Reconhecimento de movimentos humanos para
imitação e controle de um robô humanoide

Fernando Zuher Mohamad Said Cavalcante

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito: 15/10/2012

Assinatura: _____

Reconhecimento de movimentos humanos para imitação e controle de um robô humanoide

Fernando Zuher Mohamad Said Cavalcante

***Orientadora:* Profa. Dra. Roseli Aparecida Francelin Romero**

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

**USP – São Carlos
Outubro de 2012**

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

C376r Cavalcante, Fernando Zuher Mohamad Said
Reconhecimento de movimentos humanos para imitação
e controle de um robô humanoide / Fernando Zuher
Mohamad Said Cavalcante; orientadora Roseli
Aparecida Francelin Romero. -- São Carlos, 2012.
63 p.

Dissertação (Mestrado - Programa de Pós-Graduação em
Ciências de Computação e Matemática Computacional) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2012.

1. Robótica. 2. Imitação. 3. Teleoperação. 4.
Humanoide. 5. Kinect. I. Romero, Roseli Aparecida
Francelin, orient. II. Título.

Agradecimentos

Em mais uma fase acadêmica concluída, há pessoas que são dignas de agradecimentos por participarem de uma boa forma na minha vida.

Não como uma interjeição, mas com sinceridade ao dizer “Graças a Deus!” é que realizo mais uma etapa em minha carreira como pesquisador. Sou grato ao Senhor por estar em uma das melhores universidades da América Latina, pela adequação dos meus interesses acadêmicos quanto às oportunidades presentes e por todo e qualquer meio que Ele provê para que eu possa viver. Este fim de mestrado representa uma de Suas provisões.

Agora se tratando de “pessoas”. Este trabalho de mestrado pôde ser realizado por meio do suporte que meus pais me forneceram. Não acredito que viver de bolsa e morando sozinho seja o suficiente pra se viver bem em São Carlos. Também, o cuidado dos meus pais quanto à minha educação desde anos atrás ainda na escola, depois na graduação e agora no término desse mestrado, foi muito importante para mim.

À minha orientadora professora Roseli Romero. Conheci a professora ainda no final da minha graduação quando passando pelo ICMC durante duas semanas. Robótica já era meu interesse de pesquisa, mas o ramo de humanoides foi um das portas que me surgiram para poder satisfazer um interesse pessoal por dentro da robótica. A professora me proporcionou a oportunidade para trabalhar nessa área, ainda mais na responsabilidade de ser um dos primeiros brasileiros a trabalhar com uma arquitetura robótica cara e complexa. Pela confiança e orientação fica minha gratidão.

Algumas amizades formadas no Laboratório de Aprendizado de Robôs do qual faço parte e também do Biocom foram importantes, já que não acredito ser saudavelmente possível a realização de um mestrado sem boas amizades. Dessa forma, meus agradecimentos, em especial, a boas amizades que adquiri: André Rossi, Bruno Feres, Ivani Negrão, Marcelo Albertini, Renato Ramos e Valéria Carvalho.

À USP São Carlos – e especialmente ao ICMC – pela estrutura e qualidade do programa de Ciências de Computação.

Não posso deixar de lembrar algumas pessoas as quais tem sido bênçãos na minha vida em São Carlos, por meio da companhia e lembranças de oração a mim. Assim, agradeço ao casal Ana e Alexandre Albuquerque e ao pastor Jarbas Valentim.

Por fim, ao CNPQ e FAPESP pelo parcial suporte financeiro.

Muito obrigado a todos!

Resumo

Em interações humano-robô ainda existem muitas limitações a serem superadas referentes à provisão de uma comunicação natural quanto aos sentidos humanos. A capacidade de interagir com os seres humanos de maneira natural em contextos sociais (pelo uso da fala, gestos, expressões faciais, movimentos do corpo) é um ponto fundamental para garantir a aceitação de robôs em uma sociedade de pessoas não especialistas em manipulação de engenhos robóticos. Outrossim, a maioria dos robôs existentes possui habilidades limitadas de percepção, cognição e comportamento em comparação com seres humanos. Nesse contexto, este projeto de pesquisa investigou o potencial da arquitetura robótica do humanoide NAO, no tocante à capacidade de realizar interações com seres humanos através de imitação de movimentos do corpo de uma pessoa e pelo controle do robô. Quanto a sensores, foi utilizado um sensor câmera não-intrusivo de profundidade incorporado no dispositivo Kinect. Quanto às técnicas, alguns conceitos matemáticos foram abordados para abstração das configurações espaciais de algumas junções/membros do corpo humano – essas configurações foram capturadas por meio da utilização da biblioteca OpenNI. Os experimentos realizados versaram sobre a imitação e o controle do robô por meio da avaliação de vários usuários. Os resultados desses experimentos revelaram um desempenho satisfatório quanto ao sistema desenvolvido.

Abstract

In human-robot interactions there are still many restrictions to overcome regarding the provision of a communication as natural to the human senses. The ability to interact with humans in a natural way in social contexts (the use of speech, gestures, facial expressions, body movements) is a key point to ensure the acceptance of robots in a society of people not specialized in manipulation of robotic devices. Moreover, most existing robots have limited abilities of perception, cognition and behavior in comparison with humans. In this context, this research project investigated the potential of the robotic architecture of the NAO humanoid robot, in terms of ability to perform interactions with humans through imitation of body movements of a person and the robot control. As for sensors, we used a non-intrusive sensor depth-camera built into the device Kinect. As to techniques, some mathematical concepts were discussed for abstraction of the spatial configurations of some joints/members of the human body – these configurations were captured through the use of the OpenNI library. The performed experiments were about imitation and the control of the robot through the evaluation of various users. The results of these experiments showed a satisfactory performance for the developed system.

Sumário

Capítulo 1 – Introdução	13
1.1 Robô	13
1.2 Interação Humano-Robô	16
1.3 Motivação	17
1.4 Objetivos.....	18
1.5 Contribuições	18
1.6 Organização do Trabalho	18
Capítulo 2 – Revisão Bibliográfica.....	19
2.1 Sensoriamento Humano.....	19
2.1.1 Detecção/Rastreamento	22
2.1.2 Entendimento de comportamento humano.....	22
2.2 Teleoperação	22
2.3 Interação com robôs humanoides	24
2.4 Considerações finais	29
Capítulo 3 – Métodos de Pesquisa.....	31
3.1 Fundamentos matemáticos.....	31
3.1.1 Espaço de coordenadas e orientação de objetos	31
3.1.2 Funções matemáticas utilizadas na abstração das configurações	33
3.2 Microsoft Kinect.....	34
3.3 Framework OpenNI	36
3.4 Robô NAO	37
3.5 Articulações	38
3.5.1 Medições do robô e do usuário.....	38
3.5.2 Regras mapeadas para junções	40
3.6 Funcionalidades do sistema.....	43
3.7 Outras observações.....	45
3.8 Considerações Finais	45

Capítulo 4 – Experimentos Realizados	47
4.1 Avaliação subjetiva.....	47
4.2 Considerações Finais	50
Capítulo 5 – Discussão dos Resultados	51
5.1 Framework OpenNI	51
5.1.1 Limitações na matriz de rotação.....	51
5.1.2 Yaw do ombro.....	52
5.1.3 Rastreamento dos membros inferiores	52
5.2 Robô NAO	52
5.2.1 Módulo de balanço e omissão de auto-colisão do NAO	52
5.2.2 Dissimilar corpo do robô para uma pessoa	52
5.3 Imitação das pernas.....	52
5.4 Principais dificuldades.....	54
5.5 Considerações Finais	55
Capítulo 6 – Conclusão e Trabalhos Futuros.....	57
Referências Bibliográficas	59

Capítulo 1

Introdução

Desde a revolução industrial tem-se realizado com notoriedade a automação de tarefas repetitivas, entediantes, complexas ou impossíveis para serem realizadas por seres humanos. Atividades como lavar louças/roupas, aquecer alimentos/ambientes, lavar carros, preparar café, fazer chamadas por telefone sem a necessidade da intervenção de telefonistas, e muitas outras tem sido realizadas por meio de máquinas que não requerem algum tipo de inteligência criativa – grosseiramente referindo-se a apenas interpretadores de regras *se, então*.

Aqui, referimo-nos ao termo inteligência como um que envolva um mínimo de criatividade ou um entendimento/ação que não esteja relacionado a qualquer aspecto previamente definido por alguma entidade externa à máquina.

Da escravidão humana passando pela automação de atividades por meio de máquinas até sua execução por robôs, parece ser essa a linha de evolução de como o ser humano realiza algumas de suas atividades braçais.

1.1 Robô

A primeira menção do neologismo “robô” é encontrada no trabalho do escritor tcheco Karel Čapek quando este o utilizou em sua peça R.U.R. (*Rossum’s Universal Robots*) em 1921 [1]. Na literatura há divergências quanto ao criador do termo, se Karel ou seu irmão Josef Čapek. Entretanto, o termo parece ter sido criado em uma conversa informal com seu irmão no verão de 1920 [2].

Não há consenso na literatura do que seja um robô. O dicionário Houaiss [3] informa que um robô é uma “máquina, autômato de aspecto humano, capaz de se movimentar e de agir”. Segundo Russel e Norvig [4] “agentes físicos que executam tarefas manipulando o mundo físico”. Čapek [1] *apud* Thrun *et al.* [5] define como “máquinas dispostas, inteligentes e com aparência humana que tornam a vida agradável por fazer o tipo de trabalho que não gostamos de fazer”.

Dada à evolução da pesquisa em robótica, temos visto uma vulgar nomeação de máquinas/autômatos como robôs. Algumas destas indevidas nomeações são feitas a autômatos com funções muito reduzidas tais como: softwares automatizados para coleta de informações em sistemas bancários, programas de coleta e análise de informações atuais na Internet, etc.

Esses são uns poucos exemplos da confusão na utilização desse termo. Possivelmente o uso de automação em algumas tarefas que até então eram desempenhados por seres humanos, pode ser a razão do mau uso desse termo.

Em nossas palavras um robô é uma máquina composta de software/hardware que desempenha automaticamente, através da percepção do ambiente, tarefas consideradas complexas. Complexas não necessariamente intrínsecas a natureza da tarefa, mas sim pelo fato de serem realizadas por entidades que não são seres humanos, os quais são hábeis para a realização de tais (como comunicar, caminhar, reconhecer padrões, etc).

Criaturas artificiais, com características e/ou comportamentos análogos a algum ser vivente, têm estado presentes na filosofia, mitologia e religião, há séculos. De acordo com uma ordem aproximadamente cronológica [4, 6], deparamo-nos com: um ser antropomórfico gigante feito de bronze, chamado Talos, criado para proteger a ilha de Creta pelo mitológico deus grego Hefesto [7]; um ser antropomórfico que se comportava como um humano e até piscava para concubinas [8] *apud* [6]; o *golem* do folclore judeu [9]; as empregadas de ouro da obra *Ilíada* [10]; um pássaro pica-pica feito de madeira/bamboo que podia ficar suspenso no ar por até três dias [11] *apud* [6]; um boi de madeira e cavalo flutuante [6]; um esboço mecânico antropomórfico de Leonardo da Vinci [12]; pássaros mecânicos no poema *Byzantium* [13]; robôs na literatura de ficção científica de Asimov [14]. Figura 1 demonstra o período aproximado do surgimento de algumas dessas criaturas na literatura.

Ao aborda-se a ciência robótica depara-se com a falta de consenso quanto à sua origem. Focando em autômatos tangíveis há trabalhos – seguindo uma ordem aproximadamente cronológica de eventos [4, 6] – um pato mecânico construído por Vaucanson (1738) [4]; um bote mecânico teleoperado por rádio (1898) [15]; um cachorro elétrico (1923) [6]; algumas tentativas durante a Segunda Guerra Mundial de controlar bombardeadores remotamente [6]; manipuladores teleoperados para tratar materiais nucleares [16]; uma máquina mecânica (1948) aludida a uma tartaruga – este robô era capaz de se movimentar autonomamente por

Figura 1: Alguns conceitos de criaturas artificiais encontrados na literatura.

uma sala [4]; um robô nomeado *Hopkins Beast* era capaz de reconhecer placas de cobertura de tomadas de energia elétrica, conectar-se a elas e então recarregar a si próprio (início da década de 60) [4]; um braço-robô chamado *Unimate* vendido a General Motors em 1961 para ser utilizado na produção de tubos de imagem de TV [4]; o robô *Shakey* [17], desenvolvido na antiga *Stanford Research Institute*, é comumente referenciado como o primeiro robô móvel autônomo dotado de técnicas de Inteligência Artificial (IA) – também, pode-se dizer que este trabalho foi uma pedra angular no consequente desenvolvimento da IA; em 1972 a Nissan Corporation automatizou toda uma linha de montagem com robôs [4]; o desenvolvimento do robô *PUMA* em 1978, permitiu que este robô se tornasse uma base na manipulação robótica por até cerca do ano 2000 [4]; o *Stanford Cart* e o *CMU Rover* são outros avanços ocorridos em torno da década de 80 [18], etc. Na Figura 2 é exibida uma amostra da evolução dessas implementações.

Os primeiros robôs criados parecem emergir do campo de pesquisa de teleoperação onde robôs com nenhuma ou pouca autonomia eram controlados explicitamente por comandos dados por humanos [16]. Um aparte aqui: Embora inteligência autônoma possa implicitamente remeter uma característica determinante de um robô – podendo assim haver certo preconceito quanto à pesquisa em teleoperação, pois a inteligência autônoma parte do usuário e não do robô –, Jean Vertut demonstrou uma utilização mais complexa da aplicação de um manipulador robótico por meio de teleoperação [19, 20].

Em uma escala de realização de atividades muito menor comparadas a máquinas autômatos, robôs têm sido aperfeiçoados para fazer o que esperamos que eles façam: tarefas tediosas, sujas, complexas ou perigosas para seres humanos. A utilização de robôs em algumas áreas tais como salas cirúrgicas, viagens espaciais, ambientes perigosos como desativação de bombas ou averiguação de ambientes instáveis, é em grande maioria realizada por algum nível de teleoperação em vez de um robô inteligente completamente autônomo.

Em uma tentativa de categorizar a maior parte dos robôs, de acordo com Russel e Norvig [4], existem três principais distinções: braços robóticos (ou manipuladores), robôs móveis e humanoides. Na Figura 3 podemos ver exemplos destes robôs.

Braços robóticos estão limitados a um ponto fixo pelo qual se movimentam (por exemplo, manipuladores em linhas de montagem de automóveis). Robôs móveis são equipados com rodas, pernas ou variantes de locomoção (por exemplo, robôs de corpo poligonal com

Figura 2: Algumas criações robóticas ou que se assemelham a robôs.

