Robôs Socialmente Assistivos em Ambientes Públicos: desafios em plataforma, arquitetura e sistemas de percepção

Pedro Victor Vieira de Paiva (CTI) pvpaiva@cti.gov.br

Resumo

Em níveis elevados de automação, sistemas robóticos devem cooperar com humanos de forma interativa. Tal tarefa requer um alto nível de percepção, o que se traduz em elevada demanda computacional. Outros desafios para a cooperação humano-robô surgem com a proximidade exigida nessas atividades. A orientação do corpo humano é uma informação valiosa para robôs sociais devido ao seu uso no planejamento de caminhos. Todos esses desafios se ampliam quando consideramos ambientes não preparados para robôs, com características dinâmica e não-estruturadas. Uma maneira de superar as limitações de percepção dos agentes robóticos é movendo recursos de processamento para unidades externas. Uma vez superado o desafio de poder computacional, técnicas avançadas de aprendizado de máquina podem ser aplicados nas tarefas de percepção. Sendo assim, apresentamos nesse artigo soluções desenvolvidas para um sistema robótico baseado em nuvem capaz de reconhecer pessoas, estimar orientação de corpos, bem como suporte para comunicação não verbal. Tal sistema contribuirá no desenvolvimento de ajudantes robóticos para diversas aplicações como atendimento ao público, exploração de ambientes, entre outros. São apresentados estudos de caso que exemplificam a integração entre o agente e a arquitetura, comparativo entre o método de estimação de orientação proposto e o estado da arte, além de resultados simulados do mesmo.

Palavras-chave*: Interação Humano-Robô, Robótica em Nuvem, Aprendizado Profundo, Visão Computacional.*

1. Introdução

Uma das principais motivações da Interação Humano-Robô (Goodrich, 2008) é entender e modelar interações entre um ou mais humanos com um ou mais robôs. Entre os atributos que devem reger essa interação estão a autonomia e a adaptabilidade dos sistemas robóticos. A autonomia pode ser definida como capacidade de receber estímulos do ambiente e realizar ações de acordo com essas entradas. Robôs com elevada autonomia podem ser programados com consciência social para realizar várias tarefas. Robótica social é uma área muito pesquisada recentemente, entretanto, apenas uma pequena parte dos trabalhos na área consideram cenários públicos da vida real, ou como é conhecido no campo da robótica, em *in-the-wild*. Isto é um problema desafiador porque os seguintes aspectos não são considerados: (i) a dificuldade na identificação de objetos no espaço, especialmente pessoas e (ii) o tempo de resposta do robô para uma ação executada próxima de humanos.

Dentro da Robótica Socialmente Interativa (RSI) (Fong et al., 2003) busca-se o estabelecimento de conhecimento e habilidades para permitir a interação entre humanos e robôs, respeitando as regras de convivência. Tarefas como reconhecer uma pessoa, identificar objetos, aprender e localização e mapeamento, são desejáveis para um robô socialmente interativo, mas esbarram na capacidade limitada de processamento dos agentes robóticos. Uma alternativa para o problema de baixo poder computacional são as arquiteturas baseadas em nuvem, que demonstram elevado potencial em robótica (Tian *et al.*, 2018). Seguindo a

ideia de uma plataforma distribuída e com capacidades de percepção em nuvem, apresentamos ROSANA (*RObot for Social interAction in uNstructured dynAmic environments*), uma arquitetura robótica capaz de implementar elevado nível de automação graças a sua capacidade de integrar diferentes estratégias de percepção. Esta abordagem baseia-se em agente robótico móvel capacitado com processamento em nuvem.

