

KEVIN BRAATHEN DE CARVALHO

COOPERAÇÃO HUMANO-ROBÔ ATRAVÉS DE INTERAÇÕES POR AÇÕES

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

Orientador: Alexandre Santos Brandão

**VIÇOSA - MINAS GERAIS
2019**

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

C331c
2019 Carvalho, Kevin Braathen de, 1992-
Cooperação humano-robôs através de interações por ações /
Kevin Braathen de Carvalho. – Viçosa, MG, 2019.
71 f. : il. (algumas color.) ; 29 cm.

Orientador: Alexandre Santos Brandão.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 63-71.

1. Interação homem-máquina. 2. Robótica Redes neurais
(Computação). 3. Sistema de reconhecimento de padrões.
I. Universidade Federal de Viçosa. Departamento de Engenharia
Elétrica. Programa de Pós-Graduação em Ciência da
Computação. II. Título.

CDD 22. ed. 004.019

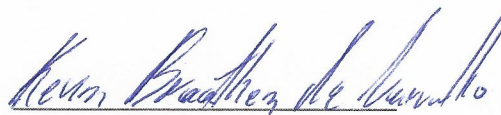
KEVIN BRAATHEN DE CARVALHO

COOPERAÇÃO HUMANO-ROBÔ ATRAVÉS DE INTERAÇÕES POR AÇÕES

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

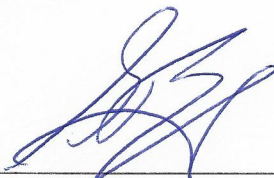
APROVADA: 18 de dezembro de 2019.

Assentimento:



Kevin Braathen de Carvalho

Autor



Alexandre Santos Brandão

Orientador

Dedico esse trabalho a todos que caminham comigo, se sou quem sou, devo isso à vocês.

Agradecimentos

A vida é uma caminhada e somos feitos daqueles que caminham conosco. Então eu agradeço todos que caminharam e que caminham comigo. Os passos só podem ser dados por mim, mas o apoio e o incentivo de quem me cerca tornam todos os fardos mais leves.

Primeiramente agradeço minha família por sempre estarem comigo. Em especial minha mãe Beth, por ter me ensinado tanto sobre a vida, sobre mim, sobre o que é amar, ser bom, gentil e carinhoso. E também por ter prontamente me ajudado com a revisão do trabalho.

Às minhas irmãs Karen e Stella, por terem me aturado quando ninguém deveria ter feito e por terem ficado por perto até que nós encontrássemos uma maneira de se amar como irmãos devem fazer. É muito bonito olhar para trás e ver onde chegamos juntos e o quanto que isso nos traz força pra seguir a vida.

Ao meu avô Per, por toda a inspiração que tive na minha vida toda, tenho certeza que meu passado, presente e futuro profissional tem uma influência enorme do senhor, por todos os ensinamentos sobre a vida, todos os incentivos diretos e indiretos sobre meus estudos. Serei eternamente grato pela sabedoria que o senhor humildemente sempre me conferiu.

À minha vó Honey, que com seu jeito único sempre fez questão de mostrar o melhor e mais puro carinho e amor que ela tem em si, seja sabendo a hora de cobrar, de abraçar e de conversar.

À minha tia Frankie, que sempre arrumou um tempo para um abraço e um conselho, e por me lembrar sempre que não precisamos deixar de brincar com o passar dos anos, isso é algo que sempre levarei comigo.

Ao meu tio Billy que em mais de uma ocasião largou o semblante brincalhão pra falar sério sobre a vida quando precisei, que me deu conselhos valorosos sobre vários setores da vida.

Ao meu orientador, Alexandre, por toda a luz que me conferiu nessa jornada. Por toda a paciência e confiança e por entender e respeitar os meus limites, me cobrando e incentivando de forma humana, isso é algo incomum de encontrar e fico feliz que tenha sido assim. O caminho que você seguiu na sua vida é de muita inspiração e digno de admiração.

Agradeço também meus amigos. Todos vocês. Se estão lendo isso sabem de quem eu falo. O amor que vem de vocês me fez enxergar a pessoa capaz que tem em mim mesmo quando os percalços da vida ofuscaram isso aos meus olhos, me fez encontrar forças e capacidade para continuar trilhando quando achei que não tinha mais. Eu realmente tenho

muita sorte de ter encontrado vocês nessa vida.

Aos dois Arthures da minha vida, Renan, Marco Antônio e Pedro Lanna, um agradecimento especial por terem me ajudado tão prontamente num momento que precisava com urgência de uma opinião externa sobre o trabalho.

Ao Vitor Thinassi, que por meses trabalhamos juntos em uma parte crucial do projeto. Sem sua colaboração esse trabalho teria demorado uma infinidade para ser realizado. Sua dedicação, esforço e paciência comigo foram muito importantes para tudo. Espero que tenha aprendido tanto quanto eu nesses meses de cooperação.

Aos Departamentos de Zootecnia e Engenharia Agrícola pelo apoio cedendo equipamentos necessários para várias etapas ao longo do mestrado.

À CAPES, pelo incentivo através da bolsa de estudos concedida nesse mestrado.

Por fim gostaria de agradecer a todo o povo do NERO que me ajudou demais, seja escutando minhas falações, reclamando de erros, ou só jogando papo fora mesmo. Foi um prazer trabalhar com todos. Certamente o dia-a-dia ficou mais leve tendo vocês lá.

É reconfortante saber que essa vitória não é só minha.

A felicidade só é real quando é compartilhada.

-Alexander Supertramp

Resumo

CARVALHO, Kevin Braathen de, M.Sc., Universidade Federal de Viçosa, dezembro de 2019. **Cooperação Humano-Robô através de Interações por Ações**. Orientador: Alexandre Santos Brandão.

A Interação Humano-Robô vem ganhando mais espaço no meio acadêmico nos últimos anos. Esta é uma área de interesse devido ao seu grande leque de aplicações, tais como busca e resgate, entretenimento, auxílio militar e policial, exploração espacial, tarefas com veículos aéreos não tripulados e robótica socialmente assistiva. O último item pode ser mais detalhado com aplicações como ajuda a idosos, pessoas física ou mentalmente debilitadas, e até mesmo suporte educacional. Esse trabalho propõe um método de reconhecimento de ações para ser utilizado como via de comunicação com uma plataforma robótica para tarefas de robótica socialmente assistiva, tais como auxiliar o usuário em carregar peso, sendo capaz de segui-lo e de ir para pontos predeterminados, enquanto desvia de obstáculos. O método proposto para interação é baseado em redes neurais artificiais e se apoia em redução da dimensionalidade da entrada pelo uso de autovalores. A base de dados utilizada é própria, devido ao fato das bases de dados encontradas na internet não possuírem classes de ações desejadas para a aplicação proposta. A validação do método se dá através de comparação com o DTW (*Dynamic Time Warping*), tanto em desempenho simples quanto em escalabilidade. Testes *online* são realizados para assegurar que a base de dados não estava enviesada ou tendenciosa. Os resultados experimentais mostraram que o método de comunicação é eficaz, tendo mais de 99% de acerto nos testes *online*, para 5 pessoas diferentes na sua base de dados com 10 amostras de cada pessoa para cada classe. A estratégia é escalável para novos usuários e requer um número reduzido de 5 amostras de treinamento de cada pessoa para cada classe para alcançar taxas de acerto superiores a 96%.

Palavras-chave: Reconhecimento de Ações. Interação Humano-Robô. Robótica Socialmente Assistiva. Redes Neurais.

Abstract

CARVALHO, Kevin Braathen de, M.Sc., Universidade Federal de Viçosa, December 2019. **Human-Robot Cooperation Through Interactions Via Actions**. Advisor: Alexandre Santos Brandão.

Human-Robot Interaction has been gaining more space in academia in recent years. This is an area of interest due to its wide range of applications such as search and rescue, entertainment, military and police assistance, space exploration, unmanned aerial vehicle tasks and socially assisted robotics. The last one may be more detailed with applications such as helping the elderly, the physically or mentally disabled people, and even educational support. This paper proposes an action recognition method to be used as a communication link with a robotic platform for socially assisted robotics tasks, such as assisting the user in carrying weight, being able to follow him/her and to go to predetermined points while avoiding obstacles. The proposed method for interaction is based on artificial neural networks and relies on reducing the dimensionality of the input by using eigenvalues. The database used is its own, due to the fact that the databases found on the Internet do not have desired classes of actions for the proposed application. Validation of the method is by comparison with DTW (Dynamic Time Warping), in both simple performance and scalability. Online tests are performed to ensure that the database was not biased. Experimental results showed that the communication method is effective, with more than 99% of successfulness, for 5 different people in the database with 10 samples per person for each class. The strategy is scalable to new users and requires a small number of 5 training samples per person per class to achieve over 96% accuracy.

Keywords: Action Recognition. Human-Robot Interaction. Socially Assistive Robotics. Neural Networks.

Lista de ilustrações

Figura 1 – Exemplos de aplicações de HRI.	15
Figura 2 – Exemplos de tipos de HRI.	16
Figura 3 – Ilustrações da natureza da comunicação em HRI.	17
Figura 4 – Exemplos de Aplicação de SAR.	22
Figura 5 – Ilustração da terminologia utilizada	28
Figura 6 – Alinhamento temporal de duas séries.	30
Figura 7 – Matriz de custo de duas sequências, X na vertical e Y na horizontal. As regiões mais escuras tem custo mais baixo e as mais claras custo mais alto.	31
Figura 8 – Caminho de menor custo para as Sequências X e Y	32
Figura 9 – Exemplos de caminho para uma sequência com 9 e outra com 7 itens.	33
Figura 10 – Neurônio Artificial.	34
Figura 11 – Funções de Ativação.	35
Figura 12 – Critério de parada no treinamento em função da validação.	36
Figura 13 – Ilustração das Zonas Sociais.	38
Figura 14 – Esquema do modelo cinemático do robô	40
Figura 15 – Esquema do Desvio Tangencial.	41
Figura 16 – <i>Kinect</i> v2.0.	43
Figura 17 – Esquema de detecção de luz estruturada.	43
Figura 18 – <i>Pioneer 3DX</i>	44
Figura 19 – Ações padrão da base de dados.	45
Figura 20 – Juntas de destaque para a etapa de classificação.	47
Figura 21 – Fluxograma da classificação por DTW.	48
Figura 22 – Fluxograma da classificação por Redes Neurais.	48
Figura 23 – Classes de ação padrão e seus respectivos lugares no Jogo da Velha.	57
Figura 24 – Experimento 1 Jogo da Velha.	58
Figura 25 – Experimento de Navegação	59
Figura 26 – Experimento de navegação aplicada.	61

Lista de tabelas

Tabela 1 – Resultados das simulações do DTW para a simulação 1.	51
Tabela 2 – Resultados das simulações da RNA para a simulação 1.	51
Tabela 3 – Resultados das simulações do DTW para a simulação 2.	53
Tabela 4 – Resultados das simulações da RNA para a simulação 2.	53
Tabela 5 – Resultados da simulação para o método proposto com RNA para a simulação 3.	55
Tabela 6 – Resultados para os testes <i>online</i>	56

Sumário

1	INTRODUÇÃO	13
1.1	O Problema e sua Importância	22
1.1.1	Objetivos	23
1.2	Trabalhos Correlatos	23
1.3	Contribuições	25
1.4	Estrutura do Trabalho	26
2	FUNDAMENTAÇÃO TEÓRICA	28
2.1	O Reconhecimento de Ações	28
2.1.1	Reconhecimento de Ações	29
2.2	<i>Dynamic Time Warping</i> - DTW	30
2.3	Redes Neurais Artificiais - RNA	34
2.4	Zonas Sociais	37
2.5	A Navegação do Robô	39
2.5.1	Modelagem e Controle	39
2.5.2	Desvio de Obstáculos	40
3	METODOLOGIA	42
3.1	Materiais e Métodos	42
3.1.1	O Sensor <i>Kinect</i>	42
3.1.2	<i>Pioneer 3-DX</i>	43
3.2	Base de Dados	44
3.3	<i>Dynamic Time Warping</i> - DTW	47
3.4	Redes Neurais Artificiais - RNA	47
4	EXPERIMENTOS E RESULTADOS	49
4.1	Simulações	50
4.1.1	Simulação 1	50
4.1.2	Simulação 2	52
4.1.3	Simulação 3	54
4.2	Experimentos	54
4.2.1	Experimento 1	55
4.2.2	Experimento 2	56
4.2.3	Experimento 3	57
5	CONSIDERAÇÕES FINAIS	62

REFERÊNCIAS 63

1 Introdução

A Interação Humano-Robô (do inglês *Human-Robot Interaction*, HRI) tem continuamente ganhado mais atenção no meio acadêmico nos últimos 30 anos (TSAROUCHI; MAKRIS; CHRYSOLOURIS, 2016). Este é um campo de estudo importante devido à sua grande gama de aplicações, tais como busca e resgate, entretenimento, auxílio militar e policial, exploração espacial ou tarefas com veículos aéreos não tripulados (VANT).

Busca e resgate é uma aplicação que inerentemente se tem interação humano-robô, tendo em vista que o agente está procurando um humano em uma situação de desastre. No trabalho de Murphy et al. (2008) para resumir a situação de robôs de busca, ele ambienta essa aplicação com as características básicas de um desastre e como que isso afeta o projeto de robótica e, em seguida, descreve os robôs utilizados em desastres famosos como o atentado ao *World Trade Center* em 2001, os furacões Katrina, Rita e Wilma, em 2005, nos Estados Unidos. Além disso, os projetos mais promissores, conceitos e métodos de avaliação de desempenho são elucidados. Por fim conclui-se sobre as dificuldades fundamentais dessa aplicação e apresenta os problemas em aberto.

Para aplicações em entretenimento, o trabalho de Yoshida et al. (2015) propõe uma interface simples que permita a usuários criarem padrões de movimentos complexos para um robô dançar. Suas maiores contribuições são: a interface desenvolvida é capaz de criar movimentações elaboradas sem a necessidade do usuário dar comandos de baixo nível, permitindo que usuários especializados em robótica ou não consigam operar o robô; a possibilidade de criação de padrões complexos através de movimentos simples predeterminados e a sincronização automática entre a dança e os movimentos do robô. O trabalho foi desenvolvido para ser utilizado com dois tipos de robôs; um móvel e um humanoide. Os autores fizeram duas demonstrações públicas da interface que validou a facilidade da aplicação para leigos e especialistas no assunto.

Uma análise sobre o presente e o futuro de aplicações militares da robótica é vista em (MIES, 2010), onde o autor fala das primeiras tecnologias utilizadas nesse cenário e como elas podem ser úteis. As aplicações são aquelas que comumente colocam em risco a vida de humanos, tais como desarmar uma bomba ou andar em lugares com escombros em perigo de desabamento. Tais missões podem ser realizadas com segurança através de agentes robóticos tele-operados. Para o futuro, o autor menciona o desenvolvimento mais profundo de robôs terrestres nas aplicações de exploração e auxílio médico, bem como de robôs marítimos com o mesmo intuito. Por fim, é mencionado que especialistas acreditam que a robótica espacial é um campo de atuação para o futuro, dado que enviar humanos é ainda muito caro, em função dos riscos.

Exploração espacial é um tópico que a interação humano-robô se dá comumente por teleoperação em algum grau do robô. O trabalho (YOSHIDA, 2009) faz um apanhado geral das aplicações de robótica no espaço. Inicialmente, é mencionado os avanços em robótica orbital, com aplicações de remoção de entulho espacial através da mudança da órbita de satélites, ou utilização de robôs para a realização de algum serviço, como manutenção ou manipulação de algo em um satélite. Em seguida, é discutida a robótica planetária, exemplificando aplicações na Lua e em Marte e explicitando dificuldades recorrentes como o sensoriamento da odometria do agente e a compensação para deslizos e controle de tração. Por fim, as dificuldades de aplicações em asteroides são trabalhadas, como a aderência e locomoção do robô em lugares com campo gravitacional muito diferente da Terra.

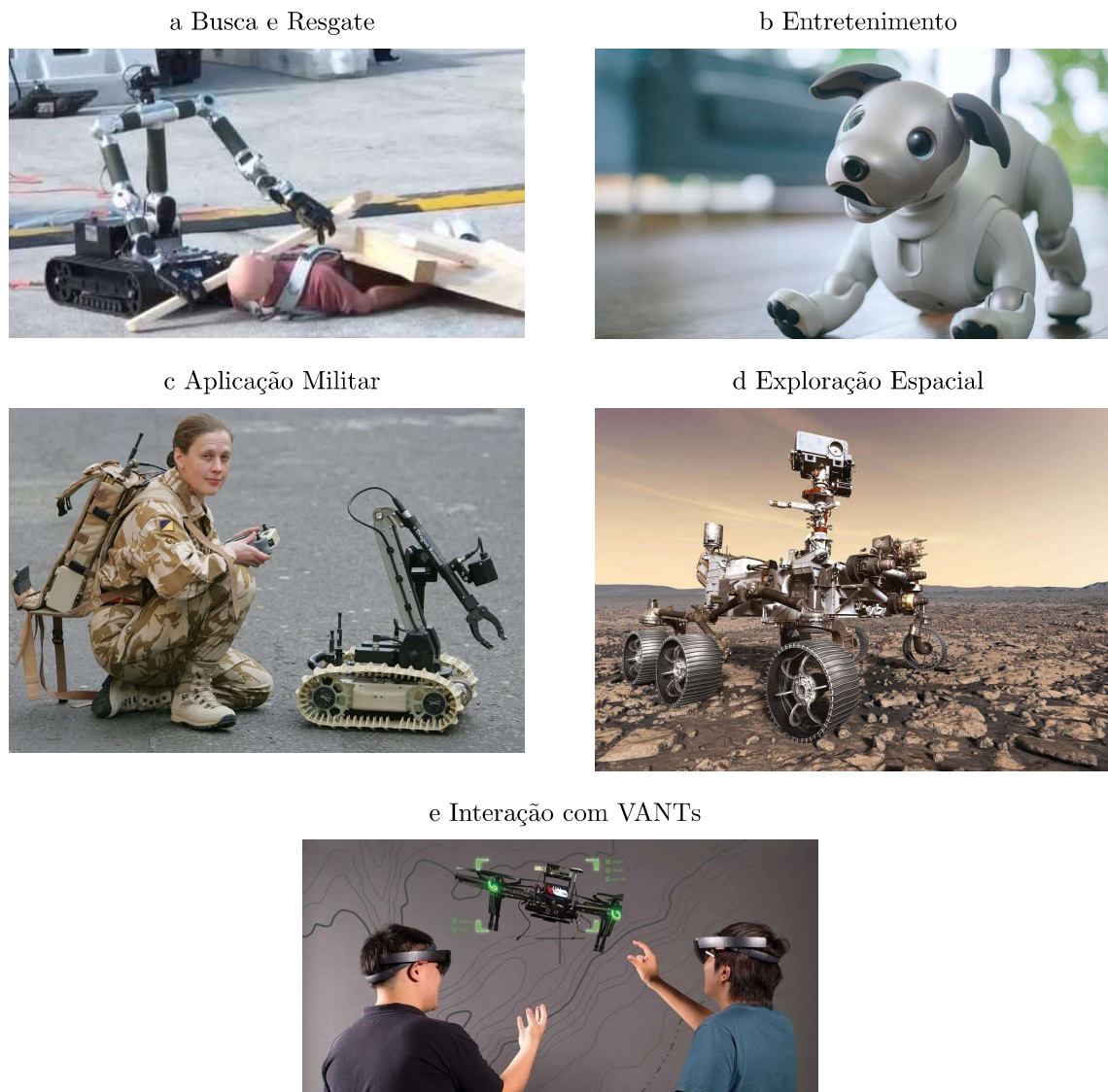
Em (MONAJJEMI et al., 2013), pode-se ver o desenvolvimento de um sistema de interação humano-robô através de gestos e de detecção de onde o usuário está olhando, a fim de comandar um grupo de VANTs. O sistema implementado permite que o usuário consiga manipular os VANTs individualmente ou adicionar e removê-los de grupos para que um mesmo comando seja dado para mais de um dos agentes. O trabalho não abrange a execução de tarefas em si, apenas a seleção de quais VANTs irão receber os comandos de se juntar ao grupo ou se separar. O sistema de comunicação consegue ser utilizado por usuários não especializados de forma intuitiva, onde se seleciona o VANT que receberá o comando olhando para ele e os comandos de se juntar, separar ou executar a ação são dados por gestos.

A Figura 1 ilustra exemplos das aplicações supracitadas.

Segundo Goodrich, Schultz et al. (2008), a interação entre humanos e robôs pode se agrupar nas seguintes categorias com alguns exemplos de trabalhos correlatos:

1. HRI Colaborativo, que envolve um robô e uma pessoa trabalhando juntos para completar uma tarefa simples (PARK; HOWARD, 2013; HINDS; ROBERTS; JONES, 2004; LÄNGLE; WÖRN, 2001; CHERUBINI et al., 2016; MAURICE et al., 2017).
2. HRI Assistivo, onde o robô é utilizado para auxiliar pessoas através de assistência física, social e/ou cognitiva (MCCOLL; NEJAT, 2013; FEIL-SEIFER; MATARIĆ, 2011; CLABAUGH et al., 2018).
3. HRI de Imitação, onde o robô ou pessoa tenta imitar comportamentos verbais ou não verbais de quem está interagindo. Isso pode ser utilizado para melhorar coordenação social entre as partes envolvidas (BOURGEOIS; HESS, 2008; RAKITA; MUTLU; GLEICHER, 2017).
4. HRI Geral ou de multi-tarefas, que envolve robôs desenvolvidos para interagir com pessoas em várias aplicações distintas utilizando comunicação bidirecional (ISHIGURO et al., 2001)

Figura 1 – Exemplos de aplicações de HRI.