Figura 3: Da esquerda à direita: 1 - Braços robóticos em uma linha de montagem de automóveis. 2 - Robô móvel Pioneer. 3 - Robô humanoide NAO.

duas rodas em sua base). Humanoides, como a própria etimologia denuncia, *humano*, apresentam uma estrutura corporal semelhante a dos humanos.

1.2 Interação Humano-Robô

Desde que *interação* lembre algum nível de relação entre pelo menos duas entidades, aborda-se aqui o campo de estudo científico denominado Interação Humano-Robô (IHR), o qual conforme Goodrich e Schultz [6] “é um campo de estudo dedicado à compreensão, elaboração e avaliação de sistemas robóticos pelo uso por ou com seres humanos”.

Esse campo relativamente novo tem em sua base a multidisciplinariedade como regra na qual se encontram pesquisadores oriundos dos campos de Ciência Cognitiva, Ciência de Computação, Interação Humano-Computador, Fatores Humanos, Psicologia, entre outros. Esses pesquisadores tem buscado a realização de uma interação eficaz e eficiente (esta se referindo ao êxito da interação em si e aquela ao meio como é desempenhada) entre seres humanos e robôs ao longo da evolução em robótica.

A IHR pode ocorrer por meio de teleoperação, robôs assistentes, interação física, interação social, etc. Também, a IHR pode ser abordada em dois tipos de interação: Remota e Próxima [6].

A Interação Remota refere-se à comunicação entre um robô e um ser humano em espaço ou até em tempos diferentes. Por exemplo, o robô enviado a exploração da superfície do planeta Marte se comunica com o ser humano na Terra fora do mesmo espaço e tempo (o tempo que se leva para a comunicação de ambos os agentes é em torno de 14 minutos devido à distância entre os mesmos). Na Interação Próxima, ambos os agentes se comunicam no mesmo espaço/tempo, por exemplo, robô e humano interagindo num mesmo local (físico) ao mesmo tempo [6].

Quanto ao ramo de teleoperação, este é comumente referenciado a interações remotas, por meio das quais um robô móvel é controlado a distância por um humano [6]. Entretanto, outros robôs, além dos móveis, podem também ser teleoperados.

Quanto aos sensores comumente utilizados para abstração do controle humano em função do robô, estes requerem contato físico entre alguma parte do corpo do agente controlador – no caso, o ser humano. Neste trabalho referimo-nos a estes tipos de sensores como intrusivos, pois eles são como objetos estranhos a morfologia humana.

Com a evolução da área de Sensoriamento Humano tem ocorrido o desenvolvimento de abordagens que abstraem o estado atual de um ambiente por meio de sensores não-intrusivos

ao objeto em estudo. A utilização de câmeras visuais/termais, sensores lasers, entre outros, em abordagens não-intrusivas, tem permitido essa abstração de uma forma muito mais cômoda e agradável para o agente em estudo.

Um termo nomeado Interface Natural refere-se, em poucas palavras, à utilização de meios antropomórficos natos de um ser humano como entrada a uma interface para interação com algum outro par. Por exemplo, se alguém deseja comunicar um estado emotivo ou uma ordem a um robô, esse alguém está restrito a utilização de seu corpo para expressão dessas informações. Assim, a expressão facial e a fala da pessoa poderiam ser uns dos meios utilizados já que esses são canais de passagem de informação concernentes a antropomorfia humana – diferente de um controle remoto na mão do usuário.

Embora sensores intrusivos, em geral, meçam mais acuradamente a postura do corpo humano com relação às medições providas por sensores não-intrusivos, tem ocorrido grandes avanços nesta área ao longo da pesquisa de análise de movimento humano com aplicações no contexto de robótica. Restringindo-se a utilização desses sensores no campo de teleoperação de robôs humanoides: há consideráveis limitações devido aos desafiantes problemas em estimar a configuração de todo o corpo de um ser humano, e pela complexidade em materializar robôs próximos o suficiente à morfologia humana.

Um sensor não-intrusivo acessível e revolucionário lançado pela Microsoft Corporation no final de 2010 para o videogame Xbox, nomeado Kinect, tem sido extensamente utilizado em plataformas computacionais nos mais criativos meios possíveis. A utilização desse dispositivo, com bibliotecas apropriadas, tem tornado possível a abstração da configuração de algumas junções do corpo do ser humano – abstração a qual alcança a supracitada Interface Natural – podendo ser utilizados para teleoperar robôs.

O atual estado de arte da robótica não corresponde o que se espera da ideia original de Čarel – um ser disposto. Embora ainda em fase de concepção dessa ideia, pode-se tomar vantagens do ramo de teleoperação, substituindo o fator inteligência, em forma artificial, por fazer uso de nossa inteligência para controlar robôs. Em um requerimento muito menor de complexidade o ramo de teleoperação tem permitido o uso de robôs em atividades que um dia espera-se que eles realizem autonomamente.

Assim, pode-se inferir que esta área há de ser ainda muito utilizada pelo fato de ser altamente complexa a criação de dispositivos dotados de inteligência artificialmente materializada por humanos.

1.3 Motivação

Embora as pesquisas sobre robôs estejam de fato evoluindo, ainda existem diversas limitações a serem superadas. A maioria dos robôs existentes possui canais limitados de percepção, cognição e ações, comparados a seres humanos. A noção de inteligência autônoma é uma questão complexa para ser implementada em máquinas.

Não é esperado que pessoas, em um contexto social, interajam com robôs através da incorporação/manipulação de dispositivos computacionais de complexidade característica à área. Dessa forma, a busca de uma interação mais natural entre humanos e robôs é fundamental para aceitação destes em uma sociedade de não especialistas em tecnologia e não familiarizadas com sua manipulação.

Portanto, prover meios naturais de comunicação entre esses agentes é oportuno e desejável.

1.4 Objetivos

Este trabalho tem por objetivo geral a teleoperação de um robô humanoide por meio de um sensor câmera não-intrusivo. Por não-intrusivo entende-se que as informações abstraídas do agente controlador (o ser humano) não são oriundas de qualquer marco/sensor/objeto sobre o corpo do agente, salvo o próprio corpo do usuário.

A teleoperação de um robô humanoide deverá torná-lo capaz de executar as seguintes tarefas: (1) a imitação de movimento de alguns membros do corpo humano; (2) e o reconhecimento de alguns pré-determinados movimentos do corpo humano para provocar a execução, por meio do robô, de alguns comportamentos pré-definidos como sentar/levantar, caminhar, e a realização de alguns gestos.

Quanto ao robô e ao sensor utilizado, estes são o humanoide NAO e o sensor Kinect. Na literatura existem trabalhos que demonstram a utilização do robô de forma satisfatória em campos de pesquisa acadêmica [21-23].

Também, será utilizado o sensor Kinect operando em conjunto com o robô humanoide NAO.

Como resultado, depreende-se que uma interação próxima e social entre um humano e um humanoide no tocante as duas tarefas citadas deveria ser realizada por meio de interfaces naturais.

1.5 Contribuições

A semelhança deste trabalho com relação aos abordados na literatura abrange a realização da imitação e controle de um robô humanoide. Quanto à diferença, é apresentada uma discussão própria sobre os interessantes achados gerados pelo desenvolvimento deste projeto, visto que não é encontrado na literatura um volume substancial de trabalhos expondo alguma considerável discussão sobre imitação e controle de robôs humanoides através de controle direto – isto é, sem qualquer tipo de abordagem de aprendizado.

1.6 Organização do Trabalho

Este trabalho está organizado do seguinte modo. No Capítulo 2, é apresentada uma revisão bibliográfica referente aos campos de pesquisa próximos a esse projeto. No Capítulo 3 são versados os métodos de pesquisa que suportaram o objetivo desse trabalho e as informações referentes ao sistema desenvolvido. No Capítulo 4, experimentos sobre o sistema desenvolvido são abordados – seguidos de discussão e limitações concernentes a este projeto no capítulo 5. Por fim, no capítulo 6, são apresentados as conclusões e os trabalhos futuros.

Capítulo 2

Revisão Bibliográfica

Na interação próxima, inicialmente, ambos os agentes devem se perceber. Neste projeto tratamos o caso da percepção de um robô, não requerendo o tratamento da percepção do ser humano na comunicação. Desta forma, abordamos algumas áreas de pesquisa quanto ao Sensoriamento Humano que permite entender as propriedades de um ser humano em um dado contexto. Uma vez conhecidas essas propriedades, explanamos a área de Teleoperação por meio de uma breve narração de sua evolução, suas aplicações e a correspondência que há com o que esse trabalho propõe. Por fim, trabalhos semelhantes ao que este propõe quanto à imitação de movimentos humanos e controle de robôs humanoides são abordados.

2.1 Sensoriamento Humano

O grande volume de trabalhos e aplicações de sucesso em Análise de Movimento Humano baseado em Visão disponibilizaram vários *surveys* na área, os quais classificam os trabalhos em taxonomias distintas baseadas em focos diferentes.

Destas várias revisões publicadas até a presente data (vide [24, 25]), escolhemos a taxonomia de Wang *et al.* [26] que se assemelha ao trabalho proposto aqui. Os autores categorizam os trabalhos desta área em Detecção, Rastreamento e Entendimento de Comportamento Humano.

Detecção refere-se à abstração de um objeto em uma imagem (quadro ou *frame*). Rastreamento, no mapeamento desse objeto nos *frames* subsequentes à (primeira) detecção. Entendimento de comportamento humano, grosso modo, aborda a classificação de um comportamento característico a um objeto, que no nosso caso é um ser humano.

Os problemas que surgem na localização de uma pessoa por meio de câmeras são vários, tais como: posição relativa da pessoa na visão da câmera (frente, perfil, costas, sentado, deitado, inclinado), posição da câmera em relação ao ambiente, tamanho/roupa das pessoas, luminosidade do ambiente, fundo (*background*), perda temporária de localização da pessoa,

ruído de sensores, velocidade do movimento do agente, objetos estáticos ou em movimento (como carro, animais, outras pessoas), movimento da câmera/pessoa. Exemplos são demonstrados na Figura 4.

Segundo Moeslund e Granum [27] há três principais aplicações na área de Análise de Movimento Humano sendo vigilância, controle e análise. O trabalho aqui abordado refere-se à área de controle na qual uma particular detecção de movimento humano¹ é utilizada para controle de algo – uma arquitetura robótica.

Consoante Teixeira *et al.* [28] pode-se classificar a maior parte das propriedades humanas passíveis de mensuração em três: espaço-temporais, comportamentais e fisiológicas.

A propriedade espaço-temporal é definida pelas características como presença, quantidade, localização, rastreamento e identidade de pessoas. Nas comportamentais, características como pose, ação, atividade, comportamento individual, comportamento em grupo, são consideradas. Por fim, as fisiológicas revelam-se por meio de informações quanto à frequência cardíaca, temperatura, peso, cor de pele, etc.

Algumas propriedades fisiológicas do ser humano são sempre passíveis de observação, independente da ação que este possa realizar – nomeadas traços intrínsecos estáticos conforme [28]. Algumas são movimentos dos batimentos cardíacos ou dos pulmões, diferença de temperatura do corpo com o ambiente, odor corporal, etc.

Outros traços intrínsecos, mas dinâmicos, se revelam quando o ser humano está em atividade. Alguns exemplos: se uma pessoa está em movimento, então pode-se detectar o movimento dos respectivos membros; se caminhando, a percepção da pressão, vibração, ou som produzido pelos pés ao se caminhar; se falando, a detecção de voz [28].

Quanto aos meios para abstração desses dados, estes variam largamente dependendo dos tipos de informações produzidas e dos problemas tratados, tais como: sensores infravermelhos passivos, pisos sensíveis à pressão, detecção de campos elétricos, sensores de

Figura 4: Imagens representando várias posições de pessoas/câmeras, ambientes e luminosidade. [29]

¹ Neste capítulo, os termos *usuário*, *pessoa*, *humano* e *controlador* são utilizados indistintamente referindo-se ao ser humano que interage com o robô, ambos agentes da comunicação. Palavras como *segmentar*, *detectar* e *localizar* com o sentido de abstrair um ser humano dos dados percebidos por um robô. Também, algumas partes do corpo humano como cabeça, pescoço, mão e pé são referenciadas como *junções*; como o ombro, cotovelo, quadril e joelho os são.

vibração, sensores laser/ultrassom/radio, câmeras, sensores inerciais, microfones, etc [28].

Outrossim, o sensoriamento humano apresenta diversos desafios a serem tratados, tais como [28]: ruído de sensores; variações ambientais não esperadas (chuva, neblina, variações de temperatura) e similaridades com o *background*; movimento de objetos como carros, cadeiras, outras pessoas; variabilidade na aparência e imprevisibilidade; fraude ativa.

O sistema ou o meio pelo qual se obtém a localização de um ser humano provém de sensores e do respectivo processamento dos dados destes [27]. Pode-se classificar os sensores como instrumentados e não-instrumentados referentes à utilização ou não de sensores no corpo do usuário, respectivamente [28].

Abordagens não-instrumentadas manifestam-se por meio de sensores binários, sensores de vibração, *Ranger-Finders*, *Doppler-shift*, câmeras (e outros imageadores), etc [28].

Sensores binários, basicamente, retornam um lógico 1 se alguma pessoa é detectada num ambiente, e 0 caso contrário. Destes, há *PIRs* (*Passive infrared*), sensores de contato, sensores de campo elétrico, etc. Sensores de vibração medem os sinais produzidos pelos passos de uma pessoa – em contextos maiores, estes podem ser encontrados em sensores geossísmicos. Varredura de *Range-Finders* transmitem um sinal e verificam o eco desse sinal com relação ao tempo ou energia. Câmeras e outros imageadores permitem a abstração de uma quantidade considerável de informações concernentes a uma pessoa (como movimento, posição, comportamento, aparência, tipos de objetos presentes no ambiente, etc) ao custo de alta dimensionalidade comparada a outras modalidades de sensores [28].

Em abordagens instrumentadas a detecção do usuário se faz claramente dada à presença de sensores no corpo de uma pessoa. Esta modalidade pode-se traduzir por meio de dispositivos operando em conjunto com uma rede externa, por exemplo, o Sistema de Posicionamento Global (GPS) utilizado através de telefones móveis com a pessoa. Sinais Wi-Fi podem ser utilizados para estimar a localização de assinaturas de força (*strength signature*) de sinal em cômodos distintos. Sinais de comunicação de telefonia móvel também podem ser trabalhados para estimação de localização de pessoa.