2. Objetivo

Este trabalho visa contribuir para as iniciativas associadas à Indústria 5.0, que no caso do NRVC-CTI são associadas por ações ligadas à combinação das forças do ser humano e do robô para uma manufatura mais eficiente. Temos como objetivo abordar os desafios associados à maior proximidade entre pessoas e robôs com o uso de sensores e sistemas de interação que propiciem essa maior interatividade, aplicados a ambientes industriais, residenciais e públicos. Neste trabalho destacamos as seguintes contribuições feitas seguindo os objetivos acima:

- Construção de uma arquitetura robótica distribuída;
- Experimentos de integração da plataforma proposta com técnicas de reconhecimento;
- Proposta de expansão das capacidades de percepção por meio de aprendizado da orientação de corpos humanos;
- Comparativo com o estado da arte e estudos de caso e experimentos em simulação do método proposto para estimação de orientação;

3. Materiais e Métodos

ROSANA é uma plataforma social interativa conforme definido por (Fong et al., 2003), capaz de perceber informações do meio ambiente. Para tanto, se comunica com a nuvem, a fim de aumentar sua capacidade de lidar com informações, reduzindo os requisitos de processamento a bordo do agente móvel. Nesta seção, os principais componentes da arquitetura ROSANA são apresentados, bem como as contribuições geradas nas áreas de plataforma e arquitetura robótica, percepção e visão computacional.

3.1 Contribuições em Plataforma e Arquitetura Robótica em Nuvem

Robótica em nuvem é definida como qualquer sistema de automação que depende de dados ou código de uma rede para suportar sua operação. Neste tipo de sistema, computação e memória são distribuídos, não centralizados em um único sistema autônomo (Kuffner, 2010). Existem várias vantagens em sistemas em nuvem, como: computação paralela sob demanda, compartilhamento de sistemas trajetórias, etc. Podemos enumerar os seguintes modelos de serviço para robótica em nuvem: infraestrutura como serviço (IaaS), plataforma como um serviço (PaaS), software como serviço (SaaS) e Robô como serviço (RaaS) (Chen *et al., 2010*). Tais estratégias levam a ganhos de desempenho em espaços estáticos, mas o desempenho degrada rapidamente em ambientes não estruturados devido a latência e outros gargalos de rede (Saha & Dasgupta, 2018). Para enfrentar o desafio da na robótica *in the wild*, propomos uma arquitetura de percepção local de alta velocidade chamada PPaaS. Percepção como Serviço (PPaaS) fornece serviços de reconhecimento em imagens RGB/RGB-D em tempo real empregando classificadores de imagem baseados em redes neurais convolucionais. Uma das vantagens do PPaaS é a possibilidade de suplantar sistemas robóticos móveis simples com poderosos sistemas de reconhecimento, possíveis apenas com GPUs e CPUs potentes. Como de costume em arquiteturas de robótica em nuvem (Saha & Dasgupta, 2018), PPaaS é dividido em dois componentes: infraestrutura em nuvem e sua base móvel. Conforme ilustrado na Figura 1, arquitetura ROSANA implementa o modelo PPaaS

fornecendo ao sistema móvel serviços de percepção rodando no módulo de nuvem. Um elemento essencial do ROSANA é a comunicação full-duplex (ver Figura 1). O sistema móvel adquire informações do ambiente de usando um sensor RGB-D, registra localmente dados de profundidade e usa Ethernet Wifi 5.0 Ghz para transmitir imagens para a nuvem. Como enfatizado anteriormente, o gargalo da rede é uma constante limitação para robótica em nuvem. Este problema foi eliminado implementando uma comunicação de soquete confiável baseada em TCP/IP (TCP NODELAY) (Choi *et al.*, 2011). A nuvem é capaz de receber transmissões múltiplas e usar várias GPUs e arquiteturas de aprendizagem profunda para processar e prever diferentes padrões visuais simultaneamente.

