nas páginas 14, 71 e 76.
- STONE, M. Cross-validatory choice and assessment of statistical predictions. *Roy. Stat. Soc.*, v. 36, p. 111–147, 1974. Citado na página 94.
- SUYKENS, J. A. K.; GESTEL, T. V.; BRABANTER, J. D. *Least squares support vector machines*. New Jersey, London, Singapore: World Scientific, 2002. ISBN 981-238-151-1. Citado na página 72.
- THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition, Fourth Edition*. 4th. ed. USA: Academic Press, 2008. ISBN 1597492728, 9781597492720. Citado 2 vezes nas páginas 65 e 66.
- VAPNIK, V. N. *Statistical Learning Theory*. Nova Jersey, EUA: John Wiley & Sons, Inc., 1998. Citado na página 67.
- VARSHAVSKY, A.; LARA, E. de; HIGHTOWER, J.; LAMARCA, A.; OTSASON, V. {GSM} indoor localization. *Pervasive and Mobile Computing*, v. 3, n. 6, p. 698 – 720, 2007. ISSN 1574-1192. PerCom 2007. Citado na página 30.
- VEDALDI, A.; FULKERSON, B. *VLFeat: An Open and Portable Library of Computer Vision Algorithms*. 2008. Citado na página 88.
- VIOLA, P.; JONES, M. J. Robust real-time face detection. *Int. J. Comput. Vision*, Kluwer Academic Publishers, Hingham, MA, USA, v. 57, n. 2, p. 137–154, maio 2004. ISSN 0920-5691. Citado na página 62.

WALKER, L. L.; MALIK, J. When is scene recognition just texture recognition. *Journal of Vision*, Citeseer, v. 2, n. 7, p. 255, 2002. Citado na página 38.

WANG, Z. L.; CAI, B. G.; YI, F. Z.; LI, M. Advances in automation and robotics, vol. 2: Selected papers from the 2011 international conference on automation and robotics (icar 2011), dubai, december 1–2, 2011. In: LEE, G. (Ed.). *Advances in Automation and Robotics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 593–601. ISBN 978-3-642-25646-2. Citado na página 27.

WILAMOWSKI, B. M.; IRWIN, J. D. *Intelligent Systems*. 2nd. ed. Boca Raton, FL, USA: CRC Press, Inc., 2011. ISBN 1439802831, 9781439802830. Citado na página 76.

WU, H.; TIAN, G. H.; LI, Y.; ZHOU, F. Y.; DUAN, P. Spatial semantic hybrid map building and application of mobile service robot. *Robotics and Autonomous Systems*, v. 62, n. 6, p. 923 – 941, 2014. ISSN 0921-8890. Citado 2 vezes nas páginas 40 e 41.

XU, R.; CHEN, W.; XU, Y.; JI, S. A new indoor positioning system architecture using gps signals. *Sensors*, v. 15, n. 5, p. 10074, 2015. Citado na página 27.

YANG, H.; SHAO, L.; ZHENG, F.; WANG, L.; SONG, Z. Recent advances and trends in visual tracking: A review. *Neurocomputing*, v. 74, n. 18, p. 3823 – 3831, 2011. ISSN 0925-2312. Citado na página 28.

YAYAN, U.; YUCEL, H.; YAZICI, A. A low cost ultrasonic based positioning system for the indoor navigation of mobile robots. *Journal of Intelligent & Robotic Systems*, v. 78, n. 3, p. 541–552, 2015. ISSN 1573-0409. Citado 2 vezes nas páginas 27 e 30.

ZANDIAN, R.; WITKOWSKI, U. Performance analysis of small size and power efficient uwb communication nodes for indoor localization. In: ALBOUL, L.; DAMIAN, D.; AITKEN, J. M. (Ed.). *Towards Autonomous Robotic Systems: 17th Annual Conference, TAROS 2016, Sheffield, UK, June 26–July 1, 2016, Proceedings*. Cham: Springer International Publishing, 2016. p. 371–382. ISBN 978-3-319-40379-3. Citado na página 30.

ZHANG, H.; LIU, Y.; TAN, J. Loop closing detection in rgb-d slam combining appearance and geometric constraints. *Sensors*, v. 15, n. 6, p. 14639, 2015. ISSN 1424-8220. Citado na página 40.

Apêndices

APÊNDICE A – Hiperparâmetros Adotados no Teste de Navegação

A Tabela 20 mostra os valores adotados para os hiperparâmetros dos classificadores utilizados no teste de navegação com o robô usando a câmera GoPro. De maneira semelhante, a Tabela 21 exibe os hiperparâmetros dos classificadores usado no robô com a câmera omnidirecional.

Tabela 20 – Valores dos hiperparâmetros dos classificadores utilizado no teste de navegação com o robô usando a câmera **GoPro**. CO significa o número de neurônios na camada oculta e K é a porcentagem dos dados de treinamento usando como pontos de referência.

	M. Espaciais	M. Centrais	M. de Hu	GLCM	LBP
LSSVM (Linear)	$C = 2048$	$C = 2048$	$C = 2048$	$C = 2048$	$C = 2048$
LSSVM (RBF)	$C = 2048; \sigma = 8$	$C = 2048; \sigma = 8$	$C = 2048; \sigma = 8$	$C = 2048; \sigma = 8$	$C = 2048; \sigma = 8$
SVM (Linear)	$C = 2048$	$C = 2048$	$C = 2048$	$C = 2048$	$C = 2048$
SVM (RBF)	$C = 2048; \sigma = 8$	$C = 2048; \sigma = 8$	$C = 2048; \sigma = 8$	$C = 2048; \sigma = 8$	$C = 2048; \sigma = 8$
MLP	$CO = 10$	$CO = 20$	$CO = 25$	$CO = 25$	$CO = 10$
MLM	$K = 100\%$	$K = 100\%$	$K = 100\%$	$K = 100\%$	$K = 100\%$
MLM-NN	$K = 100\%$	$K = 100\%$	$K = 100\%$	$K = 100\%$	$K = 100\%$

Tabela 21 – Valores dos hiperparâmetros dos classificadores utilizado no teste de navegação com o robô usando a câmera **omnidirecional**. CO significa o número de neurônios na camada oculta e K é a porcentagem dos dados de treinamento usando como pontos de referência.

	M. Espaciais	M. Centrais	M. de Hu	GLCM	LBP
LSSVM (Linear)	$C = 16384$	$C = 16384$	$C = 16384$	$C = 16384$	$C = 16384$
LSSVM (RBF)	$C = 16384; \sigma = 16$	$C = 16384; \sigma = 16$	$C = 16384; \sigma = 16$	$C = 16384; \sigma = 16$	$C = 16384; \sigma = 16$
SVM (Linear)	$C = 16384$	$C = 16384$	$C = 16384$	$C = 16384$	$C = 16384$
SVM (RBF)	$C = 16384; \sigma = 16$	$C = 16384; \sigma = 16$	$C = 16384; \sigma = 16$	$C = 16384; \sigma = 16$	$C = 16384; \sigma = 16$
MLP	$CO = 30$	$CO = 60$	$CO = 30$	$CO = 30$	$CO = 20$
MLM	$K = 100\%$	$K = 100\%$	$K = 100\%$	$K = 100\%$	$K = 100\%$
MLM-NN	$K = 100\%$	$K = 100\%$	$K = 100\%$	$K = 100\%$	$K = 100\%$