Sensores inerciais no corpo de uma pessoa estimam a trajetória de movimento do corpo com relação a sua velocidade e aceleração por meio de acelerômetros e giroscópios.

Campos magnéticos podem ser trabalhados por meio da emissão artificial de correntes DC em três eixos. Receptores são posicionados em algumas regiões do corpo humano e medem os campos magnéticos caracterizados nos três eixos.

Potenciômetros eletromecânicos podem ser utilizados diretamente na estimação da orientação de algumas junções do corpo humano ao serem posicionadas no corpo de uma pessoa.

Também, a intensidade de pulsos acústicos emitidas por transmissores ultrassônicos podem ser utilizadas para estimação de posição do corpo de uma pessoa por meio de microfones posicionados em locais específicos do usuário.

Os sensores, de modo geral, podem ser caracterizados como passivos ou ativos.

Sensores ativos são dispositivos artificiais presentes no ambiente que auxiliam no problema do rastreamento [27]. Estes podem estar acoplados no usuário e/ou ambiente, como sensores de detecção de movimento, variação de temperatura, tátil, etc. Esta abordagem implicitamente denota na emissão e recepção de sinais artificiais.

Em uma abordagem com sensores passivos utilizam-se fontes naturais como a própria luz do ambiente (natural ou elétrica), ondas infravermelhas de calor humano, etc. Esta

abordagem permite uma interação mais próxima ao contexto de comunicação de humanos, livre de auxiliares. Uma exceção a esta abordagem é que a presença de alguns auxiliares como marcos de papéis, bolas, vestimentas de cor homogênea, não deixam de serem sensores passivos, visto que os mesmos não emitem e nem recebem (processam) sinal artificial algum, diferente da abordagem ativa [27].

2.1.1 Detecção/Rastreamento

Para qualquer interação social natural com um robô, primeiramente ambos os agentes da comunicação devem se perceber. O ser humano tem vários sentidos para isso. Já o robô, independente dos sensores que venha utilizar deve saber o que é um ser humano através da abstração de características que possam descreve-lo.

Dada uma imagem – no contexto de sensores câmeras – deve-se detectar onde se localiza um ser humano. Depois de identificado em um *frame*, resta fazer o mesmo para os quadros subsequentes, ou seja, rastrear a pessoa.

Nisso surge uma pergunta: Quais características devem ser consideradas na detecção de um ser humano em uma imagem? No contexto de Reconhecimento de Formas, Costa e Cesar [30] informam que “a escolha de características é particularmente crítica, desde que isso pode impactar grandemente o resultado final de classificação”. Além da extração de características pertinentes ao problema a ser tratado, ainda há o meio pelo qual se dará a classificação a qual tem sido menos crítica que o processo de selecionar características pertinentes ao problema, conforme [30, 31].

Não há consenso das características (e também das técnicas) utilizadas na detecção de pessoas. Algumas abordagens encontradas na literatura são baseadas em cor de pele [32-34], diferenciação de mapa de ambiente por presença de usuários [35], calor corporal [36], etc.

2.1.2 Entendimento de comportamento humano

Quanto ao entendimento de comportamento humano, este refere-se, grosso modo, a como associar determinados comportamentos do ser humano com padrões de comportamento/ações pré-estabelecidos em um sistema. Algumas técnicas que são utilizadas na aprendizagem/reconhecimento de comportamentos (ações) realizadas por pessoas são *Dynamic Time Warping* [37, 38], Modelos Ocultos de Markov [39-43], Redes Neurais Artificiais [44, 45], etc.

2.2 Teleoperação

Embora não haja um consenso claro entre os pesquisadores sobre o começo da ciência Robótica, parece que a mesma teve seu início no desenvolvimento da pesquisa de teleoperação de manipuladores durante e depois da II Guerra Mundial para tratamento de material nuclear como demonstra os trabalhos de Goertz [16, 20, 46-49].

O prefixo *tele* significa *distância*, *afastamento*, *longe*. Assim, *teleoperação* refere-se a uma operação à distância. No contexto de robótica: a manipulação à distância de robôs que não

envolva contato direto entre o operador humano (denominado *master*) e o robô (denominado *slave*).

A teleoperação pode ser, grosso modo, classificada em teleoperação direta, teleoperação coordenada e controle supervísório [50]. O primeiro refere-se ao controle direto dos atuadores do robô com efeitos no ambiente em (ou quase) tempo-real. A teleoperação coordenada assemelha-se a direta, exceto pela inclusão de um tipo de controle interno de *loop* no robô com a função de tratar os atrasos (*delays*) do operador humano. No controle supervísório o robô possui certo nível de autonomia em sua execução tendo o operador humano um papel de monitor e de instruidor de ordens mais gerais.

Além do tratamento de materiais radioativos, outras áreas de pesquisa têm tomado vantagens da utilização de robôs teleoperados, tais como: exploração de profundo mar [51-54], exploração espacial [55-60], áreas militares [61], vigilância [62], telecirurgia [63], etc.

Segundo Lichiardopol [50], o desenvolvimento de um sistema de teleoperação requer a formação de três aspectos:

- Sensores e atuadores: Estabelecimento de correspondência de sensores e atuadores entre o robô e o operador. Alguns exemplos são vestimentas mecânicas, *head-mounted displays*, *joysticks*, etc.
- Meio de comunicação: Meios a fio, sem fio, ondas a rádio, infravermelho, e também protocolos de comunicação devem estabelecidos.
- Tempo de atraso: Atraso no tempo considerado entre a emissão de um comando pelo operador até a execução de ação correspondente pelo robô.

Os fatores estabilidade e telepresença podem ser designados como dois principais objetivos da teleoperação [64]. O primeiro refere-se à independência do sistema *closed-loop* quanto à conduta do operador humano ou do correr do ambiente – mantendo-se estabilidade. O segundo, quanto a prover um senso real de presença do ambiente remoto para o operador.

Seguindo uma linha de tempo conforme [64] na análise e desenvolvimento de sistemas teleoperados, um tipo de comportamento nomeado *move-and-wait* era observado nos experimentos de Sheridan e Ferrell (década de 60) [65, 66] sempre que um sistema teleoperado era submetido a atrasos no *loop* de comunicação do sistema. Esses experimentos objetivavam a mensuração do tempo requerido para a realização de determinadas tarefas sob o viés dos atrasos. Esses ensaios revelaram a direta relação linear entre os atrasos induzidos no circuito e o tempo de execução das tarefas submetidas aos operadores, e uma satisfatória execução de movimentos realizados pelos operadores ante a estratégia *move-and-wait*.

A fim de superar o tempo de execução de teleoperação proporcional ao atraso no *loop* de comunicação da estratégia anterior, Ferrell e Sheridan (1967) [67] estabeleceram que os comandos executados pelos operadores fossem tratados de forma supervísória, caracterizando-os sob natureza análoga ou simbólica. Desta forma, o mecanismo teleoperado deveria ser desenvolvido com certa autonomia para execução das pequenas tarefas compreendidas pelos comandos supervísórios. Após, Whitney em 1969 [68] apresentou a abordagem supervísória com enfoque na otimização das tarefas realizadas pelo sistema. Dessa forma, pode-se estimar o meio pelo qual se obteria desempenho ótimo na realização de tarefas durante a teleoperação.

Madni *et al.* (1983) [52] trataram de uma especial linguagem de programação voltada

ao desenvolvimento de controle supervisorio com algumas características de telepresença, tais como visual e força.

Teleoperação baseada em passividade remete a Teoria de Passividade (*Passivity Theory*) [69] abordando a troca de energia entre sistemas interconectados (ditos dinâmicos). Esta conduz ao estabelecimento de um modelo de sistema estável e sem perdas. Ao se garantir passividade no sistema de teleoperação, várias abordagens foram desenvolvidas, tais como, *2-Port networks*, abordagem *Scattering*, *Wave Variables*, Decomposição Passiva, Controle Adaptativo, etc [64].

Um fator importante no ramo de teleoperação é o de autonomia, a qual em um contexto de uma interação humano-robô pode ser representada pelo grau de solicitações de um robô ao operador durante a comunicação [70], ou pelo “volume de tempo que um robô pode ser negligenciado” [6]. Segundo Barros e Lindeman [70], um simplificado estabelecimento de níveis de autonomia de robôs para diferentes modos de operação baseados nos achados da literatura, podem ser apresentados como: (1) completamente controlado, (2) controle compartilhado (teleoperação segura, semiautônomo e alto nível de teleoperação) e (3) completamente autônomo. O trabalho desenvolvido nesta dissertação abrange o nível de controle compartilhado pela subdivisão de teleoperação segura na qual as ações no ambiente são tomadas conjuntamente entre o operador e o robô, embora o operador controle diretamente na maior parte da interação os efeitos que o robô realiza no ambiente.

Diferentemente do conceito de teleoperação estar ligado ao de interação remota, neste trabalho nós definimos também teleoperação ligado ao pensamento de interação próxima por duas razões: (1) Os mecanismos trabalhados neste projeto permitem o controle de um robô humanoide independentemente do robô estar no campo de visão direto do operador. Se fora desse campo (ou seja, distante do operador, não-tele), bastaria utilizar as câmeras do robô e a representação de sua postura por meio de um simulador. (2) Uma parte considerável da utilização do robô pelo operador neste projeto se revela por uma imitação direta das configurações abstraídas do movimento do usuário, demonstrando assim, a falta de liberdade do robô em agir autonomamente. Em outras palavras, o robô age no ambiente quase que exclusivamente pelos movimentos desempenhados por um usuário.

2.3 Interação com robôs humanoides

Cheng e Kuniyoshi [71] utilizam um humanoide para imitação dos braços de um usuário por meio de segmentação de cor de pele de apenas estas partes descobertas (veja Figura 5). A estimação da configuração da cabeça fora utilizada para orientar o movimento da cabeça/pescoço/torso do robô a fim de manter o usuário dentro do campo de visão da câmera acoplada no humanoide, e a imitação dos braços era abstraída por meio de dois graus de

Figura 5: Detecção de usuário por meio de cor de pele e imitação realizada por um robô humanoide. [71]

liberdade no ombro e um no cotovelo.

Riley *et al.* [72] usaram marcros coloridos passivos na parte superior do corpo de um usuário (demonstrado na Figura 6) e câmeras externas e montada na cabeça do robô a fim de abstraírem a configuração de um modelo representando o estado de um usuário. Esses dados eram mapeados para um robô humanoide imitar as posturas do usuário. Quanto aos experimentos realizados, fizeram uso da imitação com dados reais apenas para a parte superior do robô. Para a parte inferior, os autores comentam que basta o acréscimo de mais marcros para abstração das configurações dessa região – no entanto, a parte inferior não fora abordada com dados reais.

Figura 6: Humanoide imitando movimento da parte superior de um usuário. [72]

Um sistema de câmeras estéreo é utilizado em [73] para detecção de cor de pele em imagens e estimação de profundidade com relação a uma pessoa. Esses dados brutos são trabalhados para cálculos de ângulos das junções do ombro e cotovelo do usuário a fim de mapeá-los em um humanoide. A cabeça e mãos são as únicas partes do corpo do usuário descobertas, ou seja, apenas essas regiões revelavam cor de pele. Os autores focaram em estimar as configurações angulares das junções e não a posição relativa destas durante a imitação. Assim, os resultados revelados demonstram limitação na imitação dos movimentos dado à restrição da arquitetura física do robô utilizado – um modelo simulado demonstrou melhor imitação (visivelmente falando). Figura 7 exemplifica alguns experimentos.

Figura 7: Imitação dos braços de um usuário por meio de um sistema de câmeras estéreo. [73]

Em [74] um sistema de captura de movimento óptico, composto por algumas câmeras espalhadas em uma sala, foi utilizado para abstrair 30 marcros posicionados no corpo do usuário em experimentos que objetivavam um robô humanoide imitar os movimentos de todo o corpo de uma pessoa ao realizar uma determinada dança. Também, um módulo de balanço foi desenvolvido para permitir estabilidade do robô na execução dos movimentos – sob a restrição dos pés fixos ao chão (demonstrado na Figura 8). Nos experimentos a imitação não foi realizada em tempo real.

Figura 8: Robô mimcando movimentos realizados por um dançarino. [74]

Ott *et al.* [75] realizaram uma mímica por meio de um robô humanoide através de uma abordagem Cartesiana. Um sistema de captura de movimento foi utilizado para abstrair a presença de trinta e quatro marcos, com formas de mola, sobre o corpo de um usuário e dois marcos sobre um tipo de bastão. Os autores também definiram onze movimentos em particular para aprendizagem, reconhecimento e execução do robô por meio de onze Modelos Ocultos de Markov. Figura 9 demonstra o reconhecimento e a execução de um dos onze movimentos aprendidos.

Figura 9: Um tipo de movimento definido chamado *punch* reconhecido e reproduzido por um humanoide. [75]

Além da utilização na área de jogos, o advento do Microsoft Kinect permitiu a utilização desse dispositivo em meios computacionais para variadas propostas, inclusive em manipulação de robôs.

Suay and Chernova [76] utilizaram o Kinect junto com um humanoide para permitir que o robô imitasse o movimento dos braços do usuário, tivesse seu caminhar ordenado, e o olhar do robô fosse direcionado pela posição da mão direita para ordem de deslocamento. Estas funcionalidades foram obtidas pelo estabelecimento de alguns modos de controle com relação à posição de alguns membros do corpo do usuário. Figura 10 exemplifica uma amostra da interação realizada.

Figura 10: Robô NAO sendo controlado pela interpretação da posição do usuário. [76]

Koenemann e Bennewitz [77] utilizaram uma vestimenta de sensores inerciais no corpo de um usuário a fim de capturar as configurações das junções de seu corpo e mapeá-las para as junções do robô em tempo-real. Também, foi desenvolvido um módulo de balanço que permitisse o robô se comportar de forma estável durante a imitação conforme mostrado na Figura 11. O robô humanoide desempenhou a imitação adequadamente durante a interação, permitindo também a estável imitação do movimento de ambas as pernas sem cair no chão.

Figura 11: Humanoide imitando o movimento de todo o corpo do usuário. [77]

Veltrop [78] fez uso da fusão de vários sensores para teleoperar um humanoide NAO. Para tal, utilizou o Kinect para estimação das configurações espaciais do corpo do usuário; um controle *Wii* para melhor estimação das configurações da mão/punho e ativação de determinados comportamentos no robô; uma esteira para permitir o robô caminhar conforme o caminhar do usuário na esteira evitando, assim, que o mesmo saia do campo de visão do Kinect; e um *head-mounted display* servindo como a visão do robô no usuário. Figura 12 exibe uma amostra dessa teleoperação.