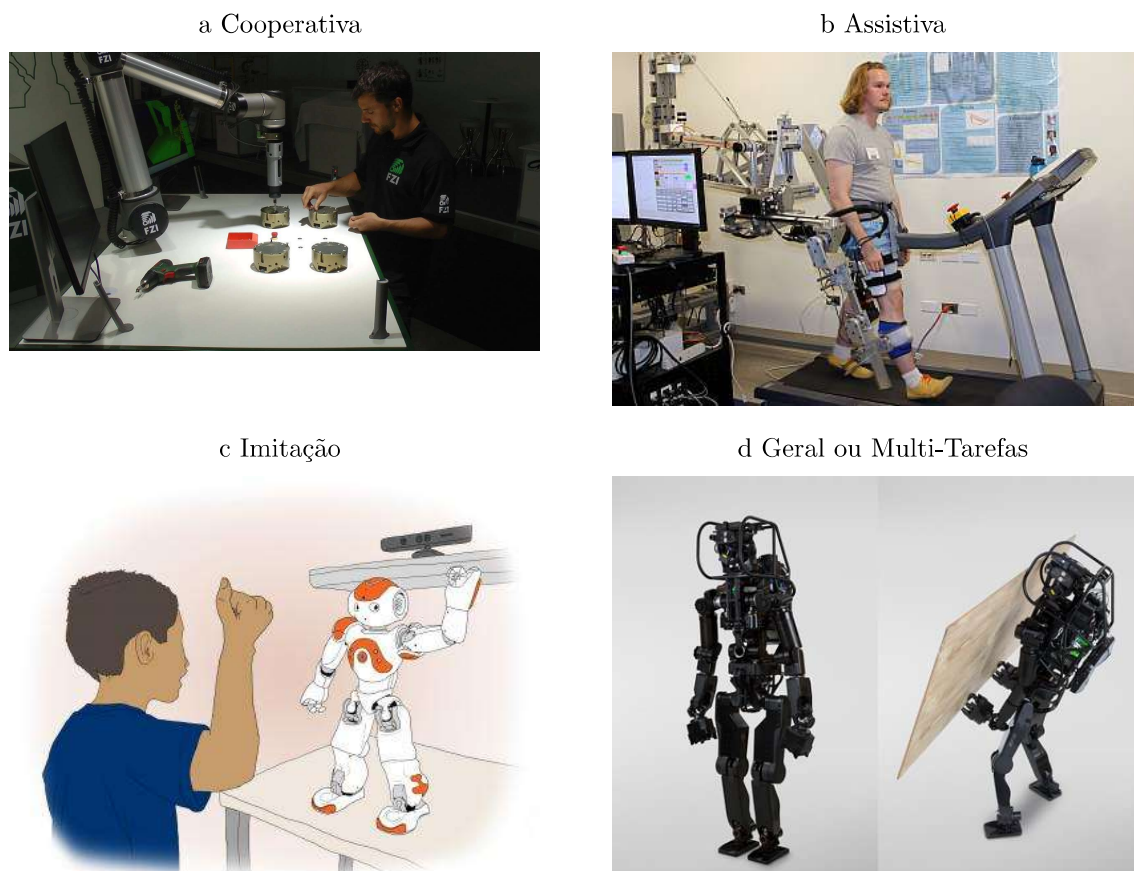
Fontes: [Seabotix \(2019\)](#), [Wxyz Detroit \(2019\)](#), [Grit Daily \(2019\)](#), [Independent UK \(2019\)](#), [Hololens \(2019\)](#)

Por sua vez, a Figura 2 ilustra exemplos dos tipos de interação humano-robô.

Sobre a natureza da comunicação ou interação humano-robô, pode-se destacar diferentes abordagens, tais como:

1. **Gestos ou ações**, adotados quando a liberdade de movimento é uma opção. Ademais, pode ser utilizado em situações tais como fisioterapia, onde o reconhecimento da posição e orientação de partes do corpo é vital, reafirmando a comunicação entre o usuário e o agente ([KANDA et al., 2004](#); [CHEN et al., 2019](#); [BODIROŽA](#); [STERN](#); [EDAN, 2012](#)).
2. **Comandos de voz**, através da linguagem natural utilizada regularmente no dia-a-

Figura 2 – Exemplos de tipos de HRI.



Fontes: [KolRob \(2019\)](#), [TechBriefs \(2019\)](#), [EurekAlert \(2019\)](#), [Phys Org \(2019\)](#).

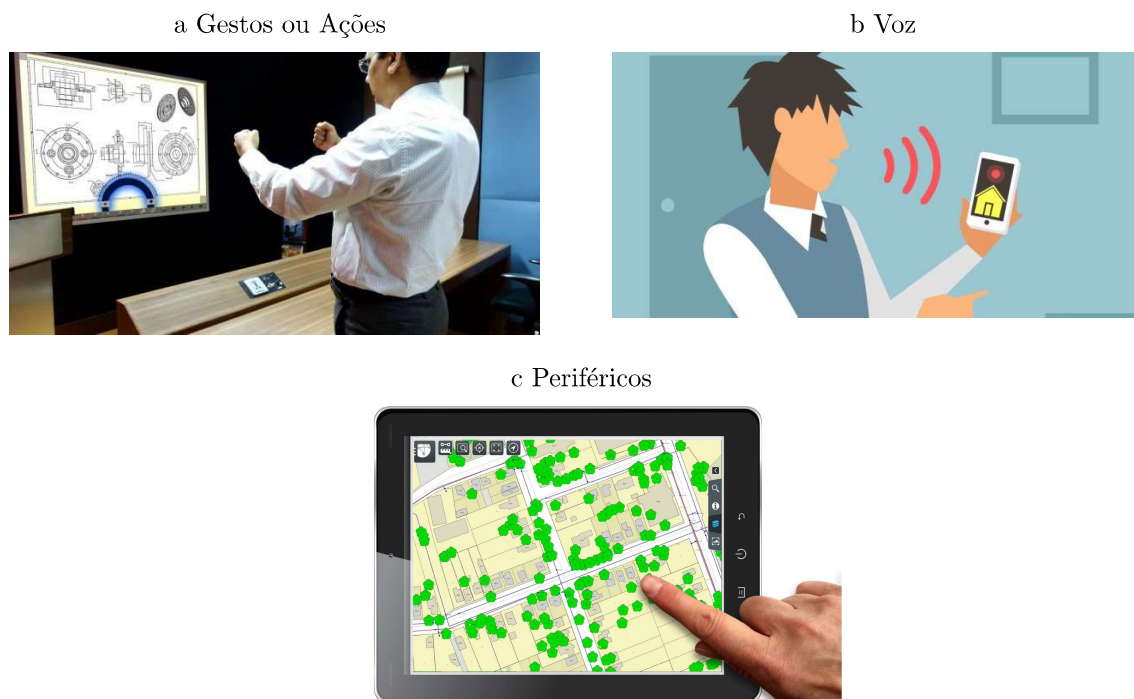
dia para; neste caso, para dialogar com o robô ([ERIKSSON, 2004](#); [LIU; ZHANG, 2019](#)).

3. **Periféricos**, tais como mouse, teclados ou *touchscreens*, os quais podem ser úteis em situações que gestos ou reconhecimento de voz seriam tediosos ([HUTTENRAUCH; EKLUNDH, 2002](#); [MONTEMERLO et al., 2002](#)).

Na Figura 3 se encontram ilustrações dos exemplos da natureza da comunicação supracitados.

Esse trabalho tem foco em uma parte específica de HRI, definida como Robótica Socialmente Assistiva (SAR), que é a intersecção entre Robótica Socialmente Interativa (do inglês, *Socially Interactive Robotics*, SIR) e a Robótica Assistiva (do inglês *Assistive Robotics*, AR). No caso, a SIR é inspirada na comunicação entre o robô e o ambiente que ele está inserido, ou até mesmo entre mais de um robô ([FONG; NOURBAKHSH; DAUTENHAHN, 2003](#); [FEIL-SEIFER; MATARIC, 2005](#)), e a AR é aquela dedicada a assistir humanos de alguma forma ([FEIL-SEIFER; MATARIC, 2005](#)). Abaixo será

Figura 3 – Ilustrações da natureza da comunicação em HRI.



Fontes: [AV Magazine \(2019\)](#), [Medium \(2019\)](#), [Bentley \(2019\)](#).

trabalhado com mais detalhes essas três áreas e também exemplificando trabalhos com suas aplicações.

Sobre robótica assistiva, o trabalho ([LOOS; REINKENSMeyer; GUGLIEMELLI, 2016](#)) traz um panorama geral de robôs de reabilitação, definindo inicialmente que esse setor desenvolve robôs que provêm terapia para pessoas a procura de recuperar suas funções físicas, sociais, comunicativas ou cognitivas e/ou assiste pessoas que têm uma debilitação crônica para a realização de atividades do cotidiano. Após um breve apanhado histórico do tópico, o autor divide essas aplicações em quatro grupos e discorre sobre suas dificuldades e os atuais avanços.

O primeiro grupo é a terapia de reabilitação, onde um dos grandes desafios é explorar o fato do sistema neuromuscular humano ter a característica de plasticidade ligada ao uso, isto é, o uso de músculos e neurônios altera suas propriedades, incluindo os padrões de conexão. Avanços nesse setor podem ser citados como o *Lokomat* da empresa *Hocoma*, que é um exoesqueleto para ser utilizado durante caminhadas em esteiras para auxílio na reabilitação de caminhada; o *Bi-Manu Track*, que é um sistema simples com resultados promissores que consiste em dois motores, um para cada mão, para ajudar na reabilitação de flexão de pulso e também na pronação e supinação do antebraço.

O segundo grupo é o auxílio para pessoas com debilitações. Quanto maior for a debilitação da função que o paciente está perdendo, maior o custo da intervenção técnica.

Cada debilitação pode ser altamente pessoal, então o maior desafio é tornar algo viável a personalização do aparelho para o usuário. Outro limitante relevante é a autonomia do robô, afinal, se a debilitação física for grande, ou ela ter algum nível cognitivo, isso implica que o robô deve ter um sensoriamento que torne sua assistência, suficientemente segura e capaz de coexistir com os humanos que o utilizam. Aparelhos nesse setor incluem cadeiras de roda motorizadas com ou sem manipuladores robóticos para ampliar o espaço de trabalho do usuário e exoesqueletos que permitam a locomoção de usuários com danos da coluna vertebral.

O terceiro grupo é o de próteses e ortoses inteligentes, em palavras mais simples, membros artificiais. Existem diversos empecilhos para o desenvolvimento nessa área, como desenvolver um meio intuitivo do usuário manipular robôs com vários graus de liberdade; os robôs ainda não estão pareados com o braço humano em termos de alcance da força, peso e duração com uma fonte de energia portátil; os membros humanos são permeados por sensores táteis, instalá-los em um membro artificial de forma que o usuário possa usá-los de maneira intuitiva é desafiador. Avanços interessantes incluem interfaces neurais para a atuação dos membros artificiais, com algumas abordagens não invasivas tendo bastante sucesso, indicando que no futuro os sistemas de controle de próteses inteligentes terão a opção de se apoiar numa interface direta com o cérebro do usuário.

O último grupo é o de ampliação para diagnóstico e monitoramento. Um dos aspectos críticos da reabilitação é a manutenção da saúde com o decaimento relacionado à idade ou à degeneração após intervenção médica. Tecnologias de sensoriamento para o cotidiano do paciente são importantes para ter informações de como o tratamento está evoluindo. Desafios para a área são que os sensores que são vestidos, devem ser compatíveis com a pele do usuário, então para um futuro próximo será necessária uma incorporação melhor desses elementos sensoriais. O autor menciona que vários sensores incorporados em camisas parecem promissores. Por fim, ele menciona exemplos do estado da arte nesse ramo, como casas inteligentes com sensoriamento para saber o estado do paciente, e com atuadores para modificar as condições ambientais para aumento do conforto desse paciente.

Robótica Socialmente Interativa é abordada de forma geral nos trabalhos (DAUTENHAHN; FONG; NOURBAKSH, 2014; FONG; NOURBAKSH; DAUTENHAHN, 2003) onde se apresenta os tipos de classes sociais para robótica.

Socialmente Evocativo é quando o robô deve se assemelhar mais à humanos, emulando emoções e tendo aparência humanoide, e com isso ele tira vantagem da tendência que humanos tem de antropomorfizar e ter sentimentos estimulados, quando ficam envolvidos com um agente próximo à eles.

Interface Social, que deve prover uma interação através de deixas sociais similares aos humanos, como contato visual e expressão de emoções, tais como raiva, medo e afins. Esses modelos em geral demandam um nível de cognição social mais raso.

Socialmente Receptivo é um robô socialmente passivo, mas que pode se beneficiar através de aprendizado por imitação. Robôs assim devem ter competências sociais mais profundas que aqueles da classe de Interface Social.

Sociável é aquele agente que deve agir proativamente com humanos para satisfazer necessidades sociais, como emoções. Esses robôs precisam de modelos sociais profundos para atingir seu propósito.

Socialmente Situado envolve robôs cercados de ambientes sociais, onde são capazes de perceber e reagir. Esses agentes devem ser capazes de entender e perceber outros agentes sociais e vários objetos no ambiente.

Socialmente Embutido diz respeito a robôs que estão ou situados em um ambiente social e interagem com outros agentes e humanos, ou estruturalmente acoplado ao ambiente social ou ao menos parcialmente consciente das estruturas de interação dos humanos.

Socialmente Inteligente discute sobre robôs que mostram aspectos da inteligência social humana baseados em modelos profundos de cognição humana e competência social.

Para os autores, robôs socialmente interativos devem ser capazes de algumas das seguintes habilidades: expressar ou perceber emoções, comunicar com um alto nível de diálogo, aprender ou reconhecer os modelos de outros agentes, estabelecer e manter relações sociais, usar ou perceber deixas sociais naturais (olhar, gestos, etc), exibir personalidade e aprender ou desenvolver competências sociais.

Como previamente mencionado, a intersecção das duas áreas previamente mencionadas, Robótica Assistiva e Robótica Socialmente Interativa é a chamada Robótica Socialmente Assistiva. O artigo ([TAPUS; MAJA; SCASSELLATTI, 2007](#)) dá um panorama geral de objetos de estudo dessa área, como cuidado de idosos, uma preocupação crescente tendo em vista que a expectativa média de vida mundial tem aumentado; auxílio em tarefas físicas, reabilitação ou pessoas com necessidade de treinamento, onde o autor ressalta que SAR pode fornecer métodos inovadores de monitoramento, motivação e guia; auxílio para pessoas com debilitações cognitivas, nesse trabalho é ressaltado que SAR pode ser usado como um suporte social de forma a ensinar habilidades sociais gradualmente para crianças com autismo. Na sequência do artigo, [Tapus, Maja e Scassellatti \(2007\)](#) fala sobre os focos de estudo dessa área, explicados nos parágrafos seguintes.

A personificação do robô tem um papel chave na efetividade da assistência, dado que pessoas atribuem intenções, objetivos, emoções e personalidades até mesmo para máquinas com noções simples de movimento ou forma que possam ser relacionados com algo vivo.

Personalidade é também determinante importante nas relações sociais humanas, dado que pesquisas mostram uma ligação direta entre personalidade e comportamento ([CHRISTENSEN; PACCHIEROTTI, 2005](#)). Com robôs sendo assistentes ou companheiros

para humanos, desenvolver a personalidade do robô dentro do escopo da sua aplicação é uma motivação importante. Para uma maior efetividade, é importante que a personalidade que o robô transmite seja compatível com a do usuário. Não se tem uma definição genérica de personalidade, mas é comum de vê-la como o padrão coletivo de traços mentais, de comportamento, temperamento e emoções que o indivíduo apresenta consistentemente ao longo do tempo e de eventos diferentes.

Empatia é a característica mais determinante em aplicações centradas em um paciente, onde o robô auxilia ou cuida dele. Esse atributo é chave na interação social pois ele implica na compreensão da situação do outro aliado à um entendimento geral de emoções. Robôs não sentem empatia, mas é possível criar padrões de comportamento que emitem sinais de empatia para emular isso para o paciente. Isso pode ser feito com o sistema sendo capaz de reconhecer o estado emocional do usuário, comunicando com ele ou então mostrar traços da emoção que o usuário está sentindo.

Engajamento é um atributo que se caracteriza pela capacidade do robô estabelecer e manter uma conexão colaborativa com o usuário. Isso implica que o robô deve ser capaz de perceber a presença humana e também entender quando o humano quer interagir. Em algumas aplicações também é importante que o robô seja capaz de chamar atenção para si, para que a comunicação soe natural.

Adaptação remete a entender como se comunicar e ser capaz de adaptar o comportamento às mudanças nas interações sociais. Encorporar capacidades sofisticadas de adaptação social em robôs tem se provado uma tarefa muito difícil, então as abordagens tem focado nas mudanças a curto prazo, que denotam diferenças do indivíduo, e a longo prazo que permitem que a interação continue a ser engajadora ao longo do período de meses ou até anos dependendo da aplicação. Entender o humor e as emoções atreladas à personalidade do usuário pode ser bastante útil nessa tarefa.

Transferência, em muitas aplicações de SAR, é mudança de comportamento de longo prazo, seja estimulando a capacidade de contar histórias de um idoso a sua família, ou seja facilitando a interação social de uma criança autista. Então uma das características mais importantes em aplicações dessa natureza é conseguir transferir habilidades para o usuário a fim de gerar mudanças efetivas em sua vida. A capacidade de transferência de um sistema é muitas vezes utilizada como métrica de sucesso para a aplicação.

No trabalho (MARTINEZ-MARTIN; CAZORLA, 2019) é ressaltado que com o aumento da expectativa de vida, a necessidade de monitoramento da capacidade cognitiva e da saúde física de idosos têm se tornado mais presentes. Nesse contexto, atividades físicas podem beneficiar significativamente a saúde geral do paciente. Então, os autores propõem uma plataforma robótica para estimular os exercícios físicos. A plataforma lembra o paciente de fazer os exercícios, acessa uma lista de exercícios previamente gerada por um fisioterapeuta, descreve o exercício que o paciente deve fazer e, através de reconhecimento

de ações, o robô identifica se o paciente está executando de forma correta e emite um diagnóstico sobre seu desempenho, armazenando no seu histórico para análises futuras. A personificação do robô é humanoide, com dois braços e uma cabeça, para fortalecer a conexão entre ele e o paciente.

Meyer e Fricke (2017) mostram um estudo de caso de um robô desenvolvido para andar nos corredores de um hospital, enquanto é capaz de interagir verbalmente e auxiliar idosos em terapia de recuperação de caminhada. O robô não é humanoide, mas é equipado com um par de olhos, para poder aumentar a interação através de contato visual. A presença do robô teve como resultado o aumento da autoconfiança do paciente na realização dos exercícios, maior satisfação dos usuários dado uma variedade maior de exercícios disponíveis e mais motivação para treinar, chegando ao cenário da maioria dos pacientes preferirem treinar com o robô a treinar sem ele.

Robótica Socialmente Assistiva tem mostrado resultados promissores ao lidar com crianças com autismo. Taheri et al. (2019) utilizam o ensino básico de música para diferentes pontos do espectro autista como ferramenta de melhoria de habilidades motoras, cognitivas e sociais. O robô utilizado, NAO H-21, tem forma humanoide, capacidade de mudar a cor dos olhos e apresentar prosódia na fala, para maior familiarização da criança. Os resultados mostraram que o tratamento tem efeitos diferentes baseado na idade, personalidade e grau do espectro autista da criança, mas que ainda assim mostrou melhoras promissoras na capacidade social, cognitiva e motora dos pacientes.

Também é possível de se ver a aplicação de robótica no ensino. Benitti (2012) faz uma revisão de 10 artigos de utilização de robótica para ensino. Ele observa que em 90% dos casos, o robô utilizado é LEGO, dado sua facilidade de utilização. Os tópicos abordados para ensino dentre os trabalhos revisados foram resolução de problemas de lógica, matemática, princípios básicos da evolução, leis de Newton, conhecimento em ciência, engenharia e tecnologia, compreensão de sistemas, técnicas de raciocínio, comunicação social e programação de computadores. Na maioria dos casos, a presença de robótica aumentou o aprendizado dos indivíduos analisados.

Ospennikova, Ershov e Iljin (2015) fazem a importante separação de três formas que a robótica pode ser utilizada no processo educacional. Primeiro, como objeto de estudo, onde os alunos estudam robótica de fato, a história da robótica, aplicações, como que os robôs conseguem “enxergar” o mundo através de diversos tipos de sensores. Segundo, como ferramenta de cognição, onde se utiliza robótica para resolução de problemas, com conjuntos de robôs como o *Mindstorm* da LEGO ou então ensinando os alunos a modelarem como o robô reage a sinais de controle, seja em ambiente virtual ou real. E, por fim, como meio de ensino e desenvolvimento, de forma que o robô contribua para mostrar um fenômeno de forma mais palpável para o aluno, tendo o robô como ferramenta de demonstração de algum processo. No próprio artigo se encontra o exemplo de uma mini planta térmica

regulada por um robô. Os autores concluem dizendo que robótica é uma peça fundamental no ensino politécnico, mas que deve ser inserida de forma orgânica no ensino.