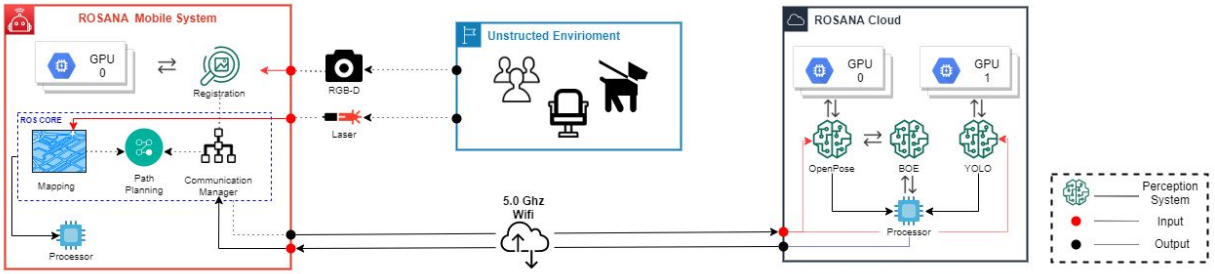


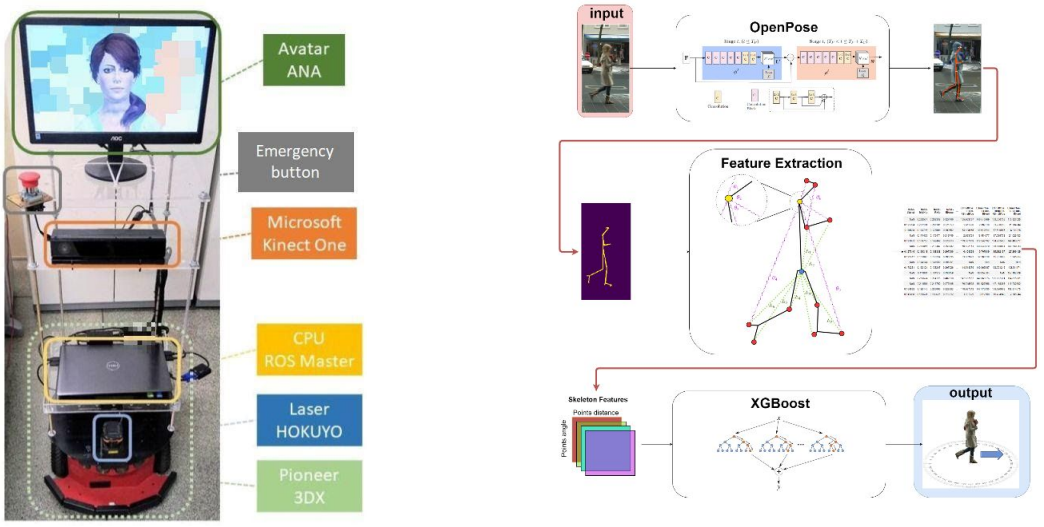
Figura 1 –Abstração em alto nível da arquitetura de robótica em nuvem ROSANA e das interações entre componentes móveis, ambiente e serviços distribuídos de percepção.

3.2 Contribuições em Percepção de Agentes Móveis

Em robótica, percepção pode ser definida como a capacidade de compreender e reagir ao ambiente circundante (Russell & Stuart, 2003). Percepção visual para sistemas robóticos tem sido o centro de atenção e em sendo utilizado em muitas plataformas. O sistema de percepção visual da ROSANA está sendo desenvolvido para superar uma limitação comum para robôs móveis: combinar um único sensor RGB/RGB-D com desempenho de percepção robusta. Um único sensor visual é conveniente para robôs em termos de consumo de energia e processamento. Embora, contar com apenas uma câmera é incomum no estado da arte (Sheridan, 2016) de arquiteturas de interação humano-robô devido à falta de informação que os processadores on-board podem extrair dessas imagens. Com o advento do aprendizado profundo, as câmeras monoculares podem detectar muitos objetos ao custo de processadores gráficos poderosos. Uma vez que a ROSANA possui hardware de computador robusto em seu lado servidor, técnicas robustas de reconhecimento são incorporadas como: (i) detecção de objetos em tempo real usando R-CNN YOLOv3 (Redmon *et al.*, 2016); (ii) detecção de corpo humano com campos de afinidade parcial OpenPose (Cao *et al.*, 2018); (iii) estimação da orientação do corpo humano com XGBoost (Paiva *et al.*, 2020).