Figura 12: Usuário teleoperando robô por meio de fusão de sensores. [78]

Abaixo na Figura 13 há uma amostra da evolução dos trabalhos abordados nessa seção com respeito a imitação (sem abordagens de aprendizado) e/ou controle em robôs humanoides.

Diante dos trabalhos expostos, a semelhança deste projeto abrange a imitação e controle de um robô humanoide. Diferentemente, é abordada uma discussão dos resultados gerados na utilização de um sensor não-intrusivo no contexto robótico.

Os dados utilizados na detecção do ser humano frente à câmera são apenas *profundidades* do ambiente onde a pessoa se encontra, isto é, cada pixel da imagem contém um valor referente a quão distante o pixel se encontra do sensor. Para o entendimento de comportamento humano foi utilizado o reconhecimento dos estados das configurações de algumas partes do corpo do usuário com relação a estados que definimos manualmente, juntamente com o reconhecimento de alguns gestos. Esse entendimento é realizado por meio de conceitos matemáticos abordados no capítulo seguinte.

Figura 13: Uma amostra da evolução de trabalhos que abordam imitação e/ou controle em robôs humanoides.

2.4 Considerações finais

Neste capítulo abordamos a literatura em Sensoriamento Humano descrevendo os multiformes meios que um ser humano pode ser compreendido através de abordagens instrumentadas ou não-instrumentadas. Após, a especificação em robótica por meio da pesquisa em Teleoperação foi abordada, juntamente com suas aplicações. Por fim, trabalhos próximos ao que este propõe, relacionados à detecção/rastreamento de uma pessoa e ao reconhecimento de alguns comportamentos para imitar e controlar um robô humanoide, foram também relatados.

Capítulo 3

Métodos de Pesquisa

Este capítulo apresenta as ferramentas adotadas para abstrair o fundamento que proporciona a interação humano-robô a ser desenvolvida neste trabalho, a qual consiste em interpretar a configuração de um ser humano quanto às suas junções e mapeá-las para uma ação apropriada a ser realizada por um robô. O robô utilizado é o robô NAO existente no LAR/ICMC/USP.

Deste modo, serão apresentados os fundamentos matemáticos referentes a espaço de coordenadas, matriz de rotação, relações trigonométricas e, em seguida, serão explanadas as ferramentas que adotamos para atingir o objetivo proposto, tais como o dispositivo Kinect, o *framework* OpenNI, e finalmente o robô NAO utilizado nos nossos experimentos.

3.1 Fundamentos matemáticos

Nesta seção, serão apresentados alguns conceitos de sistemas de coordenadas e orientação de objetos, utilizados para cálculo da configuração angular das junções existentes nos movimentos de membros do corpo humano.

3.1.1 Espaço de coordenadas e orientação de objetos

Um espaço de coordenadas define a localização de qualquer objeto localizado sobre ele através da origem e dos eixos que compõe esse espaço. Alguns tipos de espaços de coordenadas comumente utilizados [79] são:

- **Mundo:** Um espaço de coordenada que estabelece um quadro de referência geral para qualquer outro espaço que porventura seja especificado. Este espaço estabelece posições absolutas para qualquer objeto em estudo dentro desse espaço. Todavia, essas posições absolutas são apenas com respeito a esse espaço e não a uma definição literal de posição absoluta. Este espaço é também conhecido como espaço de coordenada global ou universal.

- Objeto: Qualquer objeto possui um espaço de coordenada relativo a si. Quando um objeto translada, o espaço de coordenada do mesmo também se move de acordo com o movimento ocorrido dado que ele é fixo a um ponto origem pré-definido no objeto. O espaço do objeto também é conhecido como espaço de modelo ou espaço de corpo.
- Câmera: O espaço de coordenada de câmera é um tipo especial de espaço de objeto, no qual a câmera (o objeto) tem a origem em si própria. Esse espaço possui três dimensões.

Vamos exemplificar o exposto acima por meio das Figuras 14 e 15. Em um espaço de coordenadas de três dimensões² (3D), temos um robô posicionado no ponto origem do espaço de coordenadas do mundo (Figura 14 no canto esquerdo). Após isto, este robô sofre uma rotação (Figura 14 no meio) e translação para um espaço onde se localiza uma cozinha. Na Figura 14 no canto direito, é posicionada uma câmera dentro da cozinha em frente ao robô.

Figura 14: Representação de um espaço de coordenadas. [79]

Nisto, podemos exemplificar cada um dos espaços de coordenadas abordados considerando uma perspectiva absoluta com relação ao sistema de coordenadas. O espaço de coordenadas da Figura 15A representa o quadro de referência para os demais espaços, ou seja, ele é do tipo mundo. Figura 15B representa um espaço de coordenada localizado no centro do robô, pois se refere ao espaço de objeto (robô). Por fim, o espaço de câmera é representado na Figura 15C, localizada no centro da mesma.

Figura 15: Representação dos espaços de coordenadas mundo (A), objeto (B) e câmera (C). [79]

Algumas formas de representar a orientação de um objeto em particular com relação a um espaço de coordenadas podem ser abordadas por matrizes de rotação e ângulos de Euler, cujas definições são apresentadas a seguir:

- Matriz de Rotação: Uma matriz de rotação representa a orientação de um espaço de coordenada com relação a outro espaço, por meio da descrição da base de seu espaço vetorial relativa à base do outro espaço de coordenada. Pode ser

² Considere o eixo Y projetado a partir do ponto origem perpendicular ao chão.

genericamente representada como uma matriz 3x3 (um espaço vetorial de três dimensões), conforme apresentada no canto superior direito da Figura 16.

- Ângulos de Euler: Os ângulos de Euler representam a orientação de um determinado objeto com relação a três eixos perpendiculares. Esses três ângulos são comumente representados pelos termos *yaw*, *pitch* e *roll*. A Figura 17 exemplifica os eixos, através dos quais esses ângulos são estimados.

3.1.2 Funções matemáticas utilizadas na abstração das configurações

Regras de três simples, distância Euclidiana, funções trigonométricas e transformação de matriz de rotação para ângulos de Euler são os conceitos utilizados para mapear as configurações das junções do usuário para as do robô.

Uma regra de três simples, exemplificada como:

$$\frac{x}{\limite_x} = \frac{y}{\limite_y}$$

é equivalente a

$$x = \frac{y * \limite_x}{\limite_y}.$$

Essa regra fornece uma relação entre variáveis x e y de um sistema qualquer. As variáveis $\limite_x$ e $\limite_y$ definem o limite que as variáveis x e y podem alcançar em uma dada escala numérica.

A distância Euclidiana entre dois pontos caracterizados em um espaço de três dimensões é descrita como:

$$\sqrt{(a_x - b_x)^2 + (a_y - b_y)^2 + (a_z - b_z)^2},$$

Figura 16: Uma matriz de rotação (canto superior direito) descrevendo a orientação dos três eixos do espaço de coordenada do objeto (jato) com relação a outro espaço de referência (em cinza). [79]

onde a e b são pontos localizados no espaço citado. a_x e b_x , a_y e b_y , a_z e b_z são os valores em alguma unidade de escala representando os pontos nos eixos X, Y e Z do espaço, respectivamente. Essa regra permite o cálculo do comprimento de alguns membros do corpo humano.

Uma junção humana, em um espaço de coordenadas de três dimensões, pode ter sua orientação – em um grau de liberdade – calculada através da regra do cosseno de um dos ângulos definidos por um triângulo, de acordo com:

$$\cos \alpha = \frac{b^2 + c^2 - a^2}{2bc},$$

onde a , b e c são os lados de um triângulo e a o lado oposto ao ângulo α . O ângulo correspondente a esse cosseno é obtido pela derivada de sua inversa:

$$\cos^{-1}(x) = f'(x) = \frac{-1}{\sqrt{1-x^2}}.$$

A fim de abstrair a configuração angular em três dimensões de alguma junção do corpo humano, com respeito à matriz de rotação da junção, pode-se transformar a matriz de rotação para três ângulos de Euler. A abstração de ângulos de Euler de uma matriz de rotação, conforme [79], foi codificada em linguagem de programação C++, cujo código é apresentado na Figura 18. A representação de ângulos de Euler aqui é do tipo canônico, onde *yaw* e *roll* estão na faixa de $\pm 180^\circ$ e *pitch* na faixa de $\pm 90^\circ$. Também, são tratados problemas caso haja valores fora da faixa esperada devido à aritmética de ponto flutuante e do tipo *Gimbal lock* [79].

3.2 Microsoft Kinect

O dispositivo Kinect foi lançado originalmente para o vídeo game Xbox pela Microsoft Corporation em novembro de 2010. Este mecanismo permite que o controle do vídeo game seja substituído pelo reconhecimento do movimento do corpo do usuário para interagir com alguns jogos. Em outras palavras, o usuário é o controle.

Ele é constituído de uma câmera RGB, um sensor de profundidade (*depth sensor*) composto de um emissor infravermelho e de um sensor monocromático CMOS, um conjunto de canais de microfones e um tipo de alavanca localizado abaixo do conjunto desses sensores

Figura 17: Rotação dos três eixos do espaço de coordenada de um jato demonstrando os ângulos de Euler associados. O ângulo *yaw* é associado com a rotação do eixo Y. *Pitch* associado com a rotação do eixo X. E *roll* associado com a rotação do eixo Z. [79]

```

//Assumindo que os valores da matriz de rotação são representados pelas variáveis mij.
//Cada mij representa o valor da i-ésima linha e j-ésima coluna da matriz m.
float m11,m12,m13;
float m21,m22,m23;
float m31,m32,m33;
float yaw, pitch, roll ;

float sp = -m32;

if (sp <= -1.0f)
    pitch = -1.570796f; // - $\pi/2$ 
else if(sp >= 1.0f)
    pitch = 1.570796f; //  $\pi/2$ 
else
    pitch = asin(sp);

if(fabs(sp) > 0.9999f){
    roll = 0.0f;
    yaw = atan2(-m13, m11);
}
else{
    yaw = atan2(m31,m33);
    roll = atan2(m12,m22);
}

```

Figura 18: Código em C++ da transformação da matriz de rotação para ângulos de Euler. [79]

que permite o dispositivo se acomodar ao campo de visão com relação à disponibilidade do usuário na frente do mesmo. Na Figura 19 é mostrado o dispositivo.

Logo após o lançamento deste, houveram algumas tentativas de modificações na operação do Kinect que objetivavam sua utilização por meio de plataformas *desktops*. O resultado dessas intervenções foram amostras de alguns pequenos aplicativos, que permitiam a utilização do dispositivo em um Computador Pessoal (PC). Não muito tempo depois, versões preliminares de bibliotecas foram disponibilizadas para o público por meio de alguns meios tais como a empresa PrimeSense (fabricante de parte do hardware do Kinect), e pela própria Microsoft.

Cerca de um ano e três meses após o lançamento do dispositivo, a Microsoft lançou uma versão estável de sua biblioteca chamada “Kinect for Windows SDK” [80], juntamente

Figura 19: Dispositivo Kinect.

com um hardware Kinect para utilização apenas em PCs Windows³.

Além da versão disponibilizada pela Microsoft, outras bibliotecas disponíveis para utilização em multi-plataformas são OpenNI (PrimeSense) [81] e OpenKinect [82]. O *framework* OpenNI é o utilizado neste projeto e será descrito a seguir.

3.3 Framework OpenNI

O *framework* OpenNI [81] permite a utilização do hardware Kinect em PCs para uma variedade de aplicações. Desde que uma pessoa esteja à frente do dispositivo, é possível compreender a quantidade de pessoas na frente do mesmo, as configurações espaciais e angulares com relação às junções do corpo humano, o reconhecimento de alguns gestos de mão, a utilização de dados brutos em cada um dos sensores que compõem o dispositivo (a câmera RGB, o sensor de profundidade e os canais de som), o reconhecimento de movimento da mão para controle de alguma interface, etc.

Deste *framework*, tomamos vantagens de 2 dados principais:

1. Máscara de esqueleto: As configurações espaciais e angulares citadas acima são fornecidas pela biblioteca por meio de uma “máscara de esqueleto”. Em poucas palavras, essa máscara é uma representação em forma de esqueleto onde os terminais de cada segmento desse modelo são referenciados como algumas junções do corpo do usuário. Cada uma dessas junções tem 3 propriedades, sendo elas: a posição em 3D e uma matriz de rotação com relação ao espaço da câmera-Kinect, e um valor de confiança referente ao nível de acurácia que se possui quanto a esses dois dados anteriores. Desta forma:
 - a. Posição em 3D: A posição de uma junção em 3D é representada por 3 eixos de coordenadas com origem no centro do dispositivo Kinect, onde duas coordenadas (aqui representadas por x e y) representam em milímetros a distância de uma junção com relação a um eixo horizontal (x) e vertical (y) imaginário com referência a esse ponto origem. A outra coordenada (aqui representada por z) representa o quão distante (em milímetros) a junção se encontra afastada a partir do centro do dispositivo.
 - b. Matriz de rotação: Uma matriz de rotação de 3D, cujos valores representam a orientação da base do espaço vetorial de um objeto (no caso a junção) relativo ao espaço de coordenadas do Kinect.
 - c. Fator de confiança: O valor de confiança pode ser encontrado em apenas duas formas⁴: 0% (o rastreamento da junção foi perdido ou ocluído por algum outro membro) ou 100% (quando a estimação parece estar funcionando corretamente).
2. Reconhecimento de alguns gestos pré-definidos pelo OpenNI: O *framework* possui alguns poucos gestos pré-definidos de mão passíveis de utilização. Neste trabalho são utilizados apenas dois gestos para permitirem a ativação de alguns

³ Alguns trabalhos independentes à Microsoft têm sido propostos na tentativa de forçar a utilização do “Kinect for Windows SDK” em sistemas operacionais Linux e MacOS X.

⁴ Sem a utilização de um modo heurística que a biblioteca permite.

modos de comportamentos (ações) definidos. Estes serão explanados mais adiante na seção 3.6.

Os dados utilizados por este trabalho são produzidos apenas pelo sensor de profundidade do dispositivo, o qual transmite dados a uma faixa de 30 quadros por segundo. Então, este é o único sensor do Kinect utilizado no projeto.

Figura 20 representa uma máscara de esqueleto estimada, o ponto de origem das coordenadas do dispositivo Kinect e a fusão dessas duas exemplificações.