A Figura 4 mostra exemplos das aplicações da robótica socialmente assistiva.

Figura 4 – Exemplos de Aplicação de SAR.

a Auxílio em tarefas físicas



b Auxílio à crianças do espectro autista



c Auxílio educacional



Fontes: [Mada at Portal \(2019\)](#), [TeKnoKultura \(2019\)](#), [Apple Insider \(2019\)](#).

Um conceito importante que será trabalhado com mais detalhes posteriormente, é o de zonas sociais. Ao navegar interagindo com humanos, o robô deve levar em conta a proximidade dele com as pessoas que ele está interagindo. Esse estudo é chamado de *Proximics* e almeja entender a dinâmica das distâncias sociais utilizadas nas relações humanas. Para ter uma interação socialmente agradável, o agente não deve se aproximar do usuário de forma a causar desconforto, isto é, deve-se manter uma distância compatível com a interação vigente.

1.1 O Problema e sua Importância

Carregar objetos de um lugar para o outro pode ser uma tarefa cansativa, tediosa ou inviável para um dado usuário. Um robô pode auxiliar nessa tarefa, tanto para pessoas com debilitações físicas ou não. Dado que comunicação por gestos ou ações pode tornar a interação mais direta e natural ao se comparar com a utilização de algum periférico, a

integração de um sistema de identificação de gestos ou ações com uma plataforma robótica móvel pode ser de grande utilidade.

Utilizar esse agente para receber comandos do usuário como segui-lo, se aproximar dele ou ir para lugares predeterminados, configura uma interação social. Dado que há comunicação, assume-se que eles irão compartilhar cooperativamente o mesmo ambiente. Na tarefa específica de carregar um objeto, o agente robótico está assistindo o usuário, logo esta aplicação se enquadra na superposição de Robótica Assistiva e Robótica Socialmente Interativa, sendo então Robótica Socialmente Interativa, mesmo que não haja um grau de interação social tão aprofundado.

Como há interação humano-robô, é importante que esse trabalho leve em conta as medidas para que a interação seja segura fisicamente e psicologicamente. Então a interação deve garantir que não haja colisões com o usuário e que a interação seja socialmente agradável, isto é, respeitando suas zonas sociais (ver Seção 2.4).

1.1.1 Objetivos

O objetivo geral desse trabalho é a cooperação humano-robô possibilitando que comandos sejam dados e tarefas, como auxílio no transporte de objetos, seguimento do usuário ou destinação de tais objetos para lugares predeterminados, enquanto desvia de obstáculos durante realização dessas atividades.

Esse trabalho almeja, especificamente:

1. Desenvolver e aplicar um algoritmo de reconhecimento de ações utilizando dados de esqueleto do *Kinect*;
2. Propor e implementar um classificador baseado em Redes Neurais Artificiais que utilize um banco de dados pequeno para treinamento.
3. Integrar a comunicação feita por ações com a realização de tarefas pela plataforma móvel *Pioneer 3DX*.

Em suma, a comunicação humano-robô se dará mediante ações capturados por um sensor *Kinect* e interpretados por um rede neural especializada. As ações reconhecidas serão enviadas como comandos para um robô *Pioneer 3DX*. O ambiente de navegação será parcialmente conhecido e a interação deverá ser socialmente agradável.

1.2 Trabalhos Correlatos

Esse trabalho é a junção de vários setores. Nele serão vistas técnicas de reconhecimento de ações ou gestos utilizando Redes Neurais Artificiais (RNA) e comparando

com a utilização do *Dinamic Time Warping* (DTW), em tarefas de seguir o usuário e levar objetos de um lugar para o outro, com algum nível de interação social. Não há tantos trabalhos que juntem todos esses setores, então para essa seção, serão apresentados trabalhos correlatos a um ou mais de um desses tópicos.

Nos trabalhos (LISETTI et al., 2004; LEE et al., 2009) pode ser visto exemplos de robôs com interações sociais realizando serviços de entrega em ambientes parcialmente estruturados. Lisetti et al. (2004) se atentaram para a forma como a interação social que deve ser feita. O robô tem uma tela para exibir um rosto virtual e evocar emoções humanas, cativando mais facilmente o usuário. Esse trabalho teve a intenção de entender melhor como que um robô de serviço pode ser mais facilmente aceito. No caso, o robô interage através de linguagem natural com seus usuários e é capaz de realizar tarefas como levar documentos para alguém no escritório, ou levar algum refresco. A interação para a escolha do local para enviar o robô é feita através de uma tela de toque, para selecionar o destino. O trabalho mostra uma boa integração desses dois meios de interação e o robô é dotado de formas diferentes de manifestar emoção com o rosto virtual para aumentar a aceitação com o público.

O segundo trabalho é a documentação da criação do *Snackbot*, um robô feito para entregar lanches entre dois prédios de uma universidade, sendo apto a manter relações sociais com seus usuários de longo prazo, de forma que essa interação pudesse ser objeto de estudo no sentido de entender melhor como a recepção dos usuários foi com o agente robótico. Para poder ser mais facilmente humanizado, o robô conta com uma cabeça com dois olhos e uma boca que acende com LED, no momento em que o robô for emitir sons. Ele é equipado com um mapa da região, para poder deliberar sobre a rota que irá tomar, e a interação com ele se dá por meio de linguagem natural. O foco maior do trabalho foi no *design* do robô e na natureza da interação, para que a interação social fosse mais facilmente evocada ao se relacionar com humanos. O trabalho conclui que para melhorar a interação social, a capacidade de diálogo do robô deveria ser iterativamente atualizada para dar melhor suporte para interações fluidas e naturais.

No tocante a reconhecimento de ações, a utilização de RNA é vista com abordagens diferentes. Em (LI et al., 2017b), é proposto uma abordagem utilizando redes convolucionais para classificação de ações baseado em juntas do esqueleto. Um diferencial desse trabalho é que eles implementam uma transformação linear do esqueleto antes de inseri-lo, mudando a localização de cada junta, dessa forma a rede consegue separar quais juntas são relevantes ou não no processo de treino. Utiliza-se um sistema duas vias para a rede, onde a primeira via tem como entrada as juntas em cada quadro e a segunda via tem como entrada a movimentação das juntas de um quadro para o outro. Depois, a saída das duas vias são concatenadas e passam juntas para as camadas mais profundas da rede.

Em (LI et al., 2017a), similarmente ao trabalho anterior, os autores utilizam também

de redes convolucionais, a principal diferença está na entrada da rede, que nesse trabalho é feito utilizando Mapas de Distância das Juntas. Neste caso, é calculada a distância euclidiana de cada junta para todas as outras e separado isso em vista dos planos xy , xz , yz e de forma 3D com as coordenadas xyz . Cada um desses mapas é utilizado de entrada para uma rede convolucional diferente e seus resultados são fusionados para dar a classificação da ação.

O método DTW utilizado de *baseline* para comparação de desempenho do método proposto com RNA é bem conhecido na literatura e os trabalhos a seguir são exemplos de sua utilização.

Em (SEMPENA; MAULIDEVI; ARYAN, 2011), para tratar as diferenças de forma de corpo de usuários diferentes, os autores utilizam como características das ações, a pose de cada junta em relação ao mundo. As características são extraídas utilizando o *Kinect v1* e testado num total de seis ações diferentes tendo resultados mais consistentes em ações que usam a parte de cima do corpo apenas.

Em (CHERLA et al., 2008), os autores utilizam vetores simples de características como a largura da silhueta do usuário e características espaço-temporais simples como o deslocamento do centro da silhueta e seus desvios padrão em x e y . Esses atributos são invariantes à ponto de vista, almejando garantir um reconhecimento da ação independente da posição do observador. Os autores ressaltam que utilizaram o DTW pelo fato da base de dados disponível não ser tão grande. Seu desempenho é testado no banco de dados IXMAS e para tratar as ações que não tiveram um desempenho tão bom, os autores implementam um sistema de pesos para ponderar as características para a classificação dessas ações.

No âmbito de interação humano-robô com o agente seguindo o usuário, o trabalho (HERRERA et al., 2017) propõe uma lei de controle baseada na estratégia de controle por impedância para lidar com a aceitabilidade social do robô, pois concluem que a proposição deles fará o robô andar similarmente ao humano e isso tornará a interação social mais facilmente aceita. Eles empregam uma estratégia de formação líder-seguidor para essa tarefa com o foco em manter uma distância constante do usuário, que age como líder nessa situação. O sensoriamento é feito por um LIDAR com campo de visão de 180 graus e precisão de 0,5 grau, aumentando o leque de visão para que o robô possa identificar o usuário para segui-lo.

1.3 Contribuições

As contribuições resultantes deste trabalhos de dissertação de mestrado são quatro artigos publicados e outros dois sob revisão.

Dos artigos aprovados, dois deles foram desenvolvidos na área de navegação híbrida,

unindo navegação deliberativa, utilizando o algoritmo A*, com navegação reativa utilizando Desvio Tangencial (OLIVEIRA; CARVALHO; BRANDÃO, 2018; OLIVEIRA; CARVALHO; BRANDÃO, 2019). Essas contribuições ajudaram na compreensão e implementação da navegação do robô.

O terceiro artigo é uma contribuição que facilita a implementação do Desvio Tangencial em plataformas com sensores esparsos, consertando as limitações do método para essa situação (CARVALHO; OLIVEIRA; BRANDÃO, 2018). Entretanto, o tópico nele abordado fugiu ao escopo dessa dissertação.

O quarto artigo compara o método de classificação proposto nesse trabalho com o DTW em uma análise simples em testes *offline* (BASILIO; CARVALHO; BRANDÃO, 2019).

Um dos artigos sob revisão é uma comparação mais profunda e detalhada do método de classificação de ações proposto com o DTW, com testes *offline* e *online* e com um experimento com a plataforma robótica no cenário de jogo da velha.

O outro artigo ainda sob revisão trabalha a aplicação do método de reconhecimento de ações proposto integrado à plataforma robótica *Pioneer 3DX* para a realização de tarefas auxílio ao usuário no transporte de cargas.

1.4 Estrutura do Trabalho

Esse trabalho se divide em capítulos com os seguintes conteúdos:

1. Capítulo 2: Nesse capítulo é feita a fundamentação teórica de conceitos importantes utilizados nesse trabalho, como as técnicas de reconhecimento de ações utilizadas, a navegação do robô com sua lei de controle e técnica de desvio de obstáculos, e o tratamento das zonas sociais utilizadas na navegação do robô, para que seja garantido uma navegação socialmente agradável.
2. Capítulo 3: Nesse capítulo são expostos os materiais e métodos utilizados e também a metodologia por trás da implementação do trabalho. Aqui é trabalhado com detalhes a criação da base de dados e as particularidades da implementação das técnicas de reconhecimento de ações.
3. Capítulo 4: Este capítulo expõe e discute os experimentos realizados. Os primeiros são testes *offline* comparando o desempenho do método proposto e do DTW para situações diferentes, a fim de validar a eficiência do classificador com redes neurais e, por fim, avaliar o desempenho do classificador proposto com uma base de dados para treinamento reduzido. Testes *online* também são realizados, um analisando apenas o desempenho do classificador proposto; outro validando a interação humano-robô

através de ações para ordenar o posicionamento do agente robótico em um jogo da velha; e, por fim, um experimento com a comunicação sendo utilizada para o agente realizar ações auxiliando o usuário a transportar uma carga.

4. Capítulo 5: apresenta as conclusões do trabalho, evidenciando suas principais contribuições e elucidando propostas para sua continuação.

2 Fundamentação Teórica

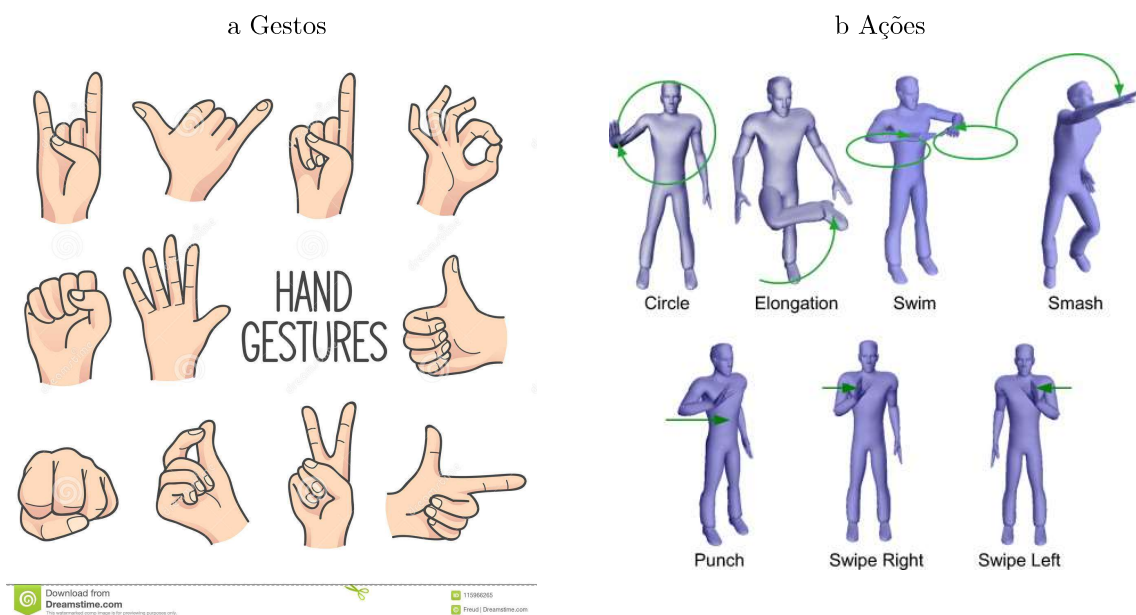
2.1 O Reconhecimento de Ações

Na literatura, nomes diferentes são usados para descrever movimentos com o corpo. Ações, gestos simples e gestos de corpo completo podem ser observados na literatura ao lidar com o que no dia a dia, é comumente referido como gestos.

No trabalho (CABIBIHAN; SO; PRAMANIK, 2012), é usado gestos de corpo completo para descrever gestos que são feitos com um ou dois braços, como apertar a mão, dirigir, andar (apenas vendo da cintura para cima) e balançar os braços. Nos trabalhos (ABDELNASSER; YOUSSEF; HARRAS, 2015; MURAKAMI; TAGUCHI, 1991) o termo gestos é utilizado para se referir a gestos feitos apenas com as mãos. Já em (LI; ZHANG; LIU, 2010) o autor se refere aos gestos feitos com o corpo inteiro como ações, enquanto em (CANAL; ESCALERA; ANGULO, 2016; IBANEZ et al., 2014) o termo gestos é utilizado para se referir a ações que são feitas com o corpo todo.

Tendo isso em vista, a terminologia não é muito bem definida na literatura, então para esse trabalho será utilizado a terminologia “ação” para os sinais utilizados. “Gestos” serão vistos apenas como feitos apenas pelas mãos. A Figura 5 ilustra essa diferença.

Figura 5 – Ilustração da terminologia utilizada



Fontes: Dreamstime (2019), Ibanez et al. (2014)

2.1.1 Reconhecimento de Ações

Atualmente, sensores RGB-D (*Red, Green, Blue – Depth*, ou seja, Vermelho, Verde, Azul – Profundidade) de baixo custo, tais como *Intel RealSense* e o *Microsoft Kinect* fornecem de forma simplificada o acesso a informação de cor e profundidade do ambiente. Aliando isso ao fato de que reconhecimento de ações tem uma grande gama de aplicações, os esforços acadêmicos em pesquisas utilizando informação de profundidade (ou tridimensionais) de esqueletos para classificação de ações ou gestos têm aumentado nas últimas décadas (PATRONA et al., 2018; YANG et al., 2017). Dentre algumas aplicações, pode-se citar: utilização em ambiente público para detectar atividades suspeitas, monitoramento médico, análise de performance atlética e também interação humano-robô avançada em aplicações de robótica socialmente assistiva (SEMPENA; MAULIDEVI; ARYAN, 2011; AGGARWAL; XIA, 2014; AGGARWAL; RYOO, 2011; MOESLUND; HILTON; KRÜGER, 2006). Reconhecimento de ação pode ser dividido em: extração das características principais (do inglês, *features*) e classificação dessas *features* (AGAHIAN; NEGIN; KÖSE, 2019).

Sensores de rastreamento em geral são Baseados em Visão (do inglês, *Vision Based*), ou de Captura de Movimento (do inglês, *Motion Caption, MoCap*). O primeiro utiliza imagens, que podem ter informação de cor (RGB) ou de profundidade, para processar o ambiente e suas particularidades para realização da classificação. O segundo baseia-se nas *features*, onde se extrai parâmetros específicos tais como informação 3D de esqueleto como posição, velocidade e/ou orientação das juntas, que podem ser adquiridas através de processamento das imagens ou tendo marcadores no corpo para extração direta dessas informações (MITRA; ACHARYA, 2007).

Após o rastreamento do esqueleto, métodos devem ser aplicados para a classificação de ações ou gestos. Dentre algumas técnicas para realizar esta função, tem-se:

1. ***Dynamic Time Warping (DTW)*** que compara séries temporais para encontrar seu grau de similaridade, isso é feito através da comparação de pontos em espaço de tempo próximo das duas séries. O cálculo pode usar várias métricas de distância, como Manhattan ou euclidiana. Isso confere uma robustez às distorções temporais entre as séries (CELEBI et al., 2013; RAHEJA et al., 2015; HANG et al., 2017; BARNACHON et al., 2014; VERCELLI et al., 2018).
2. ***Hidden Markov Model (HMM)*** que utiliza a quantificação da configuração do sistema através de um número finito de estados discretos, onde os valores armazenados representam uma aproximação da dinâmica do sistema. Esses estados podem ser um vetor de *features* contendo a informação espacial de cada característica dos objetos obtida através de uma autocalibração (KUMAR et al., 2017; ZHANG et al., 2016);
3. **Redes Neurais Artificiais (RNA)** que é um tipo de técnica de aprendizado por supervisão, a qual utiliza um conjunto de treinamento para poder refinar seus

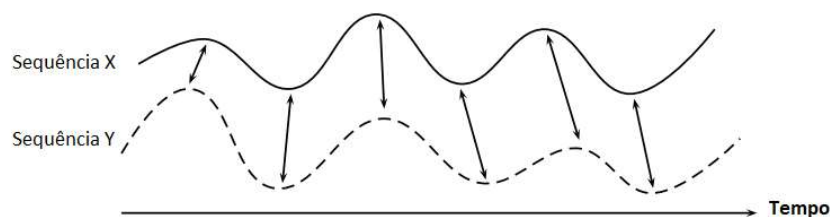
parâmetros correlacionando suas entradas com as etiquetas das suas saídas. Existem diversas estruturas para essas redes, tais como Redes Neurais Recorrentes, mais usadas quando a entrada é uma série temporal (NG; RANGANATH, 2002; DU; WANG; WANG, 2015; VEERIAH; ZHUANG; QI, 2015); Redes Neurais Profundas, onde várias camadas de neurônios são empregadas e é necessário um conjunto de treinamento extenso, mas isso confere à rede uma capacidade de extrair as *features* por conta própria, sem supervisão (WANG et al., 2015; ORDÓÑEZ; ROGGEN, 2016); as Redes Neurais Convolucionais onde as camadas subsequentes não são totalmente conectadas, permitindo que a rede seja profunda, mas com menos parâmetros do que se fosse composta apenas de camadas densas (WANG et al., 2018; YAN; XIONG; LIN, 2018).

2.2 *Dynamic Time Warping - DTW*

A ideia principal do DTW é de comparar duas sequências dependentes do tempo. Essas sequências podem ser discretas (séries temporais) mas em geral elas são sequências amostradas com um tempo de separação constante. As sequências analisadas podem, inclusive, ter tamanhos diferentes. Na situação desse trabalho, uma ação pode ser representada como uma série temporal ao colocar as informações de cada junta em sequência com o crescimento do tempo, dessa forma cada elemento subsequente será uma informação a frente no tempo da anterior.

A Figura 6 mostra o alinhamento temporal de duas sequências distintas. As setas indicam o elemento próximo mais similar das duas sequências.

Figura 6 – Alinhamento temporal de duas séries.

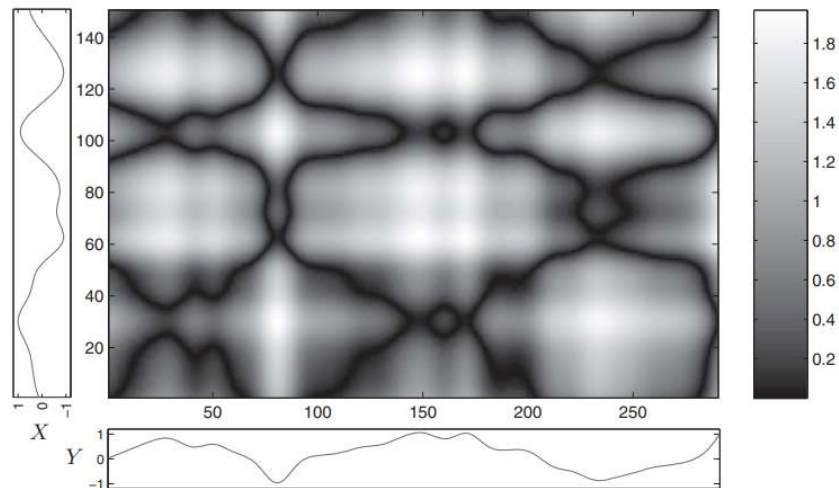


Fontes: Müller (2007)

Como pode ser visto em (MÜLLER, 2007), o DTW clássico se dá ao procurar nas proximidades dos elementos de uma sequência, o ponto com a menor diferença nos elementos da outra sequência. Isso pode ser mais facilmente visualizado com o esquema da Figura 7, também chamado de matriz de custo, onde o custo é na verdade a distância calculada pela distância euclidiana ou outras métricas.

Na Figura 7, é possível ver o custo de cada elemento de cada sequência em relação a todos os elementos da outra sequência, como por exemplo uma linha nessa matriz indica o custo do elemento da sequência X referente àquela linha com todos os elementos da sequência Y .

Figura 7 – Matriz de custo de duas sequências, X na vertical e Y na horizontal. As regiões mais escuras tem custo mais baixo e as mais claras custo mais alto.



Fontes: Müller (2007)

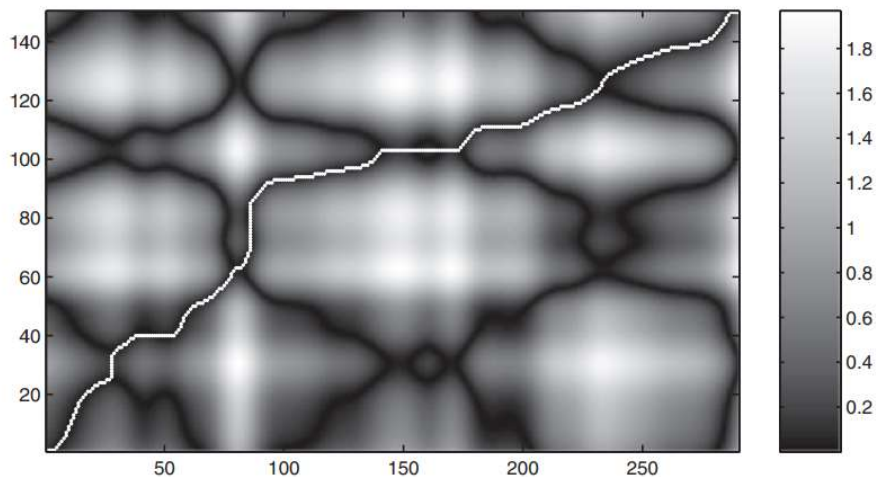
Ao analisar o custo de cada combinação de elemento, pode ser encontrado um caminho com o menor custo entre as sequências. Esse caminho consiste em uma sequência de pares, onde há um elemento de cada sequência. Esses pares representam o mapeamento que relaciona um ponto de uma sequência com o mais similar a ele nas suas proximidades com a outra sequência. Esse caminho então é como se fosse um guia para relacionar os pontos de uma sequência com os mais similares da outra. Esse caminho deve respeitar as seguintes condições:

- Fronteira - o primeiro item do caminho deve ser feito pelo par dos elementos iniciais de cada sequência e o último item do caminho pelos últimos itens de cada sequência.
- Monotonicidade - Dois pares seguidos no caminho (x_n, y_n) e (x_{n+1}, y_{n+1}) devem respeitar a seguinte relação: $x_n \leq x_{n+1}$ e $y_n \leq y_{n+1}$, garantindo que o caminho nunca compare elementos voltando no tempo e sim sempre estando no mesmo momento, ou avançando.
- Tamanho do Passo - Como dito no texto, a procura pelo item de menor custo se dá nas proximidades do elemento de cada sequência. O tamanho dessa proximidade não deve passar de um valor fixo, tanto para quanto se avança em uma sequência, quanto

na outra. Não se deve dar passos irregulares nessa procura de um correspondente de menor custo para elementos diferentes.

O caminho correto para a Figura 7 se encontra na Figura 8, onde se traça o caminho pelo “vale” de menores valores, começando no canto inferior esquerdo da matriz e terminando no canto superior direito, respeitando a condição de fronteira.

Figura 8 – Caminho de menor custo para as Sequências X e Y



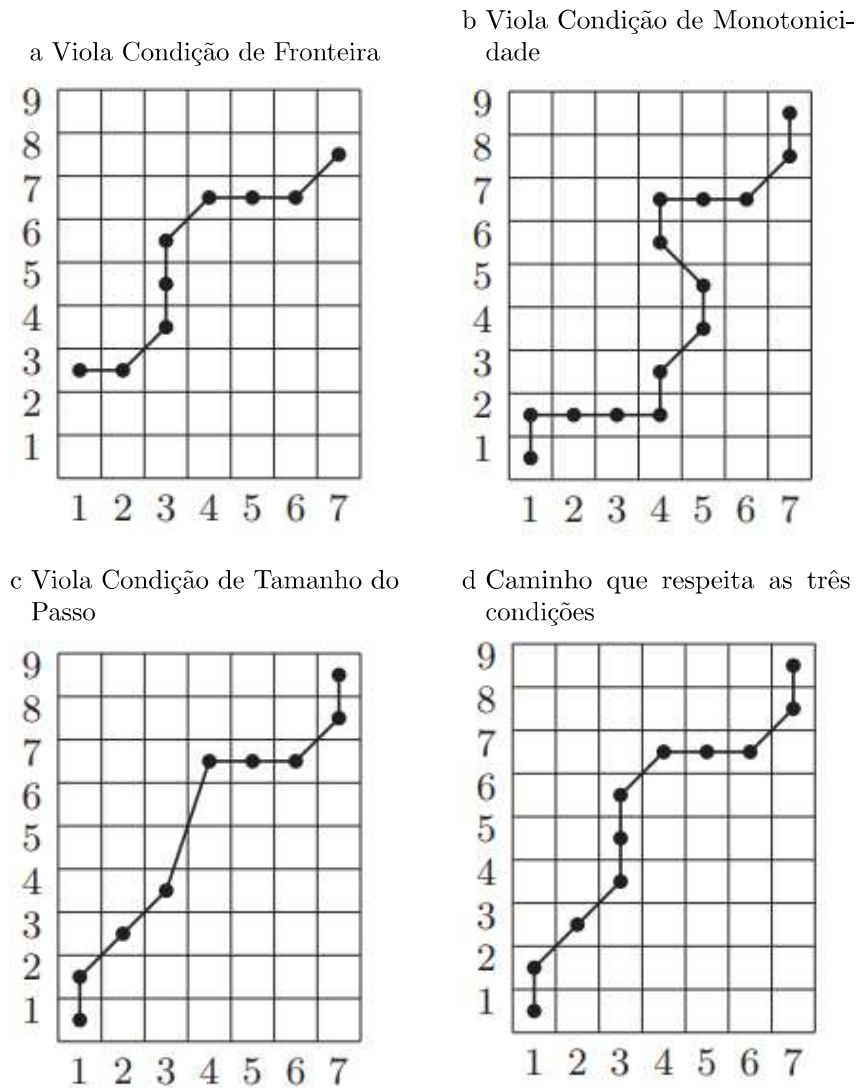
Fontes: Müller (2007)

Para melhor compreensão das regras, a Figura 9 ilustra quatro situações diferentes, mostrando caminhos que violam cada uma das condições separadamente e, por fim, um caminho que respeita todas.

No item a, o primeiro par do caminho corresponde ao item 3 da sequência na vertical e ao item 1 da sequência na horizontal, não começando no início das duas, violando a condição de fronteira. No item b, é possível ver que depois do par (5,5), que liga o quinto item de cada sequência, o par seguinte é (4,6) retrocedendo em 1 na sequência horizontal, violando a condição de monotonicidade. No item c, é possível ver que o maior passo em relação a sequência horizontal e vertical é sempre 1, como se pode ver nos primeiros pares do caminho: (1,1), (1,2), (2,3) e (3,4), onde o máximo que se avança em cada sequência é 1. Mas o par seguinte do caminho, (4,7) avança apenas 1 na sequência horizontal, mas avança 3 na vertical, violando a condição de tamanho do passo. No item d é possível ver um caminho que respeita as 3 condições, pois começa com os itens iniciais de cada sequência e termina com os itens finais, os pares subsequentes do caminho não estão retrocedendo em seus índices e os passos estão sempre dentro do mesmo limite de 1.

Para fins de exemplo, o caminho ótimo para a Figura 9d se dá pelos pares: (1,1),(1,2),(2,3),(3,4),(3,5),(3,6),(4,7),(5,7),(6,7),(7,8),(7,9). Então o custo final desse cami-

Figura 9 – Exemplos de caminho para uma sequência com 9 e outra com 7 itens.



Fontes: [AV Magazine \(2019\)](#), [Medium \(2019\)](#), [Bentley \(2019\)](#)

nho se daria, utilizando distância euclidiana por:

$$\begin{aligned}
 C(H, V) = & \sqrt{(H_1^2 - V_1^2)} + \sqrt{(H_1^2 - V_2^2)} + \sqrt{(H_2^2 - V_3^2)} \\
 & + \sqrt{(H_3^2 - V_4^2)} + \sqrt{(H_3^2 - V_5^2)} + \sqrt{(H_3^2 - V_6^2)} + \sqrt{(H_4^2 - V_7^2)}, \\
 & + \sqrt{(H_5^2 - V_7^2)} + \sqrt{(H_6^2 - V_7^2)} + \sqrt{(H_7^2 - V_8^2)} + \sqrt{(H_7^2 - V_9^2)}
 \end{aligned}
 \tag{2.1}$$

onde H é a sequência horizontal e V a sequência vertical, então $C(H, V)$ é o menor custo, seguido pelo caminho ótimo, entre as duas séries e H_n e V_n são os valores dos n -ésimos elementos dessas sequências usadas para exemplificar o método.

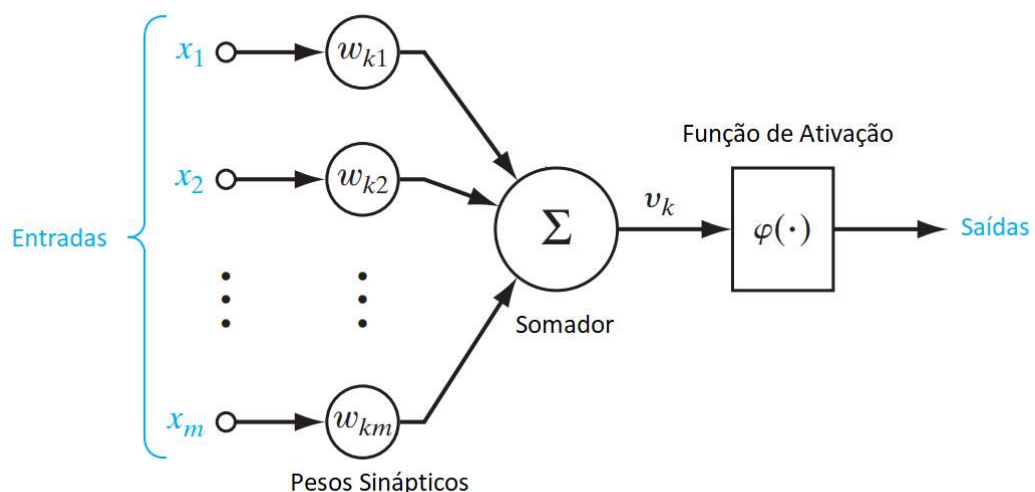
2.3 Redes Neurais Artificiais - RNA

Haykin (1994) diz que uma Rede Neural Artificial é um processador massivamente distribuído e paralelo, composto de unidades simples de processamento que tem a capacidade natural de armazenar conhecimento de experiências e de fazê-lo disponível para uso. Essa estrutura é inspirada no cérebro e se assemelha em dois aspectos fortes. O primeiro é que o conhecimento é adquirido pela rede por meio do ambiente em que está inserida e por meio de um método de aprendizagem. O segundo é que a força da conexão interneural, chamada de pesos sinápticos (ou apenas de pesos), são utilizados para armazenar o conhecimento adquirido.

O uso de redes neurais carrega consigo algumas vantagens, tais como a capacidade de atacar problemas não lineares e lineares, dado a capacidade do neurônio funcionar tanto de uma forma quanto da outra e a capacidade se adaptar, mudando os pesos de suas conexões frente à variações no seu ambiente.

A Figura 10 mostra um esquema simples de um neurônio artificial. Ele é composto de entradas, que são multiplicadas pelos seus respectivos pesos, então esses valores são somados e esse valor é aplicado à função de ativação, isso configura à saída do neurônio, que em uma rede neural pode estar interconectada a um ou mais neurônios e atrelada a pesos diferentes para cada um. A Figura 11 contém exemplos de funções de ativação.

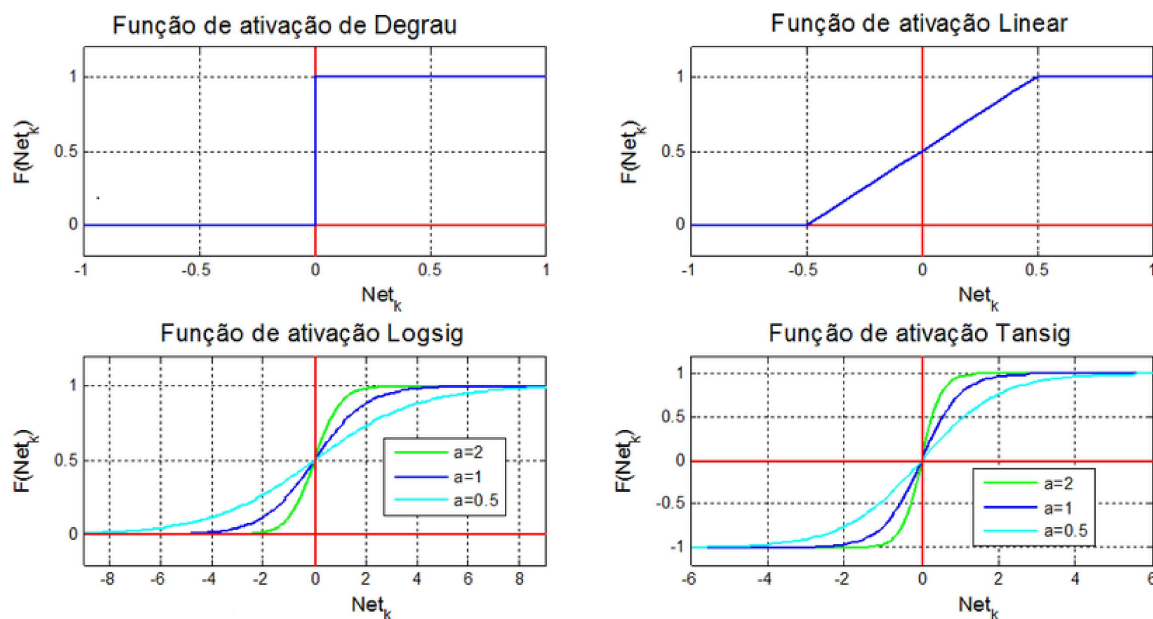
Figura 10 – Neurônio Artificial.



Fonte: Haykin (1994)

Redes neurais aprendem de forma supervisionada, através de exemplos. Para poder armazenar o conhecimento obtido dessa forma nos pesos de suas conexões, algoritmos de treinamento devem ser empregados.

Figura 11 – Funções de Ativação.



Fonte: Barrientos e Cataldo (2015)

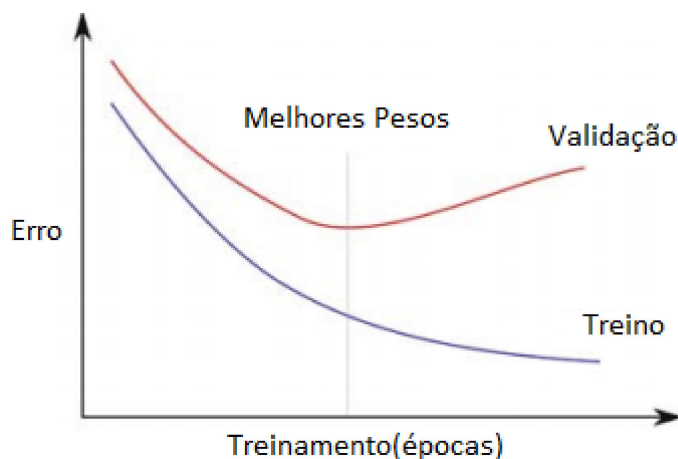
Um método bem conhecido é chamado de *Backpropagation*, que funciona com os seguintes passos extraídos e simplificados de (HAYKIN, 1994):

- **Inicialização:** O início do treinamento se dá com valores aleatórios para os pesos das conexões entre os neurônios.
- **Treinamento:** Apresenta-se para a rede uma época de amostras para treino. Para cada exemplo nesse grupo deve ser aplicados os próximos dois passos.
- **Cálculo da saída:** Aplica-se a amostra em questão à entrada da rede e compara a saída da rede com os valores obtidos na etapa de treinamento.
- **Backpropagation:** Nessa etapa o erro obtida pela diferença do resultado esperado da amostra com a saída da rede é utilizado para modificar o valor dos pesos individuais das conexões, baseado na influência que ela teve para o resultado final.
- **Iteração:** Realiza-se os últimos dois procedimentos para todas as amostras até que o critério de parada seja alcançado.

As amostras são comumente divididas em uma porcentagem para treino, onde são usadas para modificar os valores dos pesos baseados no desempenho, e uma porcentagem para validação, onde se avalia o que a rede aprendeu usando os dados que a rede não usou para aprender. Um critério de parada comumente utilizado é quando o erro quadrático absoluto da parcela de validação é mínimo preestabelecido. (BURDEN; WINKLER, 2008).

A Figura 12 ilustra o momento em que o erro de validação passa o ponto de mínimo e volta a subir, indicando o instante de parada da etapa de treinamento.

Figura 12 – Critério de parada no treinamento em função da validação.



Fonte: Burden e Winkler (2008)

Para decidir como modificar os pesos, uma abordagem conhecida é a de gradiente descendente, que faz as atualizações do peso na direção contrária do gradiente da função objetivo. Uma maneira de visualizar isso graficamente, é como se a cada iteração a rede estivesse navegando em um hiperplano procurando o menor valor de erro possível utilizando o gradiente desse hiperplano como guia. Sabendo para qual sentido o erro cresce mais (o que o gradiente aponta), a navegação deve ser dada no sentido oposto.

Outra maneira conhecida para guiar a atualização dos pesos é o método de Gauss-Newton (GAVIN, 2019), que age na direção da minimização da soma do quadrado dos erros da função objetiva. Ele presume que a função é aproximadamente quadrática quando os pesos estão próximos do estado ótimo (BJÖRCK, 1996).

Continuar o treinamento pode gerar o fenômeno de *overfitting*, onde a rede tem uma alta taxa de acerto com os exemplos utilizados para treino, porém perde capacidade de generalização para situações fora da base de dados utilizada para treinamento. Uma técnica que é mais robusta a esse fenômeno de perda de generalização é a Regularização Bayesiana (BURDEN; WINKLER, 2008), que aplica o teorema de Bayes no processo de aprendizado, inclinando o treinamento para um problema de estatística bem definido. Uma das vantagens é que durante a regularização, pesos mais relevantes são identificados e alguns são eliminados, dificultando a ocorrência de *overfitting*.

Para melhorar o desempenho do treinamento, esse trabalho utiliza o otimizador de Levenberg-Marquardt, que é um algoritmo que de forma adaptativa varia as atualizações dos pesos entre o que o método de gradiente descendente aponta e entre o que o método de Gauss-Newton aponta (RANGANATHAN, 2004).

2.4 Zonas Sociais

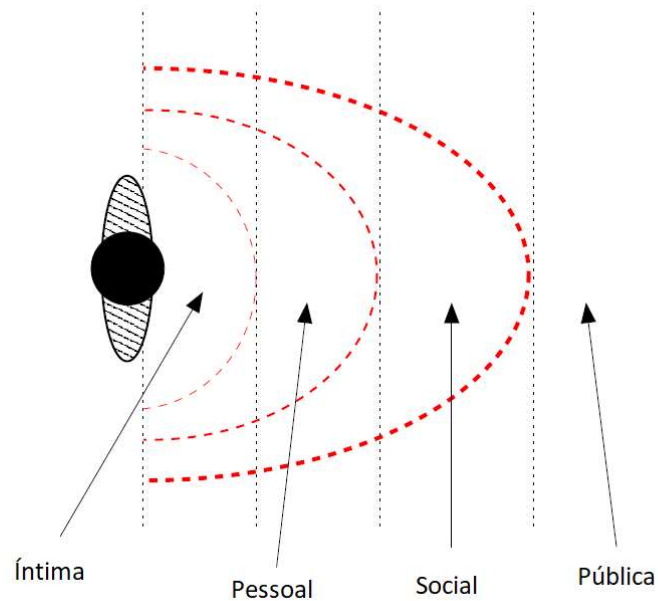
A capacidade de navegação autônoma em ambientes não estruturados e dinâmicos é crucial para robôs de serviço quando aplicadas em situações de interações sociais tais como aeroportos (TRIEBEL et al., 2016), museus (THRUN et al., 1999), situações em que o robô deve acompanhar lado a lado o usuário (FERRER et al., 2017). Em aplicações como as supracitadas, é fundamental que o robô evite colisões não só com obstáculos regulares mas com os humanos envolvidos. Nesse contexto, segurança humana em relação à navegação robótica pode ser dividida em duas categorias (NONAKA et al., 2004): segurança física e psicológica. A primeira é bem intuitiva, o agente não deve colidir com os humanos, ficando a uma distância segura mínima. Por outro lado, a segunda categoria implica que o robô deve navegar sem causar estresse ou desconforto para os humanos.