3.4 Robô NAO

O robô NAO é um humanoide de quase 60 cm de altura, com massa em torno de 5 Kg, assemelha-se ao tamanho de uma criança e possui 25 graus de liberdade. Desenvolvido por uma empresa francesa, Aldebaran Robotics, este robô passou a possuir maior expressão junto à comunidade robótica por ter sido escolhido para substituir o extinto robô Aibo da Sony Corporation, em uma categoria da competição mundial de futebol de robôs, *RoboCup Standard League* [83]. O robô tem sido utilizado em mais de 350 universidades e laboratórios de pesquisa espalhados pelo mundo. Robótica, Inteligência Artificial, Ciência de Computação divergindo para Sociologia e Cuidados Médicos são algumas áreas de pesquisa onde esse robô tem sido utilizado.

O robô NAO possibilita a interação com outros robôs ou seres humanos apresentando distintas formas, as quais podem ocorrer por meio da fala, localização sonora, padrão visual e detecção de corpo, detecção de faces, detecção de obstáculos e por meio de efeitos visuais nos *LEDs* que possui. O robô possui um sistema multimídia com dois alto-falantes (localizados nas

Figura 20: De cima para baixo. 1 - Um quadro (*frame*) do sensor de profundidade com a representação de uma máscara de esqueleto. 2 - Origem do sistema de coordenadas da câmera Kinect. 3 - Sistema de coordenadas do Kinect sobreposto ao quadro.

laterais da cabeça), quatro microfones, duas câmeras digitais (dois orifícios, um na “testa”, outro na “boca”). Também, um sensor tátil na cabeça e sonares no seu torço. A Figura 21 apresenta o robô e alguns detalhes de seu corpo.

3.5 Articulações

As junções (ou articulações) de um ser humano podem ser caracterizadas de acordo com a direção na qual podem realizar algum movimento, variando de uma a até três direções. Junções que permitem apenas movimentos como flexão e extensão (em alguns casos também hiperextensão) são definidas como mono-axiais, como o cotovelo e joelho, pois são realizadas apenas sobre um eixo de direção. As bi-axiais, além de flexão/extensão, desempenham também movimentos de adução e abdução – como o punho. Embora seja possível rotacionar o punho em torno do eixo do antebraço, o mesmo é definido como bi-axial, dado que essa rotação se estende a todo antebraço, não sendo este movimento restrito ao pulso. Por fim, ombros e quadris realizam os mesmos movimentos tomados pelas bi-axiais, juntamente com movimentos em rotação, assim sendo chamadas de tri-axiais.

3.5.1 Medições do robô e do usuário

O primeiro objetivo deste trabalho é desenvolver um sistema que torne o robô capaz de imitar o movimento do corpo de uma pessoa com referência a algumas junções. Isso é alcançado pelo cômputo dos ângulos de Euler de algumas junções e mapeá-las para as correspondentes do robô. As junções utilizadas na imitação são o pescoço, ombro, cotovelo, quadril e joelho.

Antes de partir para o cálculo das configurações angulares das junções, faz-se necessário a expressão de alguns segmentos com relação aos membros superiores (braços) e inferiores (pernas) do corpo humano, já que estes são utilizados no cálculo da orientação de

Figura 21: Robô NAO e alguns sensores do seu corpo. No sentido do topo para baixo: 1 - Sensor tátil na cabeça. 2 - Microfones. 3 - Alto-falante na lateral da cabeça. 4 - Sonares no torço.

algumas junções. O comprimento dos segmentos correspondentes ao braço superior, antebraço, segmento entre o ombro e a mão, perna superior, perna inferior, e segmento entre o quadril e o pé, são calculados utilizando-se a distância Euclidiana entre dois pontos, em cada segmento, sendo dados por, respectivamente:

$$\text{braço_superior} = \sqrt{(\text{ombro}_x - \text{cotovelo}_x)^2 + (\text{ombro}_y - \text{cotovelo}_y)^2 + (\text{ombro}_z - \text{cotovelo}_z)^2},$$

$$\text{antebraço} = \sqrt{(\text{cotovelo}_x - \text{mão}_x)^2 + (\text{cotovelo}_y - \text{mão}_y)^2 + (\text{cotovelo}_z - \text{mão}_z)^2},$$

$$\text{segmento_ombro_mão} = \sqrt{(\text{ombro}_x - \text{mão}_x)^2 + (\text{ombro}_y - \text{mão}_y)^2 + (\text{ombro}_z - \text{mão}_z)^2},$$

$$\text{perna_superior} = \sqrt{(\text{quadril}_x - \text{joelho}_x)^2 + (\text{quadril}_y - \text{joelho}_y)^2 + (\text{quadril}_z - \text{joelho}_z)^2},$$

$$\text{perna_inferior} = \sqrt{(\text{joelho}_x - \text{pé}_x)^2 + (\text{joelho}_y - \text{pé}_y)^2 + (\text{joelho}_z - \text{pé}_z)^2},$$

$$\text{segmento_perna_pé} = \sqrt{(\text{quadril}_x - \text{pé}_x)^2 + (\text{quadril}_y - \text{pé}_y)^2 + (\text{quadril}_z - \text{pé}_z)^2}.$$

Assim, na Tabela 1 são mostradas as relações entre as junções do usuário e as do robô NAO com seus intervalos angulares limites. Não definimos um intervalo angular do movimento do pescoço humano dado à variabilidade considerável dos valores encontrados na literatura conforme [84]. Também, todos os outros valores fornecidos para o humano são relacionados a “um número arredondado normalmente considerado durante a estimativa da quantidade de movimento presente” que as junções podem realizar [84].

As cores da Tabela 1 demonstram a relação das variáveis entre o par humano-robô, isto é:

- Cor verde: Existência do determinado ângulo de Euler no lado humano ou robô.
- Cor amarela e laranja: Referência do ângulo de Euler da junção humana com relação à junção do robô. Em outras palavras, o *yaw* do ombro humano é mapeado para o *yaw* do cotovelo do robô. E o *pitch* do cotovelo humano para o *roll* do cotovelo do robô.
- Cor vermelha: Sem correspondência no lado humano ou do robô, ou não utilizado no trabalho.

Tabela 1: Correspondência de junções entre uma pessoa e o robô NAO.

		<i>Humano</i> [84]	<i>Robô NAO</i>	
J u n ç õ e s	<i>Pescoço</i>	<i>Yaw</i>	-	-119,5° a 119,5°
		<i>Pitch</i>	-	-38,5° a 29,5°
		<i>Roll</i>	-	Sem correspondência
	<i>Ombro</i>	<i>Yaw</i>	-90° a 90°	Sem correspondência
		<i>Pitch</i>	-45° a 180°	Ambos: -119,5° a 119,5°
		<i>Roll</i>	0° a 180°	Esquerdo: 0,5° a 94,5° Direito: -94,5° a -0,5°
	<i>Cotovelo</i>	<i>Yaw</i>	Sem correspondência	Ambos: -119,5° a 119,5°
		<i>Pitch</i>	0° a 145°	Sem correspondência
		<i>Roll</i>	Sem correspondência	Esquerdo: -89,5° a -0,5° Direito: 0,5° a 89,5°
	<i>Quadril</i>	<i>Yaw</i>	-35° a 45°	Ambos: -65,62° a 42,44°
		<i>Pitch</i>	-10° a 120°	Esquerdo: -101,63° a 27,3° Direito: -101,54° a 27,82°
		<i>Roll</i>	0° a 45°	Esquerdo: -21,74° a 45,29° Direito: -42,3° a 23,76°
	<i>Joelho</i>	<i>Pitch</i>	-35° a 45°	Esquerdo: -5,29° a 121,04° Direito: -5,9° a 121,47°

Pescoço. O pescoço do robô NAO tem apenas 2 graus de liberdade, enquanto um ser humano tem três.

Ombro e cotovelo. (1) O humanoide não possui um grau de liberdade que corresponda ao *yaw* do ombro humano. Este grau de liberdade é localizado no cotovelo do robô, permitindo, então, que apenas o antebraço rotacione – diferente do movimento humano que corresponde à rotação de todo o braço. Também, o *yaw* abstraído da matriz de rotação do cotovelo humano retorna valores correspondentes ao *yaw* do ombro humano (comprovados por teste empíricos que realizamos). Assim, decidimos utilizar este para o correspondente *yaw* do cotovelo do robô. (2) Embora o ângulo *pitch* do cotovelo humano não tenha seu correspondente para o robô, o mesmo é mapeado para o *roll* do robô – essa divergência é apenas uma questão de nomenclatura, implementada diferentemente pela criadora do robô, já que a direção de movimento pela qual a junção realiza é a mesma.

Quadril. A rotação (*yaw-pitch*) do quadril do robô não permite o movimento de apenas um quadril independente do outro, pois os dois quadris são ligados a apenas um motor – então, eles não podem ser controlados separadamente. Deste modo, foi decidido não utilizar este grau de liberdade pela não correspondência natural para o humano.

Outras junções – como as do punho e calcanhar – não foram utilizadas para imitação porque o *framework* OpenNI não estima a orientação delas.

3.5.2 Regras mapeadas para junções

Na Tabela 2 são mostrados onde os conceitos matemáticos sobreditos foram utilizados para cálculo dos ângulos de cada junção.

Tabela 2: Regras utilizadas no cálculo das configurações das junções.

			Técnicas
J u n ç õ e s	Pescoço	<i>Yaw</i>	matriz de rotação → ângulos de Euler
		<i>Pitch</i>	matriz de rotação → ângulos de Euler
	Ombro	<i>Pitch</i>	regra de três
		<i>Roll</i>	regra de três
	Cotovelo	<i>Yaw</i>	matriz de rotação → ângulos de Euler
		<i>Roll</i>	ângulo de um triângulo
	Quadril	<i>Pitch</i>	matriz de rotação → ângulos de Euler
		<i>Roll</i>	regra de três
	Joelho	<i>Pitch</i>	ângulo de um triângulo

Para melhor compreensão dos cálculos a serem apresentados adiante, são necessárias as seguintes notações:

- Seja “ $nome_de_junção_t$ ” denotar o valor referente à distância em milímetros da junção com relação ao espaço t , onde $nome_de_junção \in [pescoço, ombro, cotovelo, mão, quadril, joelho, pé]$ e t denote algum dos três eixos de coordenadas. Por exemplo, $ombro_z$ tem seu valor referente à distância em milímetros da junção *ombro* com relação ao espaço de câmera do Kinect no eixo Z. Quando apenas a variável designa um “nome de segmento”, o valor desta refere-se ao seu comprimento em milímetros – por exemplo, *braço_superior*.
- Seja também “ $NAO_junção_{ângulo\ de\ Euler}$ ” denotar o valor referente a um dos três ângulos de Euler de uma junção do robô NAO. Por exemplo, NAO_ombro_{pitch} refere-se ao valor em grau do *pitch* da junção *ombro* do robô NAO. Quando acrescida de “*limite*” refere-se ao limite que aquele grau de liberdade, naquela junção robótica, apresenta.

Assim, são apresentados a seguir o cálculo dos ângulos de Euler de cada junção abordada na imitação.

Pescoço. O *yaw* e *pitch* do pescoço são estimados pela transformação da matriz de rotação do pescoço em ângulos de Euler (conforme Figura 18), mas testes empíricos demonstraram que a variação dos dois ombros são os fatores preponderantes na estimação desse valor.

Ombro. O *pitch* do ombro é calculado pela regra de três, dada por:

$$z = ombro_z - cotovelo_z,$$

$$NAO_ombro_{pitch} = \frac{z * NAO_ombro_{pitch_limite}}{braço_superior}.$$

O *roll* do ombro é calculado pela regra de três, dada por:

$$x = ombro_x - cotovelo_x,$$

$$NAO_ombro_{roll} = \frac{x * NAO_ombro_{roll_limite}}{braço_superior}.$$

Cotovelo. O *yaw* do cotovelo é calculado pela transformação da matriz de rotação do cotovelo em ângulos de Euler (conforme Figura 18). O *roll* do cotovelo é calculado pelo cosseno de um dos ângulos formados pelo triângulo entre as junções ombro-cotovelo-mão, como visto na Figura 22. Assim:

$$cosseno\ ângulo_cotovelo = \frac{braço_superior^2 + antebraço^2 - segmento_ombro_mão^2}{2 * braço_superior * antebraço}.$$

Para chegar ao ângulo correspondente, calcula-se a inversa do cosseno.

Quadril. O *pitch* do quadril é calculado pela transformação da matriz de rotação do quadril em ângulos de Euler (conforme Figura 18). O *roll* do quadril é calculado pela regra de três por:

$$x = quadril_x - joelho_x,$$

$$NAO_quadril_{roll} = \frac{x * NAO_quadril_{roll_limite}}{\frac{perna_superior}{2}}.$$

Joelho. O *pitch* do joelho é calculado pelo cosseno de um dos ângulos formados pelo triângulo entre as junções quadril-joelho-pé, como visto na Figura 22. Assim:

$$cosseno\ ângulo_joelho = \frac{perna_superior^2 + perna_inferior^2 - segmento_quadril_pé^2}{2 * perna_superior * perna_inferior}.$$

Para chegar ao ângulo correspondente, calcula-se a inversa do cosseno.

Quanto à razão pela qual definimos esses três caminhos para o cálculo dos ângulos: Inicialmente, buscamos utilizar apenas os ângulos de Euler abstraídos da matriz de rotação. Entretanto, na maioria dos casos esses ângulos não representavam as reais configurações angulares das junções em um determinado tempo ou eram instáveis durante o movimento. Acreditamos que isso se deva a alguma limitação do dispositivo Kinect e/ou do *framework* OpenNI. Diante dessas limitações, tivemos que prover outros meios para cálculo dessas configurações. A utilização da regra de três e a inversa do cosseno de um ângulo de um triângulo foram alternativas úteis e simples que encontramos para uma medição satisfatória de tais configurações.

3.6 Funcionalidades do sistema

Para interação com o sistema que desenvolvemos, definiu-se o reconhecimento de dois meios para engatilhamento de determinados comportamentos:

1. Reconhecimento de gestos: O *framework* OpenNI permite o reconhecimento de alguns gestos pré-determinados. Utilizamos os gestos *push* (exemplificado como o aperto de um botão invisível com a mão) e *swipe* (semelhante a um gesto realizado com a mão com o significado de “Oi”).
2. Reconhecimento da localização de algumas junções: Foram definidas algumas regiões (representadas pelas coordenadas em 3D) onde a mão do usuário deveria estar, e uma pré-determinada configuração angular do joelhos do usuário para engatilhamento de alguns comportamentos definidos.