As relações de proximidade nas interações humanas vêm sendo estudadas pela *Proximics*, termo utilizado pela primeira vez por Hall (1969). Essa área tem como objeto de estudo as distâncias sociais utilizadas nas relações humanas em várias situações sociais e interpessoais. Essas distâncias podem ser modificadas dado o contexto cultural ou o ambiente que a interação está inserida. A maior parte da comunicação humana é feita através de forma não verbal (HOGAN, 2003) e ela tem grande influência na modificação de como cada humano irá se sentir confortável com certa proximidade ou não na interação. A identificação automática dessa comunicação ainda é objeto de grande desafio na área, Vinciarelli et al. (2008) mostram um estudo mais detalhado de como isso tem sido feito.

A gestão do espaço feita por uma pessoa em uma interação social é diferente de quando se está em grupo, como visto em (KRUEGER, 2011). Esse trabalho analisa as zonas sociais individuais da interação de um humano com um robô. Para isso é considerado a divisão de quatro regiões de interação (RIOS-MARTINEZ; SPALANZANI; LAUGIER, 2015), como visto na Figura 13. Essas divisões podem ser definidas, de acordo com (PACCHIEROTTI; CHRISTENSEN; JENSFELT, 2005), como:

1. **Zona Íntima:** Onde a interação é ou diretamente física tal como um abraço ou então uma interação privada, tal como um sussurro. Interações assim acontecerão com pessoas muito íntimas ou pessoas cumprindo algum tipo de formalidade necessária. Comumente definida como distâncias de até 45 centímetros.
2. **Zona Pessoal:** Utilizada mais para interação com amigos e família, com pessoas que o usuário é mais próximo socialmente, ou então em situações de alta organização tais como uma fila, ou alguma dinâmica de grupo. Interações nessa zona são comumente vistas em situações em que há um grau de intimidade mais intenso, fazendo a interação nessa zona não gerar desconfortos. Normalmente, esta zona está compreendida nas distâncias entre 45 centímetros até 1,2 metros.

Figura 13 – Ilustração das Zonas Sociais.



Fonte: Pacchierotti, Christensen e Jensfelt (2005)

3. **Zona Social:** Utilizada para situações mais formais ou de negócios. Interações com conhecidos e pessoas com menos intimidade uma com a outra. Também utilizada como a distância de separação em ambientes públicos tais como praias ou lojas. Comumente esta zona compreende distâncias entre 1,2 metros e 3,5 metros.
4. **Zona Pública:** Utilizada quando não se há interação, de forma que as pessoas envolvidas estão no campo de visão uma da outra, mas a princípio não possuem interação social, ou quando existe, ela é de mão única. Neste último caso, quando apenas um dos lados da interação está se manifestando durante a interação, pode ser exemplificado como a relação entre um apresentador, um palestrante e sua plateia. Comumente esta zona está definida para distâncias maiores do que 3,5 metros.

Esta dissertação utiliza a noção desses espaços para a interação humano-robô, de forma que ao respeitá-los, a interação é socialmente agradável, não gerando estresse ou desconforto ao usuário.

A interação social esperada nesse trabalho é que o usuário seja capaz de se comunicar com o agente robótico e com isso enviar comandos para a execução de tarefas. É esperado que o usuário naturalmente dê comandos ao robô a partir da sua zona social. Operacionalmente, a distância limita para a interação social é de 3,5 metros, em função do sensor de profundidade utilizado, no caso, o sensor *Kinect* da Microsoft.

Dos comandos propostos para o robô, dois deles merecem uma atenção especial: seguir e aproximar.

Ao seguir o usuário, o robô para evitar qualquer desconforto social. Ademais, ele deve seguir o usuário a uma distância que um conhecido o faria, ou seja, ele não deve sair da zona social do usuário, ficando entre 1,2 metros e 3,5 metros. Para esse trabalho, a distância para seguir é de 2,25 metros. Dessa forma, o robô não estaria nem longe demais para a comunicação ser perdida e nem perto demais para o usuário se sentir desconfortável.

A aproximação do robô só ocorrerá sob comando direto do usuário, ou seja, não ocorrerá de forma inesperada. Para tal, foi definido que, ao ser requisitado para se aproximar, o robô entra na zona pessoal do usuário, estando sempre entre 45 centímetros e 1,2 metros. Em particular, nesse trabalho foi definido que essa distância é de 1 metro. Assim, o robô só entrará numa zona mais próxima do usuário com a permissão dele.

Tomada tais considerações, ao respeitar as zonas sociais do usuário, a navegação se torna socialmente agradável.

2.5 A Navegação do Robô

2.5.1 Modelagem e Controle

O modelo cinemático do robô é dado por

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} \cos(\psi) & -a \sin(\psi) \\ \sin(\psi) & a \cos(\psi) \end{bmatrix} \begin{bmatrix} u \\ \omega \end{bmatrix} \quad (2.2)$$

que pode ser escrita como

$$\dot{\mathbf{x}} = \mathbf{K}\mathbf{v}, \quad (2.3)$$

onde u e ω são a velocidade linear e angular do robô, respectivamente; a é a distância do centroide do robô para o ponto de controle; x e y são as respectivas coordenadas cartesianas da posição do robô; ψ é a orientação dele em relação ao sistema de coordenadas global $\langle 0 \rangle$; e $\dot{\psi} = \omega$ representa a cinemática da orientação.

Essas grandezas estão representadas na Figura 14.

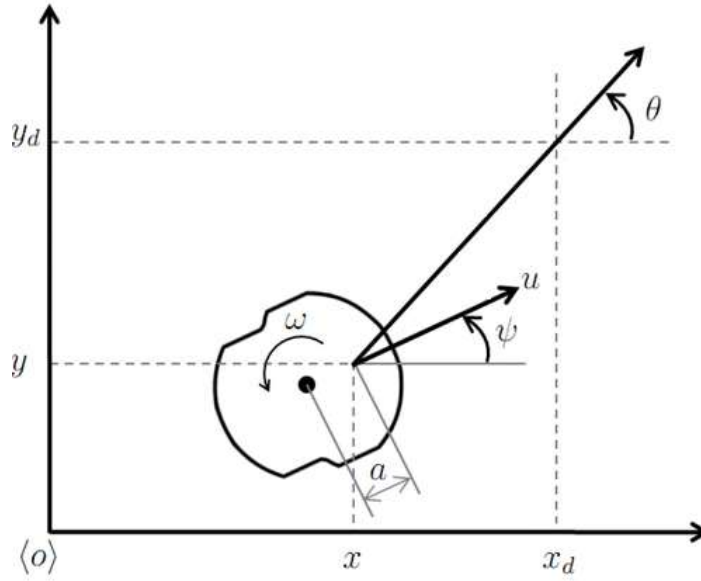
Para o controle de posição do robô, projetou-se um controlador baseado na Teoria de Lyapunov, similarmente ao feito em (BRANDÃO; SARCINELLI-FILHO; CARELLI, 2013), de tal forma que os valores de u e ω são adequadamente definidos com o intuito de fazer $\tilde{x} = x_d - x \rightarrow 0$ e $\tilde{y} = y_d - y \rightarrow 0$ com o passar do tempo.

Em outras palavras, o objetivo do controlador é fazer o vetor de estado $\tilde{\mathbf{x}} = [\tilde{x} \ \tilde{y}]^T$ ir assintoticamente para zero, dado um valor inicial qualquer. Para tal, adotou-se a candidata de Lyapunov

$$V = \frac{1}{2} \tilde{\mathbf{x}}^T \tilde{\mathbf{x}} \quad (2.4)$$

e a lei de controle deve ser feita de tal forma que a primeira derivada temporal de V , $\dot{V} = \tilde{\mathbf{x}}^T \dot{\tilde{\mathbf{x}}}$, seja negativa para todos os valores de $\tilde{\mathbf{x}}$. Para garantir a estabilidade do sistema

Figura 14 – Esquema do modelo cinemático do robô



Fonte: Brandão, Sarcinelli-Filho e Carelli (2013).

e levando em conta o modelo cinemático (2.3), adotou-se a lei de controle

$$\mathbf{v} = \mathbf{K}^{-1} [\dot{\mathbf{x}}_d - \mathbf{k}_1 \tanh(\mathbf{k}_2 \tilde{\mathbf{x}})], \quad (2.5)$$

onde \mathbf{k}_1 e \mathbf{k}_2 são matrizes diagonais definidas positivas. Logo,

$$\dot{V} = -\tilde{\mathbf{x}}^T \mathbf{k}_1 \tanh(\mathbf{k}_2 \tilde{\mathbf{x}}) < 0, \quad (2.6)$$

que por sua vez garante $\tilde{\mathbf{x}} \rightarrow \mathbf{0}$.

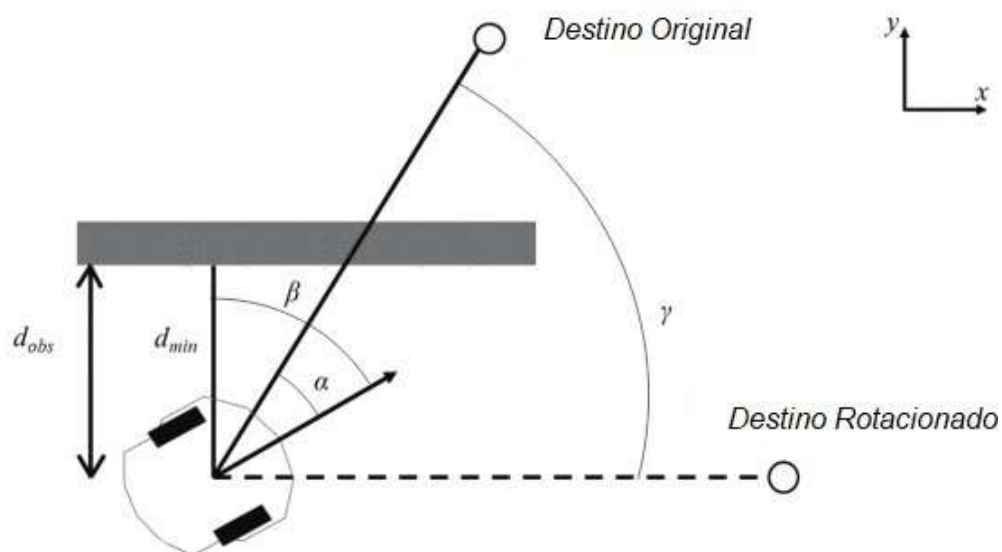
2.5.2 Desvio de Obstáculos

O desvio de obstáculos usado neste trabalho é o Desvio Tangencial, visto em (BRANDÃO; SARCINELLI-FILHO; CARELLI, 2013). Ele é um algoritmo de baixo custo computacional devido sua simplicidade matemática. Ele se pauta em esquivar dos obstáculos rotacionando o seu destino de tal forma que sua navegação passe a ser paralela ao obstáculo. Os parâmetros envolvidos são: a distância mínima para que o robô comece a desviar, d_{obs} , definida como 1 metro nesse trabalho; a menor distância lida pelos sensores do robô, d_{min} ; e o ângulo referente à orientação do robô dessa menor leitura, β . Esses dados são ilustrados na Figura 15.

Com base nessas informações, o ângulo de rotação do destino é dado por

$$\gamma = \begin{cases} -\frac{\pi}{2} + \beta - \alpha, & \text{se } \beta > 0 \\ \frac{\pi}{2} + \beta - \alpha & \text{se } \beta < 0. \end{cases} \quad (2.7)$$

Figura 15 – Esquema do Desvio Tangencial.



Fonte: Oliveira, Carvalho e Brandão (2019)

Importante explicitar que dado ângulo de abertura do *Kinect*, os sensores utilizados para a tarefa de desvio de obstáculos foram os sensores ultrassônicos de distância do robô *Pioneer 3DX*.

3 Metodologia

3.1 Materiais e Métodos

3.1.1 O Sensor *Kinect*

O sensor *Kinect* foi desenvolvido pela *Microsoft* com o intuito de permitir que os usuários interagissem com seus corpos de maneira natural para fins de entretenimento. O sensor permite que o computador identifique a profundidade dos usuários e do ambiente. Ademais, ele sabe quando as pessoas estão falando ou andando e pode rastrear o posicionamento de cada pessoa em seu campo de visão. O impacto do *Kinect* foi muito mais extenso que apenas sua aplicação em jogos. Devido ao seu baixo custo e ampla disponibilidade, tem sido usado em pesquisas em ciência da computação, eletrônica e engenharia em atividades desde ajudar crianças com autismo até auxiliar médicos em operações (ZHANG, 2012).

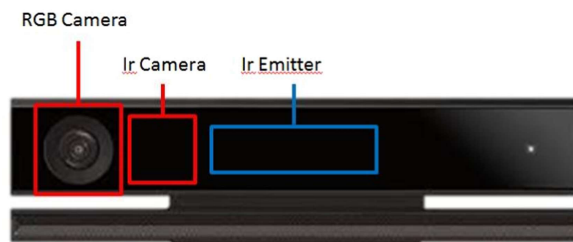
O *Kinect* é composto por vários sensores, como um projetor infravermelho, uma câmera colorida, uma câmera infravermelha e um arranjo de quatro microfones que permitem captura corporal 3D completa, reconhecimento facial e reconhecimento de voz.

Outra das inovações do *Kinect* é sua capacidade de rastrear esqueletos, o que tem um grande leque de aplicações, pois ele retorna a posição em um plano cartesiano de cada uma das vinte e cinco juntas do nosso corpo. Dessa forma, essa informação pode ser utilizada para reconhecimento de inúmeros gestos (ZHANG, 2012).

Dentre as diversas versões disponíveis no mercado, o *Kinect* utilizado neste trabalho é o v2.0. Ele tem uma abertura horizontal de 70 graus e vertical de 60 graus, com uma resolução de 1920x1080 em sua câmera RGB e de 512x424 em sua câmera de profundidade, com capacidade de detecção de objetos entre 50cm e 4.5m de distância. Uma imagem do *Kinect* se encontra na Figura 16, indicando seus sensores principais.

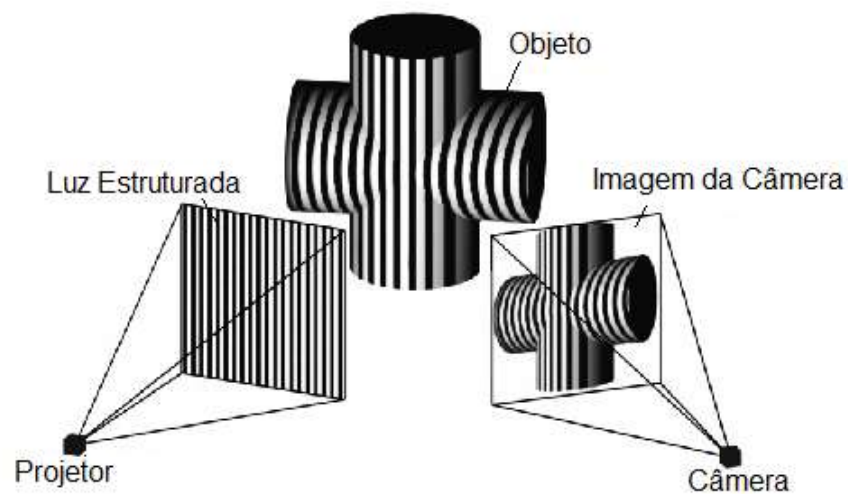
O *Kinect* utiliza a tecnologia *Light Coding* da empresa *PrimeSense*. Seu funcionamento é baseado em luz estruturada, isto é, ele emite uma luz com uma estrutura conhecida e bem definida, como linhas horizontais ou verticais, e baseado no padrão de reflexão dessa luz, é possível obter informações do ambiente com uma câmera. Na Figura 17 é possível ver um projetor de luz estruturada, um objeto e uma câmera enxergando o padrão da luz refletida pelo objeto.

Por fim, é importante ressaltar que nesse projeto será utilizado o *MatLab*, que tem uma interação direta com o *Kinect* e uma grande gama de ferramentas já implementadas, sendo de grande relevância para o desenvolvimento deste trabalho.

Figura 16 – *Kinect v2.0*.

Fonte: Caruso, Russo e Savino (2017)

Figura 17 – Esquema de detecção de luz estruturada.



Fonte: 3D Natives (2020)

3.1.2 *Pioneer 3-DX*

O *Pioneer 3-DX* é um robô terrestre de tração diferencial utilizado em pesquisas na academia devido a sua robustez e confiabilidade. Ele tem ferramentas computacionais para auxiliar o usuário em diversas aplicações, através da biblioteca *ARIA*, que permite o usuário capturar os dados sensoriais processados pelo *firmware* e também enviar sinais de controle para os atuadores utilizando linguagem de alto nível em *C++* (Omrom Adept, a).

O *Pioneer 3-DX* (ver Figura 18) tem um motor com um encoder de 500 pulso por volta, rodas de 19 centímetros de diâmetro, 8 sensores de ultrassom posicionados espaçadamente a sua frente. Sua velocidade linear máxima é de 0.75 m/s e angular de 100 °/s, e é capaz de carregar até 17 kg. Ele é adequado para pesquisas em diversas áreas como mapeamento, teleoperação, localização, reconhecimento, manipulação, cooperação multi robôs e robótica em geral (Omrom Adept, a; Omrom Adept, b).

Neste trabalho é utilizado a biblioteca *ARIA* juntamente do *Matlab* para poder enviar sinais de controle para o robô e receber o *feedback* sensorial dele.

Figura 18 – *Pioneer 3DX*.

Fonte: ROS (2019)

3.2 Base de Dados

O foco deste trabalho é a Interação Humano-Robô para a realização de tarefas de auxílio ao usuário tais como segui-lo e ir para lugares predeterminados. Tendo em vista que as bases de dados encontrados na internet tais como o utilizado por LI; ZHANG; LIAO, têm classes de ações incompatíveis para as tarefas pretendidas para este trabalho, optou-se pela criação de uma base de dados própria. No caso, nove classes de ações padrão foram escolhidas para integrar a base de dados, as quais podem ser vistas na Figura 19. Todas as ações começam da posição de descanso em pé e são feitas da seguinte forma: um gesto de “adeus” com o braço direito (A), elevar o braço direito e fazer um círculo em frente ao rosto (B), elevar o braço direito até que ele faça um ângulo de aproximadamente 45 graus com o torso (C), um gesto de “adeus” com o braço esquerdo (D), cruzar os dois braços em frente ao torso formando um X (E), elevar o braço esquerdo até que ele faça um ângulo de aproximadamente 45 graus com o torso (F), elevar o braço direito em frente de si fazendo um sinal de “pare” (G), colocar as duas mãos juntas na frente da cintura (H), elevar o braço direito com a palma direita para cima fazendo um gesto de “venha” (I).

Duas classes de ações neutras foram adicionadas, o que significa que quando elas forem identificadas o robô não receberá nenhum comando advindo dessas ações. A primeira classe contém amostras do usuário parado e a segunda, do usuário andando. Dessa forma, o agente será capaz de distinguir essas ações neutras dos movimentos naturais do usuário das ações padrão, não executando ação alguma quando uma delas for identificada.

Cada uma das ações padrão escolhidas para integrarem a base de dados tem uma motivação adicional. As classes A, D, G e I, são bastante intuitivas, dois sinais de “adeus”,

um sinal bem claro de “venha” e um de “pare”. As classe C e F podem não parecer tão intuitivas à primeira vista, mas elas foram escolhidas pelo fato de serem um comando base para usuários do *Xbox*, então são ações que o usuário já poderia ter feito ou visto antes, tornando-as mais familiares. As classes E e H não são tão presentes no dia a dia, mas são simples de fazer, logo foram escolhidas. Já a ação B é mais elaborada e mais dinâmica que as outras, mas não o suficiente para ser desconfortável para os usuários. Ela foi adicionada para testar se o classificador seria capaz de discriminar uma ação mais elaborada.

Figura 19 – Ações padrão da base de dados.



Fonte: Autor.

O *Kinect* é capaz de capturar quadros a uma taxa de 30 FPS (quadros por segundo), porém como esse projeto envolve o processamento de várias técnicas, como o pré-processamento da imagem, a classificação da ação, os sinais de controle do robô, a estimação de velocidade e o desvio de obstáculos. Logo, foi definido que as ações fossem capturadas com uma taxa de 15 FPS, não colocando o *Kinect* no seu limite de trabalho para caso as outras técnicas tomassem uma parcela maior do processamento, de modo que a taxa de captura não comprometesse a classificação da ação.

A janela de captura utilizada foi de 1,66 segundos. Isso foi obtido empiricamente, almejando uma janela grande o suficiente para fazer as ações de forma confortável, sem pressa. Em especial a ação padrão B, que é a mais complexa.

Dado a taxa de captura escolhida e a janela de captura, cada ação será representada por um total de 25 quadros (15 FPS por 1,66 segundos). Cada quadro consiste em um grupo de características, que são armazenadas em uma matriz de forma:

$$\mathbf{F}_k = \begin{bmatrix} x_1 & y_2 & z_3 \\ \vdots & \vdots & \vdots \\ x_i & y_i & z_i \end{bmatrix} \quad (3.1)$$

onde k indica o quadro, para $k = 1, \dots, 25$, e as colunas representam as coordenadas x , y e z da i -ésima junta do esqueleto, com $i = 1, \dots, 25$.

Aos olhos de um observador humano, uma ação feita em diferentes posições podem ser assumidas como idênticas, o que não ocorre quando vistas pelo sensor *Kinect*. No caso, duas ações poderão ter coordenadas diferentes para suas juntas, mesmo que pertençam a mesma classe, simplesmente por terem sido feitas em lugares distintos. Para evitar este problema, por opção, todas as características são centralizadas em relação às coordenadas da junta do ombro esquerdo. Após centralização, a ação completa é armazenada como a concatenação subsequente da matriz de características de todos os quadros, dada por

$$\mathbf{A}_n = [\mathbf{F}_1 \dots \mathbf{F}_k]. \quad (3.2)$$

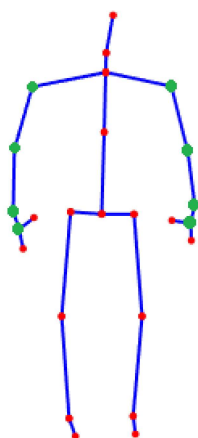
Este trabalho contou com duas bases de dados distintas, a primeira foi coletada por duas pessoas e cada uma forneceu uma quantidade de amostras diferentes para cada uma das 11 classes, gerando uma distribuição não homogênea das amostras. Essa base de dados contém 100 amostras de cada uma das classes padrão e das neutras, totalizando 1100 amostras.

A base de dados final do trabalho foi coletada por cinco pessoas e cada uma forneceu a mesma quantidade de amostras para cada uma das 11 classes. Essa base de dados contém 50 amostras de cada uma das classes padrão e das neutras, dez por pessoa, totalizando 550 amostras. Esta redução foi possível em função da capacidade de generalização da estratégia proposta, a qual será detalhada posteriormente.

3.3 *Dynamic Time Warping* - DTW

Esse método consiste na comparação de séries temporais para compreender quão similares são. Este trabalho utilizou características extraídas pelo sensor *Kinect* para representar cada quadro da ação. Com intuito de selecionar apenas as características mais relevantes para a ação, apenas 8 juntas foram utilizadas, como mostra a Figura 20, em que os ombros, cotovelos, pulsos e mãos estão destacados no esqueleto com a cor verde. Foram escolhidas tais juntas por se tratarem das partes do corpo que mais se movem durante a realização das ações apresentadas na Figura 19.

Figura 20 – Juntas de destaque para a etapa de classificação.



Fonte: Autor.

Cada quadro representa um esqueleto humano com suas juntas em um espaço tridimensional. A forma de armazenamento explicada na criação da base de dados, concatenando cada quadro subsequente de forma lateral, gera uma matriz onde ao se percorrer as colunas da i -ésima linha, está acompanhando a evolução temporal da posição tridimensional da i -ésima junta.

Para a classificação, tomou-se como exemplo uma amostra de cada classe. Para classificar uma amostra nova, ela é comparada com os exemplos de todas as classes, calculando sua função de custo. A classe com maior correspondência, isto é, com a menor função de custo, é a associada à amostra a ser classificada. O esquema de classificação utilizando o *DTW* pode ser encontrado no fluxograma da Figura 21.

3.4 Redes Neurais Artificiais - RNA

Este trabalho propõe um método de reconhecimento de ações utilizando uma rede neural simples com um pré-processamento adequado às entradas, para que dessa forma possa identificar diferentes classes com alta precisão e uma base de dados reduzida.

Figura 21 – Fluxograma da classificação por DTW.



Fonte: Autor.

De forma similar à utilizada na classificação DTW, apenas as 8 juntas da Figura 20 são utilizadas no processo de classificação, obtendo \mathbf{A}_n com dimensão 8×75 . Na sequência, \mathbf{A}_n é transformada em um vetor de característica, contendo seus autovalores λ , dados por $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_8$. O vetor λ é agora a informação de entrada do classificador.

A rede neural utilizada tem apenas uma camada escondida com 30 neurônios tendo suas camadas totalmente conectadas. A função de ativação usada é tangente *sigmoid* para a camada escondida e a função *softmax* para a camada de classificação. Por fim, o método de treinamento utilizado é Regularização Bayesiana com otimização de *Levenberg-Marquadt*. O esquema de classificação utilizando Redes Neurais pode ser encontrado no fluxograma da Figura 22.

Figura 22 – Fluxograma da classificação por Redes Neurais.



Fonte: Autor.

4 Experimentos e Resultados

Para validar o classificador proposto nesse trabalho e também a utilização dele como meio de comunicação com a plataforma *Pioneer 3DX*, foram feitos seis testes, divididos em três testes *offline* (simulações) e três testes *online* (experimentos). Eles são:

- Simulação *Offline 1* - Essa simulação compara o desempenho do classificador *baseline* DTW com o desempenho do classificador proposto. A simulação é realizada utilizando a primeira base de dados obtida contendo duas pessoas em suas amostras, com distribuição não simétrica das amostras de cada classe, isto é, não tem o mesmo número de amostras coletadas por cada uma das pessoas envolvidas para cada classe. Essa simulação visa uma comparação preliminar do desempenho dos dois classificadores. A base de dados é maior e são poucas pessoas fornecendo amostras.
- Simulação *Offline 2* - Essa simulação compara o desempenho do classificador *baseline* DTW com o desempenho do classificador proposto. A simulação é realizada usando a base de dados final, que foi obtida com uma distribuição homogênea das amostras de cada classe, tendo dez amostras de cada uma das cinco pessoas que participaram. Essa simulação visa uma comparação entre o desempenho dos dois classificadores num cenário em que a base de dados tem distribuição homogênea das suas amostras.
- Simulação *Offline 3* - Essa simulação é feita utilizando uma base de dados reduzida. O classificador proposto é treinado apenas com metade das amostras da base de dados final e o seu desempenho é testado na outra metade das amostras da base de dados. Essa simulação visa analisar o desempenho do classificador em um cenário que ele foi treinado com uma base de dados bastante reduzida, tendo apenas cinco amostras de cada pessoa para cada classe.
- Experimento *Online 1* - Esse experimento foi feito com dois usuários reproduzindo trinta vezes cada classe em uma ordem aleatória. A classificação se deu em tempo real. Esse teste visa garantir que a base de dados obtida não esteja enviesada, analisando o desempenho do classificador com os usuários executando as ações de verdade.
- Experimento *Online 2* - Esse experimento visa, de forma lúdica, testar a comunicação humano-robô por meio de ações utilizando o classificador proposto. Isso é feito atrelando uma posição predefinida a cada ação padrão. No caso, as posições representam as casas de um jogo da velha. O jogo foi jogado por dois usuários por um total de três partidas seguidas.

- Experimento *Online 3*- Esse experimento é a junção de tudo apresentado nesta dissertação, o usuário comunica com o robô utilizando ações, o robô deve realizar tarefas como seguir o usuário, se aproximar dele, ir para lugares predeterminados enquanto é capaz de desviar de obstáculos e transportar objetos para o usuário.

4.1 Simulações

As simulações foram realizados usando apenas informações da base de dados obtida nesse trabalho. Todos os resultados se encontram em matrizes de confusão, onde os resultados se dão em porcentagem em relação ao total de amostras utilizadas, a última linha da tabela mostra a porcentagem de falsos positivos (FP) e a última coluna mostra a porcentagem de falsos negativos (FN) de cada classe. As colunas são as classes desejadas e as linhas as classes preditas pelo sistema inteligente.

4.1.1 Simulação 1

Nessa simulação utilizou-se a primeira base de dados obtida para esse trabalho. Ela tem uma distribuição não homogênea entre suas amostras, isto é, cada um dos dois usuários forneceu um número diferente de amostras para cada classe. Ela contém um total de 100 amostras para cada classe (9 classes padrão e 2 classes neutras), totalizando 1.100 amostras.

Essa simulação tem o objetivo de comparar de forma preliminar o desempenho do DTW com o método proposto no cenário de uma base de dados com poucas pessoas inclusas nela e uma distribuição não homogênea das amostras de cada classe.

Os testes com DTW tomaram um exemplo de cada classe como referência para calcular a função de custo de cada outra amostra. Para a rede neural, foi utilizada validação cruzada de *10-fold*. Os resultados se encontram nas Tabelas 1 e 2, onde a primeira é dos resultados do DTW e a segunda do classificador proposto com RNA.

Inicialmente, é importante ressaltar a aplicação desse trabalho. Ele propõe um método de reconhecimento de ações para ser utilizado em uma aplicação de robótica social, onde as ações do usuário serão as fontes de comando para o agente. As ações neutras devem ter um rigor maior em relação à sua discriminação, afinal uma confusão com uma ação padrão inviabiliza a aplicação.

No âmbito de ações padrão, ter falsos negativos é algo extremamente indesejável, pois implica que ou o usuário pediu uma ação padrão e o agente executa outra, ou o usuário fez uma ação neutra e o sistema irá interpretar como uma ação padrão. Para ações neutras, falsos positivos não são tão danosos, pois isso implica que ou o usuário fez uma ação neutra e o classificador a identificou como a outra, ainda sim não realizando

Tabela 1 – Resultados das simulações do DTW para a simulação 1.

		Classes Desejadas											
		A	B	C	D	E	F	G	H	I	J	K	FN
Classes Preditas	A	99	-	-	-	-	-	-	-	-	-	-	1
	B	-	100	4	-	-	-	-	-	-	1	16	17
	C	-	-	93	-	-	-	-	-	-	-	-	-
	D	-	-	-	94	-	-	-	-	-	-	-	-
	E	-	-	-	-	96	-	-	-	-	-	-	-
	F	-	-	-	6	-	100	-	-	-	-	-	6
	G	-	-	2	-	-	-	100	-	-	-	-	2
	H	-	-	-	-	-	-	-	100	-	-	-	-
	I	-	-	-	-	-	-	-	-	62	38	1	39
	J	1	-	-	-	4	-	-	-	38	61	-	41
	K	-	-	1	-	-	-	-	-	-	-	83	1
	FP	1	-	7	6	4	-	-	-	38	39	17	89,8

Tabela 2 – Resultados das simulações da RNA para a simulação 1.

		Classes Desejadas											
		A	B	C	D	E	F	G	H	I	J	K	FN
Classes Preditas	A	100	-	-	-	-	-	-	-	-	2	-	2
	B	-	100	-	-	-	-	-	-	-	-	-	-
	C	-	-	100	-	-	-	-	-	-	-	-	-
	D	-	-	-	100	-	-	-	-	-	-	-	-
	E	-	-	-	-	100	-	-	-	-	-	1	1
	F	-	-	-	-	-	96	-	-	2	-	2	4
	G	-	-	-	-	-	-	98	-	-	-	-	-
	H	-	-	-	-	-	-	-	100	-	-	-	-
	I	-	-	-	-	-	2	-	-	98	-	2	4
	J	-	-	-	-	-	-	-	-	-	98	-	-
	K	-	-	-	-	-	2	2	-	-	-	95	4
	FP	-	-	-	-	-	4	2	-	2	-	5	98,6

nenhuma tarefa, ou então que uma ação padrão foi confundida com uma ação neutra, o que é indesejado, mas não tão grave quanto o contrário.

Comparando os valores das Tabelas 1 e 2, é possível observar que o DTW teve mais falsos negativos que o método proposto com RNA. As classes B, I, J tiveram o pior desempenho nesse aspecto.

Para a classe B, 17% das amostras classificadas como ela são falsos negativos, onde 16% das amostras da classe K sendo falso positivo, o que é bastante indesejável, pois esse cenário implica que uma ação neutra (K) será comumente interpretada como ação padrão (B), então o agente, em uma situação que ele não deveria realizar nenhuma tarefa, irá realizar uma que não foi requisitada.

A classe I teve 39% das amostras classificadas como ela como falsos negativos,

onde 38% das amostras da classe J foram falsos positivos. Isso é um cenário altamente indesejável, pois além de ter uma taxa alta de falso negativo, isso implica em uma situação onde o usuário emitiu uma ação neutra (J), e o robô identificou como uma ação padrão (I), então ele, em uma situação que deveria realizar nenhuma tarefa, irá realizar uma que não foi requisitada.

A classe J teve 41% das amostras classificadas como ela como falsos negativos, onde 38% das amostras da classe I foram falso positivo. Isso é ruim para a aplicação, especialmente devido à alta taxa de erro, porém não gera o tipo de efeito mais grave na prática, pois nesse caso o usuário está enviando um comando com uma ação padrão (I) e o agente está identificando como uma ação neutra (J), não realizando tarefas. Assim o usuário precisaria requisitar a tarefas novamente.

O método com RNA teve uma baixa quantidade de falsos negativos, sendo parte deles na classe K. Isso não é tão danoso para a aplicação, pois, quando eles ocorrerem, a maior parte do tempo será uma situação onde o usuário dará um comando e o agente não realizará nenhuma tarefa. Daí, o usuário terá que repetir o comando.

As taxas de acerto média dos dois métodos foram de 89,8% para o DTW e de 98,6% para o classificador proposto. A relação de falsos positivos e falsos negativos discutida acima, juntamente com a diferença de taxa de acerto média, mostra que o método proposto se saiu consideravelmente melhor nesse teste.

4.1.2 Simulação 2

Nesta simulação utilizou-se a base de dados final obtida para esse trabalho. Ela tem uma distribuição homogênea entre suas amostras, isto é, cada um dos cinco usuários forneceu o mesmo número de amostras para cada classe. Ela contém um total de 50 amostras para cada classe (9 classes padrão e 2 classes neutras), totalizando 550 amostras.

Esta simulação tem o objetivo de comparar o desempenho do DTW com o método proposto num cenário com mais pessoas inclusas na base de dados e com uma distribuição homogênea das amostras de cada classe coletadas por cada pessoa.

Os testes com DTW tomaram um exemplo de cada classe como referência para calcular a função de custo de cada outra amostra. Para a rede neural, foi utilizada validação cruzada de *10-fold*. Os resultados se encontram nas Tabelas 3 e 4, onde a primeira é dos resultados do DTW e a segunda do classificador proposto com RNA.

Comparando os valores das Tabelas 3 e 4, é possível observar que o DTW teve consideravelmente mais falsos negativos que o método proposto com RNA. As classes B, I, J tiveram o pior desempenho nesse aspecto.

Para a classe B, 40% das amostras classificadas como ela são falsos negativos.

Tabela 3 – Resultados das simulações do DTW para a simulação 2.

		Classes Desejadas											
		A	B	C	D	E	F	G	H	I	J	K	FN
Classes Preditas	A	50	-	-	-	-	-	-	-	-	-	-	-
	B	50	76	-	-	-	-	-	-	-	-	-	40
	C	-	-	86	-	-	-	-	-	-	-	-	-
	D	-	-	-	98	-	-	-	-	-	-	-	-
	E	-	-	-	-	100	-	-	-	-	-	-	-
	F	-	-	-	-	-	100	-	-	-	-	-	-
	G	-	-	2	-	-	-	100	-	-	-	-	2
	H	-	-	-	-	-	-	-	90	-	-	-	-
	I	-	24	-	-	-	-	-	-	98	-	-	20
	J	-	-	14	2	-	-	-	10	2	100	40	41
	K	-	-	-	-	-	-	-	-	-	-	60	-
	FP	50	24	14	2	-	-	-	10	2	-	60	87,1

Tabela 4 – Resultados das simulações da RNA para a simulação 2.

		Classes Desejadas											
		A	B	C	D	E	F	G	H	I	J	K	FN
Classes Preditas	A	96	-	-	-	-	-	-	-	-	-	-	-
	B	2	100	-	-	-	-	-	-	-	-	-	2
	C	-	-	100	-	-	-	-	-	-	-	-	-
	D	-	-	-	96	-	-	-	-	-	-	-	-
	E	-	-	-	-	100	-	-	-	-	-	-	-
	F	-	-	-	-	-	100	-	-	-	-	-	-
	G	2	-	-	4	-	-	100	-	2	-	-	5,7
	H	-	-	-	-	-	-	-	100	-	-	-	-
	I	-	-	-	-	-	-	-	-	98	-	-	-
	J	-	-	-	-	-	-	-	-	-	100	-	-
	K	-	-	-	-	-	-	-	-	-	-	100	-
	FP	4	-	-	4	-	-	-	-	2	-	-	99,1

50% das amostras da classe A são falsos positivos sendo classificadas como B, o que é bastante indesejável, pois esse cenário implica que uma ação padrão (A) será comumente interpretada como outra ação padrão (B). Então, o agente em uma situação que ele deveria realizar uma tarefa, irá realizar uma outra que não foi requisitada.

A classe I teve 20% das amostras classificadas como falsos negativos. 24% das amostras da classe B são falsos positivos sendo classificadas como I. Isso é um cenário altamente indesejável, pois além de ter uma taxa alta de falso negativo, o que, pelo mesmo motivo da análise dos falsos negativos da classe B, é bastante indesejável para a aplicação desse trabalho.

A classe J teve 41% das amostras classificadas como falsos negativos. 40% das amostras da classe K são falsos positivos sendo classificadas como J. Isso não é tão ruim

para a aplicação, pois gera uma situação em que uma ação neutra é confundida por outra, então o objetivo principal é que o agente não realize tarefa alguma e é o que acontece, mesmo confundindo as duas ações neutras. Outras parcelas importantes dos falsos negativos da classe J são os falsos positivos das classes C e H, estando em 14% e 10%, respectivamente, do seu total de amostras. Isso não é desejável pois nesse cenário, o robô deveria realizar tarefa alguma, dado que J é uma ação neutra, e ele irá realizar alguma, dado que C e H são ações padrão.

O método com RNA teve poucos falsos negativos e positivos tendo uma precisão de 99,1% e o DTW apenas 87,1%. Com isso pode ser concluído que o DTW não escalou tão bem com mais pessoas na base de dados, quando comparada a RNA proposta. Além disso, a RNA não teve nenhum falso negativo nas classes neutras, o que sugere que aumentar o número de pessoas de duas para cinco na base de dados surtiu um efeito positivo na generalização das classes neutras e uma leve melhoria na generalização.

4.1.3 Simulação 3

Nessa simulação utilizou-se a base de dados final obtida para este trabalho. Ela tem uma distribuição homogênea entre suas amostras, isto é, cada um dos cinco usuários forneceu o mesmo número de amostras para cada classe. Ela contém um total de 50 amostras para cada classe (9 classes padrão e 2 classes neutras), totalizando 550 amostras.

Essa simulação tem o objetivo de avaliar a capacidade de classificação da RNA com uma base de dados reduzida, usando apenas 25 amostras (5 de cada pessoa) de cada classe para seu treinamento.

A RNA foi treinada com apenas metade da base de dados e validada com a outra metade. Utilizou-se validação cruzada de *10-fold* para obter os resultados da Tabela 5.

Analisando as Tabelas 4 e 5 e, pode-se ver que o desempenho em relação ao treinamento utilizando a metade da base de dados e a base de dados inteira não foi tão discrepante. Houve uma leve piora no desempenho geral, caindo a média de acerto de 99,1% para 96,7%, porém pode se dizer que ainda teve um desempenho satisfatório dado o baixo número de falsos positivos e negativos e uma parcela significativa dos falsos negativos foram na classe neutra K, implicando que esses erros de classificação não irão levar o agente a fazer um comando incorreto ou não solicitado, apenas não irá realizar o comando que o usuário pediu.

4.2 Experimentos

Para realização dos experimentos reais, alguns pontos devem ser destacados. Primeiramente, os experimentos foram realizados com a RNA treinada com a base de dados

Tabela 5 – Resultados da simulação para o método proposto com RNA para a simulação 3.

		Classes Desejadas											
		A	B	C	D	E	F	G	H	I	J	K	FN
Classes Preditas	A	96	-	-	-	-	-	-	-	-	-	-	-
	B	4	100	-	4	-	-	-	-	-	-	-	7,5
	C	-	-	96	-	-	-	-	-	-	-	-	-
	D	-	-	-	96	4	-	8	-	-	-	-	11,2
	E	-	-	-	-	96	-	-	-	-	-	-	-
	F	-	-	-	-	-	96	-	-	-	-	-	-
	G	-	-	-	-	-	-	92	-	-	-	-	-
	H	-	-	-	-	-	-	-	96	-	-	-	-
	I	-	-	-	-	-	-	-	-	96	-	-	-
	J	-	-	-	-	-	-	-	-	4	100	-	3,9
	K	-	-	4	-	-	4	-	4	-	-	100	10,8
	FP	4	-	4	4	4	4	8	4	4	-	1	96,7

completa. Segundo, uma aplicação de HRI requer que o agente seja capaz de identificar as ações do usuário assim que elas são realizadas. Neste caso, um gatilho foi implementado com o intuito de impedir que o algoritmo de classificação fique em execução ininterruptamente, gerando possíveis classificações erradas. Então para sinalizar para o sistema que o usuário irá iniciar um gesto, ele deve abrir qualquer uma de suas mãos. Dessa forma, o algoritmo entende que deve começar a extrair as características para classificação. Esse gatilho é possível devido à uma funcionalidade do *Kinect v2.0*, que é capaz de retornar se a mão do usuário esta fechada, aberta, ou apenas com os dedos indicador e do meio levantados.