Além da imitação do movimento da cabeça, braços e pernas, foi desenvolvido um meio para controlar o caminhar do robô e a execução de alguns comportamentos. Assim, há sete modos de controle mais três comportamentos que podem ser realizados, tais como:

1. Estabilize NAO – mão esquerda sobre o ombro direito: Este modo é usado para fazer o robô ir para uma postura inicial de seu corpo e deixá-lo estável o suficiente para começar qualquer outro tipo de comportamento. Isso funciona como um modo de transição entre um modo sendo executado e outro qualquer.
2. Pausar – mão direita sobre o ombro esquerdo: O atual estado do sistema é pausado em apenas 2 modos. Se o robô estava imitando antes da pausa, a última postura exata realizada através da imitação é mantida. Ou caminhando, a última velocidade configurada é mantida. Para desativar o modo Pausar é necessário apenas fazer o mesmo para ativá-lo, exceto pela espera de um intervalo mínimo de 3 segundos entre a última ativação.
3. Imitação do pescoço – mão direita próxima e em frente à face: Começa a imitar o movimento do pescoço.
4. Imitação dos braços: Mão esquerda sobre o cotovelo direito para imitar o braço

J1 – Junção do ombro.

J2 – Junção do quadril.

T1 – Triângulo formado pelas junções do ombro, cotovelo e mão. O ângulo destacado no triângulo é o *pitch* do cotovelo.

T2 – Triângulo formado pelas junções do quadril, joelho e pé. O ângulo destacado no triângulo é o *pitch* do joelho.

1 – O *pitch* do ombro é estimado com relação à quão distante o cotovelo está do seu respectivo ombro no eixo Z.

2 – O *roll* do ombro é estimado com relação à quão distante o cotovelo está do seu respectivo ombro no eixo X.

3 – O *roll* do quadril é estimado com relação à quão distante o joelho está do seu respectivo quadril no eixo X.

Figura 22: Máscara de esqueleto e demonstração de ângulos de algumas junções.

direito. E mão direita sobre o cotovelo esquerdo para imitar o braço esquerdo. Também, aqui é possível abrir (ou fechar) os dedos se eles estiverem fechados (ou abertos) quando há pelo menos um braço desativado. O abrir/fechar das mãos se dá pelo reconhecimento da mão do braço desativado em uma região do espaço localizada a mais de 90% com relação ao tamanho do braço, a partir do ombro, através do eixo Z.

5. Perna suspensa – apenas um dos quadris levantados em pelo menos 60 graus com relação ao seu *pitch*: Se é o quadril direito levantado, começa a imitar o movimento das pernas mantendo o pé esquerdo fixado no chão enquanto a perna direita permanece livre para se movimentar. Se é o quadril esquerdo levantado, começa a imitar o movimento das pernas mantendo o pé direito fixado no chão enquanto a perna esquerda permanece livre para se movimentar.
6. Ativar modo de comportamento – um gesto *push* efetuado com a mão direita em uma área acima e ao lado do ombro direito: Permite a execução de alguns comportamentos feitos disponíveis pelo *framework* do robô. Há 3 comportamentos dentro deste modo:
 - a. Caminhar. O robô pode caminhar por seis formas: Para frente/trás, lateralmente para a esquerda/direita, e girar para a esquerda/direita. Para ativar o caminhar para frente ou lateral para direita, a mão direita do usuário deve estar distante do ombro a um intervalo de pelo menos 90% do tamanho do braço com relação ao eixo Z ou X, respectivamente. Para ativar a rotação para a direita o usuário deve posicionar a mão esquerda próxima ao cotovelo direito⁵.
 - b. Gesto Oi. O robô produz um gesto “Oi” através do reconhecimento de um gesto *swipe* realizado com a mão direita acima e ao lado do ombro direito.
 - c. Levantar/sentar. Se o robô está sentando, ele pode levantar. Caso contrário, ele pode sentar. Este comportamento é ativado quando o usuário flexiona seus joelhos em pelo menos 60 graus.
7. Desligar o robô. O reconhecimento de um gesto *push* realizado na região do torso faz com que o robô realize um gesto representando-se pela “limpeza de sua testa” demonstrando certo cansaço, e em seguida ele senta e desativa a rigidez de suas junções. Assim, o sistema é encerrado.

Durante o caminhar do robô, o comprimento do passo pode ser definido. Nós definimos cinco comprimentos (de zero até o máximo permitido), levando-se em conta o quão distante a mão do usuário está a partir do ombro através de uma das três dimensões. Eixo Z para o movimento para frente/trás, eixo X para as laterais, e eixo Y para girar. Também, a mão deve estar dentro de um paralelepípedo retângulo imaginário, o qual apresenta certas dimensões conforme vistas na Figura 23. O ombro é o ponto central do lado do paralelepípedo que o intercepta.

⁵ Para caminhar para trás, lateral para esquerda e rotação para esquerda basta trocar apenas o sentido da mão e do cotovelo – direita por esquerda e vice-versa.

Figura 23: Representação do paralelepípedo retângulo para controle do tamanho do passo do robô durante o caminhar. L1 = comprimento do braço. L2 = L3 = duas vezes o comprimento entre o pescoço e o ombro.

Todavia, se o comportamento de caminhar estiver em curso e a mão do usuário sai para fora da região do paralelepípedo retângulo, então o robô continua a caminhar pela última velocidade definida.

3.7 Outras observações

O projeto atual foi implementado em um sistema operacional Windows 7 64 bits, utilizando a *IDE* Visual Studio 2010 em C++, com a versão PrimeSense Package Build for Windows x64 v20.4.2.20 Development Edition do OpenNI, juntamente com o PrimeSense Sensor Module for OpenNI x64 v5.1.0.25. A arquitetura do robô NAO utilizada é a versão 3.2.

3.8 Considerações Finais

Neste capítulo introduzimos os mecanismos que viabilizaram a realização dos objetivos deste trabalho. Quanto ao aspecto de hardware, abordamos o sensor não-intrusivo Kinect e o robô humanoide NAO. Quanto às técnicas empregadas, versamos sobre alguns fundamentos matemáticos juntamente com a utilização do *framework* OpenNI. Também, explanamos as relações postas entre as funções matemáticas e às junções/membros utilizadas na imitação e no controle do robô. Por fim, o sistema desenvolvido foi abordado por meio das funcionalidades que este dispõe.

Capítulo 4

Experimentos Realizados

A fim de verificar o desempenho do sistema proposto, realizamos experimentos que foram conduzidos por meio de avaliações subjetivas com dez usuários. Para tal, utilizamos a técnica *Mean Opinion Score* (MOS) a qual atribui uma nota dentre uma escala de qualidade que varia de 1 (ruim) a 5 (excelente) a um dado objeto em estudo. Essa técnica foi inicialmente criada para avaliar a qualidade de comunicação entre duas pessoas em sistemas telefônicos. A avaliação era realizada por meio de respostas a um questionário estruturado concernente ao tipo de comunicação de um sistema telefônico.

4.1 Avaliação subjetiva

Como em uma interação humano-robô sempre existe um humano, com um papel mais fortemente ativo no ramo de teleoperação, podemos tomar vantagens de uma avaliação subjetiva vinda de usuários por serem estes os autores do controle do robô para execução de uma determinada tarefa.

A técnica *Mean Opinion Score* [85], em sua forma original, define um valor dentre 1 a 5 a algumas questões formuladas para análise da qualidade de um objeto em estudo. A característica qualitativa dessa escala é: 1 - Ruim, 2 - fraco, 3 - razoável, 4 - bom, e 5 - excelente.

Assim, foram definidos 4 aspectos a serem avaliados:

1. A imitação do pescoço;
2. A imitação dos braços;
3. O controle do caminhar do robô;
4. E a utilização da mão do robô através de um cenário no qual o humanoide deveria pegar uma sacola plástica em um ponto e levá-la à lixeira localizada em outro ponto.

Para avaliação dos 4 itens acima, foram considerados 10 usuários para interagir com o robô. Desta interação, foi gravado um vídeo e uma edição dele foi enviada a cada avaliador para examinar o mesmo. Assim, as perguntas formuladas para os quatro itens anteriores foram:

1. Imitação do pescoço e braços: Quão boa a imitação do movimento se revela? O robô NAO parece realmente me imitar durante o modo de imitação? A imitação é devagar, moderada ou rápida?
2. Caminhar: Quão bom é o controle do caminhar?
3. Agarrar: Quão bom se revela o agarrar da sacola plástica? Quão bom o robô consegue levar a sacola plástica para a lixeira?

Também, foi solicitado aos usuários para comentarem quanto à interação realizada, caso eles desejassem.

A fim de evitar uma tendenciosa avaliação foi solicitado aos usuários para retornarem os resultados da avaliação sem que os avaliadores se identificassem no documento.

Como resultado, Figura 24 expõe os dez usuários, Figura 25 os histogramas dos 4 aspectos avaliados na interação, e Tabela 3 revela a média das avaliações realizadas.

Considerando as 4 avaliações, obteve-se 80% como média geral de satisfação dos usuários.

Antes de prosseguirmos para os comentários dos avaliadores quanto à interação, uma definição do fator *impersistência* faz-se necessária para o entendimento de algumas justificativas com relação a alguns erros apresentados pelo robô.

O comportamento reativo do robô quanto às configurações das junções do usuário apresenta uma característica chamada *impersistência* [86]. Em vez de procurar a conclusão da execução de cada postura reconhecida no usuário, o robô procura executar o reconhecimento da última postura que foi abstraída, falhando em seguir com precisão a trajetória que o usuário realizou durante a imitação. Isto é devido ao contínuo ciclo restrito sensor→motor. Este fator é claramente percebido quando o usuário realiza movimentos muito rápidos onde o robô não consegue acompanhar todo o trajeto do movimento a tempo para execução.

Dessa forma, os comentários fornecidos pelos usuários com relação a alguns problemas ocorridos na interação são apresentados, seguidos de algumas ponderações.

Pescoço. A imitação do pescoço por meio da rotação dos ombros foi o principal fator apontado pelos usuários como errôneo nessa imitação.

Figura 24: Usuários interagindo com o robô por meio da imitação do pescoço (A e B), imitação dos braços (C, D e E), controle do caminhar (F e G) e "agarrar" (H, I e J).

Figura 25: Histogramas dos 4 aspectos avaliados por 10 usuários.

Como já informado, o *framework* OpenNI considera o deslocamento angular dos ombros (observado por testes empíricos) na estimação da orientação do pescoço. Todavia, essa limitação não impediu esse tópico de obter o melhor grau de estimação dentre os outros.

Braços. Alguns movimentos foram considerados trêmulos, abruptos, perdidos – em parte do rastreamento – e mais lentos que aqueles realizados pelo usuário.

Alguns movimentos abruptos e trêmulos podem ser explicados pela característica de impersistência descrita anteriormente. Quanto ao fato da aparente perda ou falta de correspondência em parte da imitação, esta é causada pelas limitações do rastreamento desempenhadas por meio do Kinect e da biblioteca OpenNI. A pouca lentidão que o robô apresentava para executar a imitação com relação ao tempo-real em que o usuário realizava algum movimento é explicada por meio da velocidade que definimos (até 50% da máxima permitida) – isso foi definido para reduzir a probabilidade de uma possível queda do robô caso o usuário movesse rapidamente seus braços e o robô (dado o fator impersistência), assim, procurasse apresentar sua postura tão rapidamente como aquela demonstrada pelo usuário.

Caminhar. O fator limitante nesta avaliação foi a dificuldade em controlar o caminhar do robô pela forma de controle definida.

Tabela 3: Resultado da avaliação MOS.

O que avaliar		MOS	Porcentagem
Imitação	Pescoço	4,4	88%
	Braços	4,1	82%
Caminhar		3,7	74%
Agarrar		3,8	76%

Nos testes realizados pelos autores não foi observada tamanha dificuldade para controlar o caminhar do robô NAO, comparada a apresentada pelos usuários. Acredita-se que isso se deve a falta de experiência dos avaliadores no controle definido. Possivelmente, haveria uma estimação melhor desse tópico caso fosse dado um período maior de testes para o usuário.

Agarrar. Aqui, o fator impersistência mais uma vez se demonstrou por meio da falta de suavidade e precisão no controle dos braços para que o robô pudesse agarrar uma sacola plástica.

Isto tornou a interação um pouco incômoda nesse quesito. Também, não foi definido⁶ um modo de controle da rotação do punho do robô – o qual corresponderia à rotação do antebraço do ser humano. Deste modo, os usuários deveriam realizar tal tarefa com uma fixa orientação do punho do robô.

Os outros comportamentos definidos para execução – gesto “Oi”, “Sentar/Levantar” e “Desligar o robô” – foram utilizados neste trabalho apenas como um meio de fazer o robô realizar determinados comportamentos amigáveis através do reconhecimento de alguns gestos e posição de algumas junções. Os mesmos não foram avaliados por serem comportamentos já incorporados no *framework* do robô, não havendo, assim, algum nível de complexidade considerável para serem expostos a exame. Quanto à falta da avaliação da imitação das pernas, esta é justificada na seção 5.3 do Capítulo 5.

Alguns experimentos podem ser encontrados em vídeos nos links:

- Caso 1: Imitação de braços e controle de comportamentos:
<https://www.youtube.com/watch?v=eQYChfMrQd0>
- Caso 2: Imitação de pernas (direita levantada):
<https://www.youtube.com/watch?v=Dwu5730o5YA>
- Caso 3: Cenário jogando sacola plástica na lixeira; testando “agarrar”:
<https://www.youtube.com/watch?v=3csdncKHmHs>

4.2 Considerações Finais

Abordamos neste capítulo a avaliação da interação de dez usuários com o robô por meio da imitação e do controle do humanoide NAO utilizando a técnica *Mean Opinion Score*. Foram obtidos bons resultados comparados a trabalhos relacionados à área de reconhecimento de padrões e expomos alguns comentários quanto aos quesitos examinados.

⁶ Diante da relativa dificuldade demonstrada na interação com o robô no cenário para agarrar uma sacola plástica, foi desenvolvido – após os experimentos realizados – um meio para controlar a rotação do punho do robô. O controle dessa junção era realizado durante a imitação dos braços no seguinte contexto: Se há um braço sendo imitado e outro não, pode-se controlar a rotação do punho (do braço realizando imitação) pelo engatilhamento desse controle ao reconhecer o *pitch* do cotovelo (do braço não imitado) no valor de 80 graus ou mais. Após esse reconhecimento, a configuração angular do punho era definida por meio da configuração do *yaw* do ombro (do braço não imitado). A rotação do punho do robô varia entre 104.5 a -104.5 graus.