4.2.1 Experimento 1

Esse teste visa garantir que a base de dados obtida para esse trabalho não está enviesada, invalidando os bons resultados das simulações.

Para isso, duas pessoas que fazem parte da base de dados usada para o treinamento da RNA, foram requisitadas para realizar 30 vezes uma cada das 11 classes de ações. Com o intuito de evitar que ao longo do experimento, o usuário possa corrigir os erros de repetições recentes no caso de ter que fazer a mesma ação seguidas vezes, foi gerado uma lista de ordem aleatória com as 30 repetições de cada classe que cada usuário deveria fazer. Essa abordagem evita que ações repetidas sejam realizadas em sequência. Os resultados se encontram na Tabela 6.

É possível observar nos resultados desse experimento, que o método proposto nesse trabalho é confiável para aplicações práticas, dado sua taxa de acerto de 98%. Com poucos falsos negativos, sendo que a maior parte deles é nas classes neutras J e K, o que leva a uma situação não tão grave, onde o usuário pede uma tarefa e o agente não realiza

Tabela 6 – Resultados para os testes *online*

	Classes Desejadas												
	A	B	C	D	E	F	G	H	I	J	K	FN	
Classes Preditas	A	100	5	-	-	-	-	6,7	-	-	-	-	10
B	-	91,6	-	-	-	-	-	-	-	-	-	-	-
C	-	-	96,7	-	-	-	-	-	-	-	-	1,7	1,7
D	-	-	-	100	-	-	-	-	-	-	-	-	-
E	-	-	-	-	98,3	-	-	-	-	-	-	-	-
F	-	-	-	-	-	100	-	-	-	-	-	-	-
G	-	-	-	-	-	-	93,3	-	-	-	-	-	-
H	-	-	-	-	1,7	-	-	100	-	-	-	-	1,7
I	-	1,6	-	-	-	-	-	-	100	-	-	-	1,6
J	-	-	-	-	-	-	-	-	-	100	-	-	-
K	-	1,8	3,3	-	-	-	-	-	-	-	98,3	5	5
FP	-	8,4	3,3	-	1,7	-	6,7	-	-	-	40	98	98

nenhuma.

4.2.2 Experimento 2

Com a intenção de implementar de forma lúdica a comunicação utilizando o classificador proposto e atrelar os comandos à posições predeterminadas, uma competição do jogo da velha foi realizada por dois usuários que fazem parte da base de dados usada para treinar a RNA, os quais enviavam comandos de posição através das ações. Em outras palavras, os usuários enviam comandos via ação e o robô vai para o lugar que aquela classe de ação representa. Em suma, existem nove classes de ação padrão para nove lugares diferentes no jogo da velha. Três jogos seguidos foram jogados e não houve nenhuma interpretação errada das ações, mesmo tendo dois usuários distintos envolvidos. Cada espaço no jogo da velha é ligado à uma classe de ação padrão, como indicado na Figura 23.

No tocante à navegação do robô, os valores utilizados para as matrizes de ganhos do controlador de posição foram $k_1 = 0,35$ e $k_2 = 0,45$. Em adição, para melhorar a navegação, a compensação dinâmica proposta por Martins, Sarcinelli-Filho e Carelli (2017) foi utilizada, uma vez que contempla os parâmetros físicos do robô, tais como massa, momento de inércia, resistência elétrica do motor, constantes para torque, dentre outras.

A Figura 24 mostra fotos do experimento da primeira partida com a ação realizada para o comando e o lugar que o robô iria, marcado com um círculo verde.

É possível notar que todos os comandos foram identificados e realizados com sucesso, indo para o ponto desejado no campo do jogo. O experimento completo se encontra no vídeo: <<https://youtu.be/NVXGA58JhtM>>.

Figura 23 – Classes de ação padrão e seus respectivos lugares no Jogo da Velha.

G	D	A
H	E	B
I	F	C

Fonte: Autor.

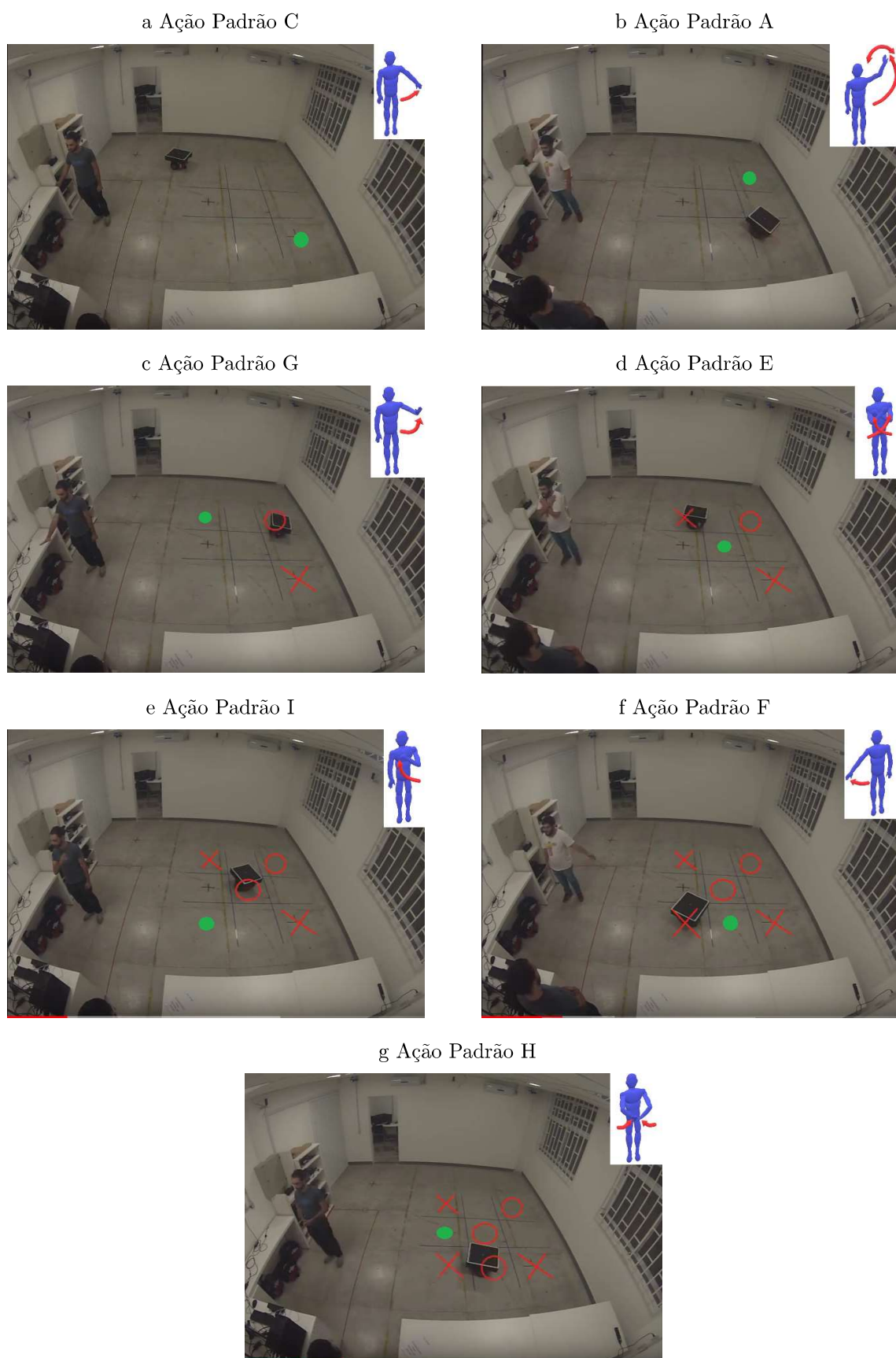
4.2.3 Experimento 3

Para finalizar a validação da proposta, o experimento aqui realizado envolve comandar o robô para a realização de sete tarefas. São elas:

1. Teste de Comunicação (A e D). O agente irá sinalizar que a comunicação está funcionando emitindo um som pré-programado.
2. Seguir (I). O robô irá começar a rotina de seguir o usuário, onde seu destino desejado é 2.25 metros atrás do usuário, dentro de sua zona social, para ter uma zona de segurança durante a execução da tarefa.
3. Aproximar (H). O agente irá se aproximar do usuário, mudando a distância do destino desejado para 1 metro do usuário, mudando para sua zona pessoal.
4. Parar (G). O robô irá parar a tarefa que está sendo executada e ficará parado até que algum outro comando seja dado.
5. Ir até a Localização A (C e F). O robô irá até uma localização predefinida. Após atingir o destino, ele irá girar em torno de si mesmo a procura do usuário.
6. Ir até a Localização B (E). O agente irá até uma localização predefinida. Após atingir o destino, ele irá girar em torno de si mesmo a procura do usuário.
7. Retornar à origem (B). O robô irá retornar à origem, isto é, onde ele estava ao começar a navegação.

Figura 25 são fotos dos momentos em que os comandos foram dados e a respectiva ação relacionada a ele.

Figura 24 – Experimento 1 Jogo da Velha.



Fonte: Autor.

Figura 25 – Experimento de Navegação .

a Ação Padrão A



b Ação Padrão I



c Ação Padrão H



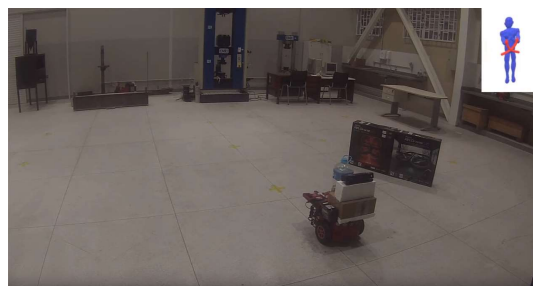
d Ação Padrão C



e Ação Padrão H



f Ação Padrão E



g Ação Padrão G



h Ação Padrão B



Fonte: Autor.

O primeiro comando é um teste de comunicação, onde o agente responde com um som ao comando do usuário. O som emitido corresponde a um dos sons característicos do robô R2D2 da franquia *Star Wars*. Esse som foi escolhido por ser famoso e para que, ao se familiarizar com a resposta, o usuário possa se sentir mais a vontade com o robô navegando próximo dele.

O segundo é o comando para o robô seguir o usuário e este vai até próximo de uma mesa, onde o robô mantém uma distância de 2,25m, estando dentro de sua zona social para evitar desconfortos.

O terceiro comando é o de aproximação, para que o robô chegue mais perto e altere da zona social para a zona pessoal, para que o usuário possa colocar uma carga sobre ele.

O quarto comando é dado para o robô ir até uma das localizações pré definidas, levando tal carga. Após chegar ao local, o agente gira em torno de seu eixo a procura do usuário e para ao encontrá-lo.

O quinto comando é para se aproximar novamente para o usuário colocar uma segunda carga sobre o robô.

O sexto comando é dado para o robô ir até uma outra localização predefinida, próximo de uma mesa. Após chegar ao local, ele gira em torno de seu eixo a procura do usuário e para ao encontrá-lo. Durante o trajeto robô detecta e desvia de um obstáculo.

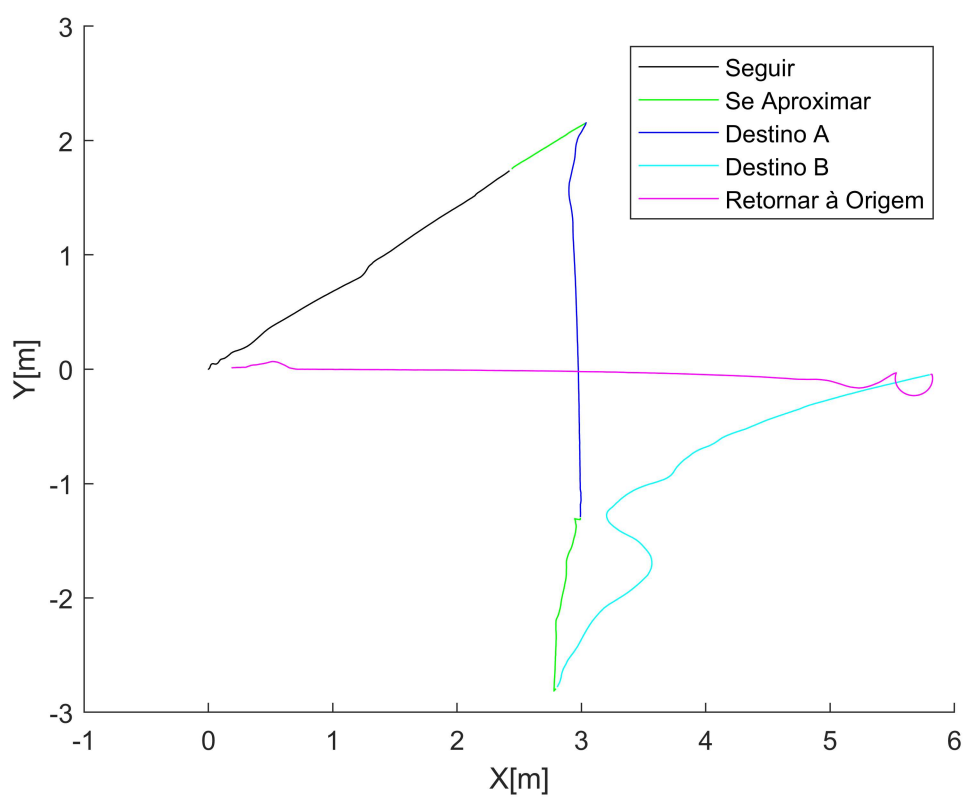
O sétimo comando é para que o agente fique parado, para que o usuário se aproxime, coloque uma terceira carga e colocar uma das cargas, anteriormente transportada, sobre a mesa.

Por fim, o oitavo comando é dado para que o agente volte à posição de origem.

O caminho seguido pelo agente se encontra na Figura 23 e o experimento como um todo em <<https://youtu.be/uKUX08v4990>>.

Na Figura 26, é possível perceber que o caminho todo foi percorrido sem oscilações bruscas. Pode se observar uma distorção na rota do robô entre as últimas ações, durante o retorno à origem. Isto é decorrência dos erros de odometria do robô, que são agravados em manobras de curva ou giro. No vídeo é possível observar que não ocorre essa oscilação, nas etapas prévias ao retorno a origem.

Figura 26 – Experimento de navegação aplicada.



Fonte: Autor.

5 Considerações Finais

Esta dissertação apresenta um método de reconhecimento de ações baseado em redes neurais artificiais que se apoia em redução da dimensionalidade das entradas. O método proposto é utilizado como via de comunicação para a Interação Humano-Robô, e a comunicação e aplicação são validadas através de experimentos, onde comandos são dados via ações e tarefas são cumpridas por um robô móvel.

A validação se deu através da comparação do desempenho do método proposto com o DTW, que já é consolidado na literatura. Analisou-se a escalabilidade dos dois métodos para pessoas diferentes fazendo as mesmas ações. Testes *online* foram realizados para assegurar a viabilidade de aplicação do método proposto e garantir que a base de dados não estava enviesada. A RNA necessitou de poucas amostras na base de dados para ter um desempenho adequado para aplicações em experimentos reais .

Dadas as necessidades particulares da aplicação em questão, onde é essencial que a taxa de acerto na classificação das ações padrões seja alta e a taxa de falsos positivos nas ações neutras seja baixa, conclui-se que o método proposto tem aplicabilidade e é eficiente, uma vez que ele teve um desempenho superior em todas as comparações com o DTW.

A tarefa de interação Humano-Robô foi bem sucedida, dado que na comunicação via ações não houve erros, e que durante a navegação o robô respeitou as zonas de interação do usuário, mantendo-se na sua zona social durante a tarefa de seguir e só aproximando para a zona pessoal com a demanda do usuário. O robô ainda se mostrou capaz de desviar de obstáculos utilizando o método de desvio tangencial.

Pontos em abertos que podem ser endereçados no futuro incluem a definição de gestos ou ações pelo usuário, isto é, criar um mecanismo que permita que a base de dados seja criada diretamente pelo usuário, onde ele escolhe quais gestos irão ser atrelados a qual tarefa realizada pelo robô. Outros pontos relevantes são a implementação da funcionalidade de SLAM (Localização e Mapeamento Simultâneos, do inglês *Simultaneous Location and Mapping*), e navegação híbrida, a fim de corrigir os erros de odometria acumulados durante a navegação.

Referências

- 3D NATIVES. *3D scanning through Structured Light*. 2020. Disponível em: <<https://www.3dnatives.com/en/structured-light-projection-3d-scanning/>>. Acesso em: 08 de janeiro de 2020. Citado na página 43.
- ABDELNASSER, H.; YOUSSEF, M.; HARRAS, K. A. Wigest: A ubiquitous wifi-based gesture recognition system. In: IEEE. *2015 IEEE Conference on Computer Communications (INFOCOM)*. Hong Kong, China, 2015. p. 1472–1480. Citado na página 28.
- AGAHIAN, S.; NEGIN, F.; KÖSE, C. An efficient human action recognition framework with pose-based spatiotemporal features. *Engineering Science and Technology, an International Journal*, Elsevier, 2019. Citado na página 29.
- AGGARWAL, J. K.; RYOO, M. S. Human activity analysis: A review. *ACM Computing Surveys (CSUR)*, ACM, v. 43, n. 3, p. 16, 2011. Citado na página 29.
- AGGARWAL, J. K.; XIA, L. Human activity recognition from 3d data: A review. *Pattern Recognition Letters*, Elsevier, v. 48, p. 70–80, 2014. Citado na página 29.
- APPLE INSIDER. *MeeBot 2.0 Educational Robotics*. 2019. Disponível em: <<https://appleinsider.com/articles/19/10/08/meebot-20-educational-robotics-kit-ships-as-apple-store-exclusive>>. Acesso em: 25 dezembro de 2019. Citado na página 22.
- AV MAGAZINE. *AV Magazine Gesture Control*. 2019. Disponível em: <<https://www.avinteractive.com/news/actis-tech-unveils-gesture-control-solutions-29-08-2012/#!>> Acesso em: 25 dezembro de 2019. Citado 2 vezes nas páginas 17 e 33.
- BARNACHON, M. et al. Ongoing human action recognition with motion capture. *Pattern Recognition*, Elsevier, v. 47, n. 1, p. 238–247, 2014. Citado na página 29.
- BARRIENTOS, J.; CATALDO, E. *Modelagem Estocástica da Voz e Classificação dos Sinais Obtidos Usando Redes Neurais Artificiais*. 2015. Citado na página 35.
- BASILIO, V. T.; CARVALHO, K. B. de; BRANDÃO, A. S. Reconhecimento de ações por rna em aplicações de robótica social. In: GALOÁ 2019. *Anais do 14º Simpósio Brasileiro de Automação Inteligente(SBAI)*. Ouro Preto, MG, Brasil, 2019. p. 23–28. Citado na página 26.
- BENITTI, F. B. V. Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education*, Elsevier, v. 58, n. 3, p. 978–988, 2012. Citado na página 21.
- BENTLEY. *Bentley Mobile Mapping App*. 2019. Disponível em: <<https://www.bentley.com/en/products/product-line/asset-performance/bentley-map-mobile>>. Acesso em: 25 dezembro de 2019. Citado 2 vezes nas páginas 17 e 33.
- BJÖRCK, Å. *Numerical methods for least squares problems*. Filadélfia, PA, USA: SIAM, 1996. Citado na página 36.

- BODIROŽA, S.; STERN, H. I.; EDAN, Y. Dynamic gesture vocabulary design for intuitive human-robot dialog. In: ACM. *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. Boston, MA, USA, 2012. p. 111–112. Citado na página 15.
- BOURGEOIS, P.; HESS, U. The impact of social context on mimicry. *Biological psychology*, Elsevier, v. 77, n. 3, p. 343–352, 2008. Citado na página 14.
- BRANDÃO, A. S.; SARCINELLI-FILHO, M.; CARELLI, R. An analytical approach to avoid obstacles in mobile robot navigation. *International Journal of Advanced Robotic Systems*, SAGE Publications Sage UK: London, England, v. 10, n. 6, p. 278, 2013. Citado 2 vezes nas páginas 39 e 40.
- BURDEN, F.; WINKLER, D. Bayesian regularization of neural networks. In: *Artificial neural networks*. USA: Springer, 2008. p. 23–42. Citado 2 vezes nas páginas 35 e 36.
- CABIBIHAN, J.-J.; SO, W.-C.; PRAMANIK, S. Human-recognizable robotic gestures. *IEEE Transactions on Autonomous Mental Development*, IEEE, v. 4, n. 4, p. 305–314, 2012. Citado na página 28.
- CANAL, G.; ESCALERA, S.; ANGULO, C. A real-time human-robot interaction system based on gestures for assistive scenarios. *Computer Vision and Image Understanding*, Elsevier, v. 149, p. 65–77, 2016. Citado na página 28.
- CARUSO, L.; RUSSO, R.; SAVINO, S. Microsoft kinect v2 vision system in a manufacturing application. *Robotics and Computer-Integrated Manufacturing*, Elsevier, v. 48, p. 174–181, 2017. Citado na página 43.
- CARVALHO, K. B. de; OLIVEIRA, G. C. de; BRANDÃO, A. S. Local planning using obstacle estimation and tangential escape. In: IEEE. *2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)*. João Pessoa, PB, Brasil, 2018. p. 303–307. Citado na página 26.
- CELEBI, S. et al. Gesture recognition using skeleton data with weighted dynamic time warping. In: *VISAPP (1)*. Barcelona, Espanha: [s.n.], 2013. p. 620–625. Citado na página 29.
- CHEN, B. et al. Online control programming algorithm for human-robot interaction system with a novel real-time human gesture recognition method. *International Journal of Advanced Robotic Systems*, SAGE Publications Sage UK: London, England, v. 16, n. 4, p. 1729881419861764, 2019. Citado na página 15.
- CHERLA, S. et al. Towards fast, view-invariant human action recognition. In: IEEE. *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Anchorage, AK, USA, 2008. p. 1–8. Citado na página 25.
- CHERUBINI, A. et al. Collaborative manufacturing with physical human-robot interaction. *Robotics and Computer-Integrated Manufacturing*, Elsevier, v. 40, p. 1–13, 2016. Citado na página 14.
- CHRISTENSEN, H. I.; PACCHIEROTTI, E. Embodied social interaction for robots. *AISB-05*, p. 40–45, 2005. Citado na página 19.