Capítulo 5

Discussão dos Resultados

Aqui são descritas as experiências, achados e algumas limitações resultantes da utilização do dispositivo Kinect por meio do *framework* OpenNI, e do humanoide NAO. Também, a imitação das pernas e algumas das principais dificuldades encontradas no desenvolvimento deste trabalho são abordadas.

5.1 *Framework* OpenNI

De todas as funcionalidades que o *framework* OpenNI disponibiliza, este trabalho ateu-se a utilização da máscara de esqueleto e o reconhecimento de dois gestos.

5.1.1 Limitações na matriz de rotação

Como já abordado, a estimação da matriz de rotação por meio desse *framework* não se revelou estável na maioria dos testes realizados com as junções utilizadas nesse trabalho. Igualmente, o próprio manual do OpenNI informa a limitação da biblioteca na estimação da máscara de esqueleto quando as mãos/braços estão próximas ao torso do corpo. Esta limitação foi percebida por algumas dificuldades apresentadas durante o reconhecimento das posições de algumas junções para ativação de alguns comportamentos.

O *framework* não estima a matriz de rotação do punho e calcanhar. Caso houvesse uma estimação estável, poder-se-ia tomar vantagens desta para controlar essas junções no lado do robô, tornando assim mais natural o processo de imitação.

Quanto ao pescoço, a estimação de sua matriz de rotação é igual em valor à matriz de rotação da cabeça. Como já mencionado, a estimação da orientação da base do espaço vetorial do pescoço parece se basear nas bases de ambos os ombros. Obviamente, esta orientação não reflete a esperada do pescoço.

Algoritmos de estimação de postura por meio de câmeras podem ser bons substitutos para a estimação da orientação da cabeça podendo ser desenvolvidos pelo suporte de bibliotecas de visão computacional como o OpenCV [87].

5.1.2 *Yaw* do ombro

O *yaw* do ombro é na verdade abstraído pelo *roll* do cotovelo no OpenNI. Por meio de testes observa-se que não é possível estimar adequadamente a orientação dessa junção, quando o cotovelo está flexionado a cerca de 40 graus para menos. Para uma boa estimação considera-se a flexão do cotovelo em torno de 50 graus para mais. Entende-se que essa limitação ocorre não exclusivamente pela biblioteca, mas sim por apenas ser utilizado o sensor de profundidade, já que a rotação do ombro com o cotovelo não flexionado (ou pouco) não reflete em mudança significativa nos dados tomados pelo sensor.

5.1.3 Rastreamento dos membros inferiores

Durante o rastreamento das pernas foi percebido que o mesmo era parcialmente perdido quando o *pitch* do quadril estava em torno de 90 graus ou mais, ou quando a perna livre estava razoavelmente para trás do quadro do seu torso, ou quando o joelho estava oculto/ocluído da visão do sensor.

5.2 Robô NAO

Nesta seção, são descritas algumas características relevantes neste robô que influenciaram de alguma forma nos resultados obtidos neste trabalho.

5.2.1 Módulo de balanço e omissão de auto-colisão do NAO

O robô possui um módulo de balanço que objetiva mantê-lo estável sobre um ou ambos os pés, fixo(s) no chão, por meio da manutenção de seu centro de massa (localizado no torso) dentro de um polígono de suporte composto pelos seus pés. Apesar da utilização desse módulo, foi verificado que o mesmo não proveu estabilidade suficiente ao robô para realizar a imitação das pernas.

Também, o *framework* do robô possui um módulo de omissão de auto-colisão, apenas para os braços com relação ao corpo – não há para as pernas.

5.2.2 Dissimilar corpo do robô para uma pessoa

Embora o robô NAO esteja na categoria de humanoides, seu corpo é dissimilar no número de junções e na quantidade de graus de liberdade que algumas delas apresentam comparadas ao ser humano. Também, sua baixa estatura assemelha-se ao corpo de uma criança, dificultando, assim, uma interação com o ambiente mais próxima à estrutura de uma pessoa adulta.

5.3 Imitação das pernas

Neste trabalho é permitido que a imitação da perna seja realizada em apenas um cenário restrito: Um pé de uma perna do robô é fixo sobre o chão, enquanto a outra perna está livre para se movimentar. Na imitação do pé da perna fixa são considerados dois graus de liberdade

– o *pitch* do quadril e do joelho. Na perna livre são considerados os mesmos ângulos da perna fixa acrescido do *roll* do quadril.

Algumas questões podem surgir quanto essa restrição que foi imposta, tais como:

1) Por que a restrição do cenário quanto uma perna fixa e outra livre?

Para omissão de colisão dos pés contra o chão. Aqui, dois fatores são causadores dessa limitação:

- O dispositivo Kinect: Na estimação das configurações angulares das junções, há uma contínua variação dos valores estimados pelo dispositivo, mesmo se o usuário está estático em frente ao Kinect. Por exemplo, se o usuário se encontra com postura estática em frente à câmera, espera-se que a estimação da configuração angular dos quadris e joelhos seja constante em um determinado valor. Contudo, esse valor varia continuamente em torno de 7 graus entre um valor médio⁷.
- Falta da aplicação de um módulo de omissão de auto-colisão entre os membros inferiores e um módulo de balanço estável: Um módulo de omissão de auto-colisão poderia ser obtido pela definição de um valor mínimo do *roll* do quadril para que a perna livre não colidisse com a perna fixa. Contudo, o problema principal se revela na colisão do pé com o chão.

A fim de reduzir a colisão do pé com o chão, durante a imitação das pernas, procurou-se tomar cuidado mantendo a perna livre acima ou suavemente próxima do chão. Também, uma simples redução nessa colisão se deu por procurar manter sempre os pés do robô paralelos ao chão. Essa redução foi implementada pela simples atribuição das variações das orientações do quadril e joelho somadas para o pé. Por exemplo, se em um dado contexto a perna livre do robô apresenta um quadril com *pitch* em 40^0 e *roll* em 20^0 , e o joelho com *pitch* em 70^0 , bastaria apenas subtrair o *pitch* atual da junção do calcanhar do robô pela soma dos *pitchs* do quadril e joelho. Para o *roll* do calcanhar, bastaria subtrair-lo pelo *roll* do quadril, respeitando-se o intervalo angular da orientação do calcanhar do robô que varia em torno de -68 a $52,9$ graus do *pitch* e $44,1$ a $22,3$ graus do *roll*.

Aos usuários da avaliação subjetiva não foram permitidos interagir com o robô NAO pela imitação da perna devido a certa instabilidade que o robô apresentava durante essa imitação, podendo o robô cair.

2) Por que o grau de liberdade do *yaw-pitch* do quadril do robô não foi utilizado e qual a razão da não utilização do calcanhar na imitação?

Abordamos essas dúvidas na seção 3.5.1 e 5.1.1, respectivamente.

3) Por que não foram utilizados o *roll* do quadril da perna do pé fixado e outras partes do corpo em conjunto, como o pescoço e os braços durante a imitação das pernas?

Estes não foram utilizados para evitar maior instabilidade durante a imitação diminuindo, assim, a probabilidade do robô cair no chão.

⁷ É claro que esta variação poderia ser controlada por meio de um tratamento antecipado dessas configurações, buscando estipular um valor médio entre essa mutabilidade para serem repassadas para os motores das junções do robô. Um módulo de balanço eficaz e a omissão de auto-colisão de suas pernas suportariam isso.

Durante a imitação, o módulo de balanço do robô requer a transmissão com antecedência da mudança de orientação das junções, para que o robô se prepare em reproduzir o movimento de uma forma relativamente estável. Quando os dados não são passados com antecedência (isto é, cada nova entrada é esperada para ser executada em tempo real), a postura inicial do robô, encontrada no início da ativação desse módulo, é repassada para os motores do mesmo entre cada mudança de orientação das junções.

A fim de tentar omitir essa característica do módulo de balanço, uma adaptação foi realizada: Um quadro (*frame*) é tomado a cada segundo para obter as configurações das pernas humanas para serem mapeadas para as do robô. Então, essas configurações são transmitidas para os motores do robô para serem executadas durante um segundo. Após isso, o módulo de balanço é desativado e, após, imediatamente ativado. Esta desativação é feita para evitar que o robô volte à postura inicial, e a ativação para mantê-lo estável durante a imitação. Como resultado, o robô NAO realiza com certa instabilidade a imitação das pernas – isto é, há momentos em que o robô não suporta o balanço do seu corpo, caindo assim.

5.4 Principais dificuldades

Algumas das principais dificuldades encontradas durante o desenvolvimento deste projeto foram:

1. Imitação da perna: A imitação da perna pelo robô levou um tempo considerável para seu desenvolvimento visto que era necessário evitar o desbalanceamento do robô ao retirar o pé do chão no início da imitação, e verificar o quão estável ele permanecia durante a imitação das pernas sem cair. Mesmo diante do esforço empreendido, a imitação desses membros tem certa instabilidade pela falta do desenvolvimento de um módulo de balanço mais estável que o proporcionado pelo *framework* do robô.
2. Documentação fraca do *framework* do robô: A documentação do *framework* do robô disponibilizado pela empresa Aldebaran Robotics não era adequada o suficiente quando este projeto iniciou. Dessa forma, foi necessária a averiguação por força-bruta de algumas funcionalidades desse *framework* as quais seriam supostamente utilizadas no trabalho. Diante disso, foi necessária uma quantidade de tempo considerável para entendimento do funcionamento do robô.
3. Confusão/demora no reconhecimento de alguns modos: A confusão (reconhecimento de outro padrão em vez do esperado) no reconhecimento de alguns modos se revela pela proximidade de algumas posições que definimos para tais (por exemplo, mão esquerda próxima ao cotovelo direito ou ombro direito) e por alguma instabilidade que o rastreamento do usuário apresentava quando os membros do corpo estavam próximos à região do torso. Quanto à relativa demora no reconhecimento – e também no correr do sistema –, esta se revelou pelo atraso causado por alguns padrões de programação na linguagem C++, tais como:

- a. Chamada *cout* – impressão no *prompt*: A chamada desta função para informar o status do sistema gerava um relativo atraso no correr do mesmo, reduzindo assim, o desempenho do sistema.
 - b. Chamada *system("cls")* - limpeza de informação do *prompt* (para Windows): esta chamada degradava consideravelmente o reconhecimento de alguns gestos do OpenNI e também a leitura de dados do sensor de profundidade. A taxa média de 30 *fps* do sensor de profundidade era reduzida para um limite de 15 *fps* na utilização dessa chamada.
 - c. Outros – Antivírus/firewall: A utilização em segundo plano de um antivírus/firewall, junto com a chamada *system("cls")* ocasionalmente reduzia a taxa de *fps* para cerca de 5.
4. Tipo de piso onde o robô se encontrava: Ao desempenhar a imitação e a execução de alguns comportamentos, o robô NAO apresentava certa instabilidade com relação ao tipo de piso onde ele se encontrava. É sugerida a utilização do robô em pisos planos, estáveis e não escorregadios para suportar de forma estável o movimento do robô.
 5. Comunicação Wi-Fi não estável: A comunicação entre o robô e o PC foi realizada por meio de um cabo Ethernet. Dessa forma, houve um cuidado maior na operação do robô já que ele poderia se enroscar no cabo ou este chegar ao seu limite. A causa de não ser utilizada a comunicação Wi-Fi é devida a instabilidade nesse tipo de comunicação presente no hardware do robô.

5.5 Considerações Finais

Neste capítulo, foi apresentada uma discussão sobre alguns resultados não mensuráveis comparados aos abordados no Capítulo 4, mas valiosos quanto à experiência apreendida, destacando algumas limitações existentes nas informações produzidas pelo *framework* OpenNI, alguns entraves encontrados na utilização do robô NAO, a restrição da imitação das pernas e algumas das principais dificuldades abordadas no desenvolvimento deste trabalho.

Capítulo 6

Conclusão e Trabalhos Futuros

Neste projeto foi abordado a utilização de um dispositivo não-intrusivo para permitir que um robô humanoide imite o movimento do corpo de uma pessoa, e que o robô seja controlado por um usuário por meio do reconhecimento de posturas e gestos pré-definidos com relação a partes do seu corpo.

O dispositivo não-intrusivo e o robô utilizados foram o Kinect e o humanoide NAO, respectivamente.

A imitação do robô se revela por meio da imitação do pescoço, braços e pernas de uma pessoa. Algumas restrições durante a imitação foram definidas visando manter certa estabilidade no robô. Essas restrições impostas compreendem uma redução pela metade na velocidade máxima com a qual o robô é capaz de desempenhar durante a imitação dos seus membros, e um cenário restrito relativo à imitação das pernas quanto à fixação de um pé no chão e outro livre para se movimentar.

Quanto ao controle do robô, este trabalho permitiu que uma pessoa por meio de suas mãos ordene o robô a andar em sentido para frente/trás, lateralmente à esquerda/direita e em rotação para esquerda-direita. Alguns outros comportamentos como ordenar o robô sentar, levantar, e expressar alguns gestos foram também realizados.

Experimentos foram conduzidos quanto à avaliação do sistema desenvolvido focando no aspecto da imitação do robô. Adotando a avaliação subjetiva *Mean Opinion Score*, considerou-se 10 usuários para interagir com o robô por meio da imitação e do controle em determinados cenários. Dessa forma, foi gravado um vídeo dessa interação e repassado a cada usuário para avaliar o desempenho do robô conforme algumas diretrizes definidas. Na imitação foi avaliada a correspondência do pescoço e braços do usuário mimificados pelo robô. Quanto aos comportamentos, o controle do caminhar e a utilização do robô em um cenário para transportar um objeto de um local para outro foram também examinados.

Os resultados dessas avaliações revelam uma precisão de 80%, a qual consideramos satisfatória para o que foi proposto neste projeto.

Por meio da experiência adquirida, discorreremos também sobre as limitações averiguadas na utilização do sensor não-intrusivo juntamente com o robô humanoide.

Discursos quanto à experiência e limitações na utilização dessas abordagens não são consideravelmente encontrados na literatura.

Para usuários não acostumados com tecnologia – muito menos especialistas em robótica – acreditamos que quanto menor a interrupção no nível de conhecimento deles, melhor se revela uma interação social e natural com robôs. Este trabalho materializou uma interação humano-robô por meio de uma interface natural obtendo resultados satisfatórios.

O ser humano possui complexidade nata quanto à sua estrutura corporal, tornando complexa a reprodução desta em um robô humanoide. O estado da arte no hardware de robôs humanoides desta geração ainda precisa evoluir quanto à incorporação do lado humano. Mais representações de articulações e de seus respectivos graus de liberdade necessitam ser materializadas para permitir uma maior aproximação relacionada à anatomia humana. No tocante ao nosso conhecimento, não existe ainda um robô humanoide que tenha tantas junções e graus de liberdade como um ser humano apresenta.

Quanto ao nível de inteligência representado em uma parcela dos robôs, este ainda não permite que robôs sejam completamente autônomos para desempenhar algumas atividades que nós realizamos naturalmente. Desta forma, pode-se tomar vantagem da manipulação desses robôs para resolver tarefas que não se deseja desempenhar, impossíveis ou perigosas para seres humanos.

Para trabalhos futuros, nota-se claramente a necessidade de um sistema de balanço estável e de um módulo de omissão de auto-colisão entre os membros do corpo de um robô para que este possa realizar uma imitação de todo o corpo sem as restrições definidas neste trabalho. O ganho em estabilidade consequentemente inclui um viés no mapeamento das configurações do usuário para o robô, o qual deve ser minimizado para não se distanciar consideravelmente das posturas do usuário.

Outrossim, acreditamos que seria de útil aplicação o rastreamento das junções de um usuário em contextos onde o mesmo não deva ficar restrito ao campo de visão de uma câmera estaticamente posicionada. Em um cenário mais claro: uma pessoa poderia manipular determinados objetos em um ambiente considerando o deslocamento do usuário por meio de todo um cômodo, no qual um robô humanoide em outro cômodo realizaria o mesmo comportamento por mimicar o agente. Sensores intrusivos poderiam suportar esse tipo de cenário, mas não é isso que esperamos de uma interação social dita natural com relação a seres humanos.

O trabalho desenvolvido nesta dissertação foi publicado de forma resumida no artigo [88].

Referências Bibliográficas

- [1] K. Capek, "Rossum's Universal Robots.," ed. República Tcheca, Praga, 1921.
- [2] J. Horáková and J. Kelemen, "Robot: The True Story of the Word."
- [3] A. Houaiss and M. S. Villar, "Dicionário da língua portuguesa," *Rio de janeiro: Objetiva*, p. 734, 2001.
- [4] S. Russell, P. Norvig, and V. D. de Souza, *Inteligência Artificial: tradução da segunda edição*: Elsevier: Campus, 2004.
- [5] W. Burgard, D. Fox, and S. Thrun, "Probabilistic robotics," ed: Cambridge, MA: MIT Press, 2005.
- [6] M. A. Goodrich and A. C. Schultz, "Human-robot interaction: a survey," *Foundations and Trends in Human-Computer Interaction*, vol. 1, pp. 203-275, 2007.
- [7] P. M. Abraham and S. Kenter, "Tik-Tok and the Three Laws of Robotics," *Science Fiction Studies*, pp. 67-80, 1978.
- [8] L. Youkou, "Book of Master Lie," ed, por volta de 500 AC.
- [9] N. Wiener, *God and Golem, inc: a comment on certain points where cybernetics impinges on religion* vol. 42: The MIT Press, 1966.
- [10] Homero, "Ilíada," vol. Livro XVIII, ed, Por volta de 800 AC.
- [11] M. Tzu, "Book of Master Mo," ed, por volta de 400 AC.
- [12] M. E. Rosheim, *Leonardo's lost robots*: Springer Verlag, 2006.
- [13] W. B. Yeats, *The winding stair and other poems*: Macmillan, 1933.
- [14] I. Asimov, *I, robot*: Voyager, 1950.
- [15] J. Vujic, A. Marincic, M. Ercegovac, and B. Milovanovic, "Nikola Tesla: 145 years of visionary ideas," 2001, pp. 323-326 vol. 1.
- [16] R. C. Goertz, "Fundamentals of general-purpose remote manipulators," *Nucleonics (US) Ceased publication*, vol. 10, 1952.
- [17] N. J. Nilsson, "Shakey the robot," DTIC Document, 1984.
- [18] H. P. Moravec, "The Stanford cart and the CMU rover," *Proceedings of the IEEE*, vol. 71, pp. 872-884, 1983.
- [19] J. Vertut, "Evolution of manipulators and robotics," 1985, pp. 506-532.
- [20] M. T. Mason, "Creation Myths: The Beginnings of Robotics Research," *Robotics & Automation Magazine, IEEE*, vol. 19, pp. 72-77, 2012.
- [21] A. Thobbi and W. Sheng, "Imitation learning of hand gestures and its evaluation for humanoid robots," 2010, pp. 60-65.

- [22] F. A. Bertsch and V. V. Hafner, "Real-time dynamic visual gesture recognition in human-robot interaction," 2009, pp. 447-453.
- [23] J. Monceaux, J. Becker, C. Boudier, and A. Mazel, "Demonstration: first steps in emotional expression of the humanoid robot Nao," 2009, pp. 235-236.
- [24] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer vision and image understanding*, vol. 104, pp. 90-126, 2006.
- [25] R. Poppe, "Vision-based human motion analysis: An overview," *Computer vision and image understanding*, vol. 108, pp. 4-18, 2007.
- [26] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern recognition*, vol. 36, pp. 585-601, 2003.
- [27] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol. 81, pp. 231-268, 2001.
- [28] T. Teixeira, G. Dublon, and A. Savvides, "A Survey of Human-Sensing: Methods for Detecting Presence, Count, Location, Track, and Identity," *ACM Computing Surveys*, vol. 5, 2010.
- [29] R. Poppe, "A survey on vision-based human action recognition," *Image and Vision Computing*, vol. 28, pp. 976-990, 2010.
- [30] L. F. Costa and R. Cesar, "Shape classification and analysis: theory and practice," *Image Processing Series, PA Laplants, ed.(Academic, 2009)*, pp. 411-414, 2009.
- [31] B. Ripley, "Pattern recognition and neural networks, 1996," *Cambridge Uni. Press, Cambridge*.
- [32] S. Waldherr, R. Romero, and S. Thrun, "A gesture based interface for human-robot interaction," *Autonomous Robots*, vol. 9, pp. 151-173, 2000.
- [33] R. Stiefelhagen, C. Fugen, R. Giesemann, H. Holzapfel, K. Nickel, and A. Waibel, "Natural human-robot interaction using speech, head pose and gestures," 2004, pp. 2422-2427 vol. 3.
- [34] I. M. Jackin, "Wireless Vision Based Mobile Robot Control Using Hand Gesture Recognition Through Perceptual Color Space," 2010, pp. 95-99.
- [35] J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun, "Towards robotic assistants in nursing homes: Challenges and results," *Robotics and Autonomous Systems*, vol. 42, pp. 271-281, 2003.
- [36] S. Iwasawa, K. Ebihara, J. Ohya, and S. Morishima, "Real-time estimation of human body posture from monocular thermal images," 1997, pp. 15-20.
- [37] A. F. Bobick and A. D. Wilson, "A state-based technique for the summarization and recognition of gesture," 1995, pp. 382-388.
- [38] K. Takahashi, S. Seki, E. Kojima, and R. Oka, "Recognition of dexterous manipulations from time-varying images," 1994, pp. 23-28.
- [39] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 257-286, 1989.
- [40] T. Starner and A. Pentland, "Real-time american sign language recognition from video using hidden markov models," 1995, pp. 265-270.
- [41] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden Markov model," 1992, pp. 379-385.

- [42] M. Brand, N. Oliver, and A. Pentland, "Coupled hidden Markov models for complex action recognition," 1997, pp. 994-999.
- [43] C. Vogler and D. Metaxas, "ASL recognition based on a coupling between HMMs and 3D motion analysis," 1998, pp. 363-369.
- [44] Y. Guo, G. Xu, and S. Tsuji, "Understanding human motion patterns," 1994, pp. 325-329 vol. 2.
- [45] R. Mark, Y. Yaser, and D. Larry, "Human Emotion Recognition From Motion Using A Radial Basis Function Network Architecture," 1994.
- [46] R. C. Goertz and F. Bevilacqua, "A force-reflecting positional servomechanism," *Nucleonics (US) Ceased publication*, vol. 10, 1952.
- [47] R. C. Goertz and W. M. Thompson, "Electronically controlled manipulator," *Nucleonics (US) Ceased publication*, vol. 12, 1954.
- [48] R. C. Goertz, "Mechanical master-slave manipulator," *Nucleonics (US) Ceased publication*, vol. 12, 1954.
- [49] R. C. Goertz, "Manipulators used for handling radioactive materials," *Human factors in technology*, pp. 425-443, 1963.
- [50] S. Lichiardopol, "A Survey on Teleoperation," *Technische Universiteit Eindhoven*, 2007.
- [51] J. Funda and R. Paul, "A symbolic teleoperator interface for time-delayed underwater robot manipulation," 1991, pp. 1526-1533.
- [52] A. M. Madni and A. Freedy, "Intelligent interface for remote supervision and control of underwater manipulation," 1983.
- [53] D. Yoerger and J. J. Slotine, "Supervisory control architecture for underwater teleoperation," 1987, pp. 2068-2073.
- [54] D. Yoerger, J. Newman, and J. J. Slotine, "Supervisory control system for the JASON ROV," *Oceanic Engineering, IEEE Journal of*, vol. 11, pp. 392-400, 1986.
- [55] S. B. Skaar and C. F. Ruoff, *Teleoperation and robotics in space* vol. 161: AIAA (American Institute of Aeronautics & Astronautics), 1994.
- [56] A. Bejczy and Z. Szakaly, "Universal computer control systems (uccs) for space telerobots," 1987, pp. 318-324.
- [57] G. Hirzinger, "The space and telerobotic concepts of the DFVLR rotex," 1987, pp. 443-449.
- [58] G. Hirzinger, B. Brunner, J. Dietrich, and J. Heindl, "Sensor-based space robotics-ROTEX and its telerobotic features," *Robotics and Automation, IEEE Transactions on*, vol. 9, pp. 649-663, 1993.
- [59] G. Hirzinger, J. Heindl, and K. Landzettel, "Predictive and knowledge-based telerobotic control concepts," 1989, pp. 1768-1777 vol. 3.
- [60] T. Imaida, Y. Yokokohji, M. Oda, and T. Yoshikawa, "Ground-space bilateral teleoperation of ETS-VII robot arm by direct bilateral coupling under 7-s time delay condition," *Robotics and Automation, IEEE Transactions on*, vol. 20, pp. 499-511, 2004.
- [61] S. G. Hill and B. Bodt, "A field experiment of autonomous mobility: operator workload for one and two robots," 2007, pp. 169-176.
- [62] R. H. Armour and J. F. V. Vincent, "Rolling in nature and robotics: A review," *Journal of Bionic Engineering*, vol. 3, pp. 195-208, 2006.

- [63] A. R. Lanfranco, A. E. Castellanos, J. P. Desai, and W. C. Meyers, "Robotic surgery: a current perspective," *Annals of Surgery*, vol. 239, p. 14, 2004.
- [64] P. F. Hokayem and M. W. Spong, "Bilateral teleoperation: An historical survey," *Automatica*, vol. 42, pp. 2035-2057, 2006.
- [65] T. B. Sheridan and W. R. Ferrell, "Remote manipulative control with transmission delay," *Human Factors in Electronics, IEEE Transactions on*, pp. 25-29, 1963.
- [66] W. R. Ferrell and U. S. N. A. a. S. Administration, "Remote manipulation with transmission delay," Massachusetts Institute of Technology, Dept. of Mechanical Engineering, 1964.
- [67] W. R. Ferrell and T. B. Sheridan, "Supervisory control of remote manipulation," *Spectrum, IEEE*, vol. 4, pp. 81-88, 1967.
- [68] D. Whitney, "State space models of remote manipulation tasks," *Automatic Control, IEEE Transactions on*, vol. 14, pp. 617-623, 1969.
- [69] C. Desoer and M. Vidyasagar, "Feedback systems: input-output properties, 1975," *Academic, New York*.
- [70] P. G. De Barros and R. W. Lindeman, "A Survey of User Interfaces for Robot Teleoperation," 2008.
- [71] G. Cheng and Y. Kuniyoshi, "Real-time mimicking of human body motion by a humanoid robot," 2000, pp. 273-280.
- [72] M. Riley, A. Ude, K. Wade, and C. G. Atkeson, "Enabling real-time full-body imitation: A natural way of transferring human movement to humanoids," 2003, pp. 2368-2374 vol. 2.
- [73] L. Molina-Tanco, J. Bandera, R. Marfil, and F. Sandoval, "Real-time human motion analysis for human-robot interaction," 2005, pp. 1402-1407.
- [74] S. Nakaoka, A. Nakazawa, K. Yokoi, H. Hirukawa, and K. Ikeuchi, "Generating whole body motions for a biped humanoid robot from captured human dances," in *Robotics and Automation, 2003. Proceedings. ICRA '03. IEEE International Conference on*, 2003, pp. 3905-3910 vol.3.
- [75] C. Ott, D. Lee, and Y. Nakamura, "Motion capture based human motion recognition and imitation by direct marker control," 2008, pp. 399-405.
- [76] H. B. Suay and S. Chernova, "Humanoid robot control using depth camera," presented at the Proceedings of the 6th international conference on Human-robot interaction, Lausanne, Switzerland, 2011.
- [77] J. Koenemann and M. Bennewitz, "Whole-body imitation of human motions with a nao humanoid," presented at the Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, Boston, Massachusetts, USA, 2012.
- [78] T. Veltrop. (2012, June). Available: <http://taylor.veltrop.com/>
- [79] F. Dunn and I. Parberry, *3D math primer for graphics and game development*: AK Peters Ltd, 2011.
- [80] (2012). *Kinect for Windows*. Available: <http://www.microsoft.com/en-us/kinectforwindows/>
- [81] (2012). *OpenNI Framework*. Available: <http://www.openni.org/>
- [82] (2012). *OpenKinect Library*. Available: <http://openkinect.org/>

- [83] D. Gouaillier, V. Hugel, P. Blazevic, C. Kilner, J. Monceaux, P. Lafourcade, *et al.*, "The nao humanoid: a combination of performance and affordability," *CoRR*, vol. *abs/0807.3223*, 2008.
- [84] A. Maciel, "Modelagem de articulações para humanos virtuais baseada em anatomia," 2001.
- [85] I. Rec, "P. 800: Methods for subjective determination of transmission quality," *International Telecommunication Union, Geneva*, 1996.
- [86] J. Saunders, C. L. Nehaniv, and K. Dautenhahn, "An experimental comparison of imitation paradigms used in social robotics," 2004, pp. 691-696.
- [87] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*: O'Reilly Media, 2008.
- [88] F. Zuher and R. Romero, "Recognition of Human Motions for Imitation and Control of a Humanoid Robot," in *Robotics Symposium and Latin American Robotics Symposium (SBR-LARS), 2012 Brazilian*, 2012, pp. 190-195.