CLABAUGH, C. et al. Month-long, in-home case study of a socially assistive robot for children with autism spectrum disorder. In: ACM. *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. Chicago, IL, USA, 2018. p. 87–88. Citado na página 14.

DAUTENHAHN, K.; FONG, T.; NOURBAKHSI, I. *A survey of socially interactive robots*. 2014. Citado na página 18.

DREAMSTIME. *Hand Gestures*. 2019. Disponível em: <<https://www.dreamstime.com/human-hand-gesture-hand-gesture-illustration-vector-collection-hands-various-gesture-activity-simple>>. Acesso em: 25 dezembro de 2019. Citado na página 28.

DU, Y.; WANG, W.; WANG, L. Hierarchical recurrent neural network for skeleton based action recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Boston, MA, USA: [s.n.], 2015. p. 1110–1118. Citado na página 30.

ERIKSSON, J. Hands-off robotics for post-stroke arm rehabilitation. *Technical Report*, 2004. Citado na página 16.

EUREKALERT. *EurekaAlert: Socially Assistive Robotics*. 2019. Disponível em: <https://www.eurekaalert.org/pub_releases/2014-08/uosc-srh082814.php>. Acesso em: 25 dezembro de 2019. Citado na página 16.

FEIL-SEIFER, D.; MATARIC, M. J. Defining socially assistive robotics. In: IEEE. *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005*. Chicago, IL, USA, 2005. p. 465–468. Citado na página 16.

FEIL-SEIFER, D.; MATARIĆ, M. J. Socially assistive robotics. *IEEE Robotics & Automation Magazine*, IEEE, v. 18, n. 1, p. 24–31, 2011. Citado na página 14.

FERRER, G. et al. Robot social-aware navigation framework to accompany people walking side-by-side. *Autonomous robots*, Springer, v. 41, n. 4, p. 775–793, 2017. Citado na página 37.

FONG, T.; NOURBAKHSI, I.; DAUTENHAHN, K. A survey of socially interactive robots. *Robotics and autonomous systems*, Elsevier, v. 42, n. 3-4, p. 143–166, 2003. Citado 2 vezes nas páginas 16 e 18.

GAVIN, P. Henri. the levenberg-marquardt algorithm for nonlinear least squares curve-fitting problems: Department of civil and environmental engineering. *Duke University*, v. 10, n. 1, 2019. Citado na página 36.

GOODRICH, M. A.; SCHULTZ, A. C. et al. Human–robot interaction: a survey. *Foundations and Trends® in Human–Computer Interaction*, Now Publishers, Inc., v. 1, n. 3, p. 203–275, 2008. Citado na página 14.

GRIT DAILY. *Grit Daily Military*. 2019. Disponível em: <<https://gritdaily.com/who-will-win-the-war-of-tomorrow-look-to-the-advances-in-robotics-and-ai/>>. Acesso em: 25 dezembro de 2019. Citado na página 15.

HALL, E. T. The hidden dimension: man’s use of space in public and private the bodley head. *London, Sydney, Toronto*, v. 121, 1969. Citado na página 37.

- HANG, C. et al. Dynamic gesture recognition method based on improved dtw algorithm. In: IEEE. *2017 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII)*. Wuhan, China, 2017. p. 71–74. Citado na página 29.
- HAYKIN, S. *Neural networks: a comprehensive foundation*. USA: Prentice Hall PTR, 1994. Citado 2 vezes nas páginas 34 e 35.
- HERRERA, D. et al. Human interaction dynamics for its use in mobile robotics: Impedance control for leader-follower formation. *IEEE/CAA Journal of Automatica Sinica*, IEEE, v. 4, n. 4, p. 696–703, 2017. Citado na página 25.
- HINDS, P. J.; ROBERTS, T. L.; JONES, H. Whose job is it anyway? a study of human-robot interaction in a collaborative task. *Human-Computer Interaction*, L. Erlbaum Associates Inc., v. 19, n. 1, p. 151–181, 2004. Citado na página 14.
- HOGAN, K. *Can't get through: eight barriers to communication*. Gretna, Luisiana, EUA: Pelican Publishing, 2003. Citado na página 37.
- HOLOLENS. *Hololens Uc Berkeley*. 2019. Disponível em: <<https://hololens.reality.news/news/uc-berkeley-team-commands-drone-fleet-with-hololens-0176671/>>. Acesso em: 25 dezembro de 2019. Citado na página 15.
- HUTTENRAUCH, H.; EKLUNDH, K. S. Fetch-and-carry with cero: Observations from a long-term user study with a service robot. In: IEEE. *Proceedings. 11th IEEE International Workshop on Robot and Human Interactive Communication*. Berlim, Alemanha, 2002. p. 158–163. Citado na página 16.
- IBANEZ, R. et al. Easy gesture recognition for kinect. *Advances in Engineering Software*, Elsevier, v. 76, p. 171–180, 2014. Citado na página 28.
- INDEPENDENT UK. *Independent UK Mars Rover*. 2019. Disponível em: <<https://www.independent.co.uk/news/science/life-on-mars-rover-2020-mission-jezero-crater-nasa-a9206141.html>>. Acesso em: 25 dezembro de 2019. Citado na página 15.
- ISHIGURO, H. et al. Robovie: an interactive humanoid robot. *Industrial robot: An international journal*, MCB UP Ltd, v. 28, n. 6, p. 498–504, 2001. Citado na página 14.
- KANDA, T. et al. Development and evaluation of interactive humanoid robots. *Proceedings of the IEEE*, IEEE, v. 92, n. 11, p. 1839–1850, 2004. Citado na página 15.
- KOLROB. *KolRob: Collaborative Screw Assembly*. 2019. Disponível em: <<https://www.youtube.com/watch?v=tPZQSKHbyq8>>. Acesso em: 25 dezembro de 2019. Citado na página 16.
- KRUEGER, J. Extended cognition and the space of social interaction. *Consciousness and cognition*, Elsevier, v. 20, n. 3, p. 643–657, 2011. Citado na página 37.
- KUMAR, P. et al. Coupled hmm-based multi-sensor data fusion for sign language recognition. *Pattern Recognition Letters*, Elsevier, v. 86, p. 1–8, 2017. Citado na página 29.

- LÄNGLE, T.; WÖRN, H. Human-robot cooperation using multi-agent-systems. *Journal of intelligent and Robotic Systems*, Springer, v. 32, n. 2, p. 143–160, 2001. Citado na página 14.
- LEE, M. K. et al. The snackbot: documenting the design of a robot for long-term human-robot interaction. In: ACM. *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. San Diego, CA, USA, 2009. p. 7–14. Citado na página 24.
- LI, C. et al. Joint distance maps based action recognition with convolutional neural networks. *IEEE Signal Processing Letters*, IEEE, v. 24, n. 5, p. 624–628, 2017. Citado na página 24.
- LI, C. et al. Skeleton-based action recognition with convolutional neural networks. In: IEEE. *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. Hong Kong, China, 2017. p. 597–600. Citado na página 24.
- LI, W.; ZHANG, Z.; LIU, Z. Action recognition based on a bag of 3d points. In: IEEE. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. San Francisco, CA, USA, 2010. p. 9–14. Citado na página 28.
- LI, X.; ZHANG, Y.; LIAO, D. Mining key skeleton poses with latent svm for action recognition. *Applied Computational Intelligence and Soft Computing*, Hindawi, v. 2017, 2017. Citado na página 44.
- LISETTI, C. L. et al. A social informatics approach to human-robot interaction with a service social robot. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, IEEE, v. 34, n. 2, p. 195–209, 2004. Citado na página 24.
- LIU, R.; ZHANG, X. A review of methodologies for natural-language-facilitated human-robot cooperation. *International Journal of Advanced Robotic Systems*, SAGE Publications Sage UK: London, England, v. 16, n. 3, p. 1729881419851402, 2019. Citado na página 16.
- LOOS, H. M. Van der; REINKENSMEYER, D. J.; GUGLIELMELLI, E. Rehabilitation and health care robotics. In: *Springer handbook of robotics*. Nova Iorque, NY, USA: Springer, 2016. p. 1685–1728. Citado na página 17.
- MADA AT PORTAL. *Mada at Portal Guide Dog Robot*. 2019. Disponível em: <<https://madaportal.org/blogs/guide-dog-robot-future-mobility-aid-blind-and-visually-impaired>>. Acesso em: 25 dezembro de 2019. Citado na página 22.
- MARTINEZ-MARTIN, E.; CAZORLA, M. A socially assistive robot for elderly exercise promotion. *IEEE Access*, IEEE, v. 7, p. 75515–75529, 2019. Citado na página 20.
- MARTINS, F. N.; SARCINELLI-FILHO, M.; CARELLI, R. A velocity-based dynamic model and its properties for differential drive mobile robots. *Journal of Intelligent & Robotic Systems*, Springer, v. 85, n. 2, p. 277–292, 2017. Citado na página 56.
- MAURICE, P. et al. Human-oriented design of collaborative robots. *International Journal of Industrial Ergonomics*, Elsevier, v. 57, p. 88–102, 2017. Citado na página 14.

MCCOLL, D.; NEJAT, G. Meal-time with a socially assistive robot and older adults at a long-term care facility. *Journal of Human-Robot Interaction*, Journal of Human-Robot Interaction Steering Committee, v. 2, n. 1, p. 152–171, 2013. Citado na página 14.

MEDIUM. *Medium Voice Control*. 2019. Disponível em: <<https://medium.com/bolt-iot/controlling-bolt-devices-via-voice-commands-c7a0608b1ccb>>. Acesso em: 25 dezembro de 2019. Citado 2 vezes nas páginas 17 e 33.

MEYER, S.; FRICKE, C. Robotic companions in stroke therapy: A user study on the efficacy of assistive robotics among 30 patients in neurological rehabilitation. In: IEEE. *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. Lisboa, Portugal, 2017. p. 135–142. Citado na página 21.

MIES, G. Military robots of the present and the future. *Technology*, v. 9, n. 1, p. 125–137, 2010. Citado na página 13.

MITRA, S.; ACHARYA, T. Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, IEEE, v. 37, n. 3, p. 311–324, 2007. Citado na página 29.

MOESLUND, T. B.; HILTON, A.; KRÜGER, V. A survey of advances in vision-based human motion capture and analysis. *Computer vision and image understanding*, Elsevier, v. 104, n. 2-3, p. 90–126, 2006. Citado na página 29.

MONAJJEMI, V. M. et al. Hri in the sky: Creating and commanding teams of uavs with a vision-mediated gestural interface. In: IEEE. *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Tóquio, Japão, 2013. p. 617–623. Citado na página 14.

MONTEMERLO, M. et al. Experiences with a mobile robotic guide for the elderly. *AAAI/IAAI*, v. 2002, p. 587–592, 2002. Citado na página 16.

MÜLLER, M. Dynamic time warping. *Information retrieval for music and motion*, Springer, p. 69–84, 2007. Citado 3 vezes nas páginas 30, 31 e 32.

MURAKAMI, K.; TAGUCHI, H. Gesture recognition using recurrent neural networks. In: ACM. *Proceedings of the SIGCHI conference on Human factors in computing systems*. Kawasaki, Japão, 1991. p. 237–242. Citado na página 28.

MURPHY, R. R. et al. Search and rescue robotics. *Springer handbook of robotics*, Springer, p. 1151–1173, 2008. Citado na página 13.

NG, C. W.; RANGANATH, S. Real-time gesture recognition system and application. *Image and Vision computing*, Elsevier, v. 20, n. 13-14, p. 993–1007, 2002. Citado na página 30.

NONAKA, S. et al. Evaluation of human sense of security for coexisting robots using virtual reality. 1st report: evaluation of pick and place motion of humanoid robots. In: IEEE. *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*. New Orleans, LA, USA, 2004. v. 3, p. 2770–2775. Citado na página 37.

OLIVEIRA, G. C. de; CARVALHO, K. B. de; BRANDÃO, A. S. A hybrid strategy for robot navigation in semi-structured environments. In: IEEE. *2018 IEEE International Conference on Industrial Technology (ICIT)*. Lyon, França, 2018. p. 23–28. Citado na página 26.

OLIVEIRA, G. C. R. de; CARVALHO, K. B. de; BRANDÃO, A. S. A hybrid path-planning strategy for mobile robots with limited sensor capabilities. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 19, n. 5, p. 1049, 2019. Citado 2 vezes nas páginas 26 e 41.

Omron Adept. *Pioneer 3-DX*. Disponível em: <<http://www.mobilerobots.com/ResearchRobots/PioneerP3DX.aspx>>. Citado na página 43.

Omron Adept. *Pioneer 3-DX Data Sheet*. Disponível em: <<http://www.mobilerobots.com/Libraries/Downloads/Pioneer3DX-P3DX-RevA.sffb.ashx>>. Citado na página 43.

ORDÓÑEZ, F.; ROGGEN, D. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 16, n. 1, p. 115, 2016. Citado na página 30.

OSPENNIKOVA, E.; ERSHOV, M.; ILJIN, I. Educational robotics as an inovative educational technology. *Procedia-Social and Behavioral Sciences*, Elsevier, v. 214, p. 18–26, 2015. Citado na página 21.

PACCHIEROTTI, E.; CHRISTENSEN, H. I.; JENSFELT, P. Human-robot embodied interaction in hallway settings: a pilot user study. In: IEEE. *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005*. Nashville, TN, USA, 2005. p. 164–171. Citado 2 vezes nas páginas 37 e 38.

PARK, H. W.; HOWARD, A. Providing tablets as collaborative-task workspace for human-robot interaction. In: IEEE PRESS. *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*. Tóquio, Japão, 2013. p. 207–208. Citado na página 14.

PATRONA, F. et al. Motion analysis: Action detection, recognition and evaluation based on motion capture data. *Pattern Recognition*, Elsevier, v. 76, p. 612–622, 2018. Citado na página 29.

PHYS ORG. *Phys Org Humanoid Robot*. 2019. Disponível em: <<https://phys.org/news/2018-11-humanoid-robot-prototype-hrp-5p-capable.html>>. Acesso em: 25 dezembro de 2019. Citado na página 16.

RAHEJA, J. et al. Robust gesture recognition using kinect: A comparison between dtw and hmm. *Optik*, Elsevier, v. 126, n. 11-12, p. 1098–1104, 2015. Citado na página 29.

RAKITA, D.; MUTLU, B.; GLEICHER, M. A motion retargeting method for effective mimicry-based teleoperation of robot arms. In: ACM. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. Viena, Áustria, 2017. p. 361–370. Citado na página 14.

RANGANATHAN, A. The levenberg-marquardt algorithm. *Tutorial on LM algorithm*, v. 11, n. 1, p. 101–110, 2004. Citado na página 36.

RIOS-MARTINEZ, J.; SPALANZANI, A.; LAUGIER, C. From proxemics theory to socially-aware navigation: A survey. *International Journal of Social Robotics*, Springer, v. 7, n. 2, p. 137–153, 2015. Citado na página 37.

ROS. *Pioneer 3DX*. 2019. Disponível em: <<https://robots.ros.org/pioneer-3-dx/>>. Acesso em: 10 de janeiro de 2020. Citado na página 44.

- SEABOTIX. *Seabotix ROV Search and Rescue*. 2019. Disponível em: <<https://www.online-sciences.com/tag/seabotix-rov/>>. Acesso em: 25 dezembro de 2019. Citado na página 15.
- SEMPENA, S.; MAULIDEVI, N. U.; ARYAN, P. R. Human action recognition using dynamic time warping. In: IEEE. *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*. Bandung, Indonésia, 2011. p. 1–5. Citado 2 vezes nas páginas 25 e 29.
- TAHERI, A. et al. Teaching music to children with autism: a social robotics challenge. *Scientia Iranica*, v. 26, n. 1, p. 40–58, 2019. Citado na página 21.
- TAPUS, A.; MAJA, M.; SCASSELLATTI, B. The grand challenges in socially assistive robotics. *IEEE Robotics and Automation Magazine*, v. 14, n. 1, p. N–A, 2007. Citado na página 19.
- TECHBRIEFS. *Techbriefs Assistive Robotics*. 2019. Disponível em: <<https://www.techbriefs.com/component/content/article/tb/stories/blog/28820>>. Acesso em: 25 dezembro de 2019. Citado na página 16.
- TEKNOKULTURA. *TeKnoKultura Robots*. 2019. Disponível em: <<https://appleinsider.com/articles/19/10/08/meebot-20-educational-robotics-kit-ships-as-apple-store-exclusive>>. Acesso em: 25 dezembro de 2019. Citado na página 22.
- THRUN, S. et al. Minerva: A second-generation museum tour-guide robot. In: IEEE. *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C)*. Detroit, MI, USA, 1999. v. 3. Citado na página 37.
- TRIEBEL, R. et al. Spencer: A socially aware service robot for passenger guidance and help in busy airports. In: SPRINGER. *Field and service robotics*. Cham, Suíça, 2016. p. 607–622. Citado na página 37.
- TSAROUCHI, P.; MAKRIS, S.; CHRYSOLOURIS, G. Human–robot interaction review and challenges on task planning and programming. *International Journal of Computer Integrated Manufacturing*, Taylor & Francis, v. 29, n. 8, p. 916–931, 2016. Citado na página 13.
- VEERIAH, V.; ZHUANG, N.; QI, G.-J. Differential recurrent neural networks for action recognition. In: *Proceedings of the IEEE international conference on computer vision*. Santiago, Chile: [s.n.], 2015. p. 4041–4049. Citado na página 30.
- VERCELLI, A. et al. Robots in elderly care. *DigitCult-Scientific Journal on Digital Cultures*, v. 2, n. 2, p. 37–50, 2018. Citado na página 29.
- VINCIARELLI, A. et al. Social signal processing: state-of-the-art and future perspectives of an emerging domain. In: ACM. *Proceedings of the 16th ACM international conference on Multimedia*. Vancouver, Canada, 2008. p. 1061–1070. Citado na página 37.
- WANG, P. et al. Action recognition from depth maps using deep convolutional neural networks. *IEEE transactions on human-machine systems*, IEEE, v. 46, n. 4, p. 498–509, 2015. Citado na página 30.

- WANG, P. et al. Depth pooling based large-scale 3-d action recognition with convolutional neural networks. *IEEE Transactions on Multimedia*, IEEE, v. 20, n. 5, p. 1051–1061, 2018. Citado na página 30.
- WXYZ DETROIT. *Wxyz Detroit*. 2019. Disponível em: <<https://www.wxyz.com/news/national/sony-unleashes-the-cuteness-with-new-robot-dog>>. Acesso em: 25 dezembro de 2019. Citado na página 15.
- YAN, S.; XIONG, Y.; LIN, D. Spatial temporal graph convolutional networks for skeleton-based action recognition. In: *Thirty-Second AAAI Conference on Artificial Intelligence*. Nova Orleans, LA, USA: [s.n.], 2018. Citado na página 30.
- YANG, Y. et al. Ultrasonic robotic system for noncontact small object manipulation based on kinect gesture control. *International Journal of Advanced Robotic Systems*, SAGE Publications, London, England, v. 14, n. 6, p. 1729881417738739, 2017. Citado na página 29.
- YOSHIDA, K. Achievements in space robotics. *IEEE Robotics & Automation Magazine*, IEEE, v. 16, n. 4, p. 20–28, 2009. Citado na página 14.
- YOSHIDA, S. et al. Robojockey: Designing an entertainment experience with robots. *IEEE computer graphics and applications*, IEEE, v. 36, n. 1, p. 62–69, 2015. Citado na página 13.
- ZHANG, X.-H. et al. Improvement of dynamic hand gesture recognition based on hmm algorithm. In: *2016 International Conference on Information System and Artificial Intelligence (ISAI)*. Hong Kong, China: [s.n.], 2016. p. 401–406. Citado na página 29.
- ZHANG, Z. Microsoft kinect sensor and its effect. *IEEE multimedia*, IEEE, v. 19, n. 2, p. 4–10, 2012. Citado na página 42.