3.3 Contribuições em Estimação de Orientação Corporal

Uma vez estabelecida a arquitetura e plataforma capaz de ser equipada com técnicas avançadas de percepção, propomos atacar um problema comum em robótica social: identificar corretamente a orientação espacial de pessoas em ambientes diversos, também conhecido como *Body Orientation Estimation* (BOE). Nossa abordagem é baseada na hipótese de que o esqueleto de saída do OpenPose (Cao *et al.*, 2018) é um conjunto de dados suficientes para essa estimação.Nossa abordagem está resumida na Figura 2 (b), podemos encapsular o método em 3 etapas: (a) obtenção de esqueletos 2D de imagens usando o OpenPose; (b) cálculo de ângulos e distâncias entre partes do corpo; e (c) treinamento de um classificador de gradiente extremo para inferir orientação.A rede OpenPose tem como saída um esqueleto formado por 25 pontos corporais. Distâncias e ângulos entre esses pontos são calculados e usados como características em uma fase posterior de treinamento.



(a) (b)

Figura 2 – Ilustração dos componentes de hardware e software: (a) Agente robótico móvel ROSANA e seus principais componentes físicos; (b) Diagrama geral da abordagem proposta para BOE.

O número de combinações possíveis entre pontos é dado por uma equação padrão da teoria dos conjuntos, conforme apresentada na Equação 1 (Zwillinger, 2002). Algo que deve ser considerado quando a distância é calculada é o sistema de coordenadas e sua faixa de valores. A distância de Bray-Curtis (Bray & Curtis, 1957) é um sistema de coordenadas normalizado para valores positivos, variaram entre 0 e 1, Equação 2. Projetando os ângulos naturais para um $\mathbb{R}^2$ espaço, como em imagens RGB, a triangulação vetorial 2D uma escolha óbvia para calcular ângulos entre 3 pontos. Para três vetores A, B e C $\in \mathbb{R}^2$ dados por OpenPose como coordenadas x e y de pixel, o ângulo interno θ entre esses vetores é dado pela Equação 3.

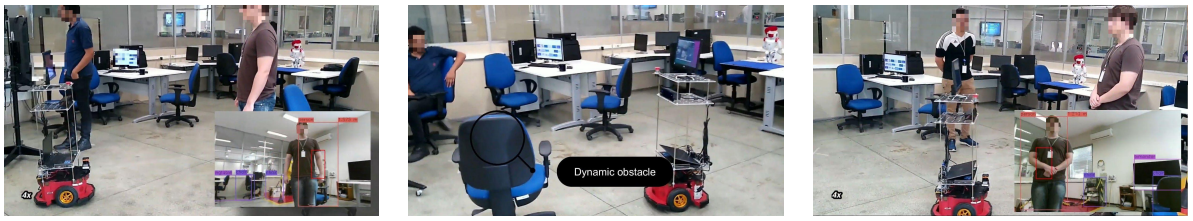
$$C(n,r) = \frac{n!}{r!(n-r)!} \quad (1) \qquad d_{i,j} = \frac{\sum_{k=1}^{n} |x_{ik} - x_{jk}|}{\sum_{k=1}^{n} (x_{ik} + x_{jk})} \quad (2) \qquad \theta = \arccos\left(\frac{\vec{BA} \cdot \vec{BC}}{\vec{BA}\,\vec{BC}}\right) \quad (3)$$

Quando todas as distâncias e ângulos possíveis são calculados, seus valores são usados como atributos para um algoritmo de aprendizado supervisionado. O algoritmo Extreme Gradient Boosting (XGBoost) é usado para classificação. Um elemento chave para a adoção do XGBoost é a sua capacidade para lidar com valores ausentes. No mundo real, existem casos em que partes do corpo estão ausentes, XGBoost é uma otimização de árvores aleatórias que define a direção dos nós para valores ausentes com base no ganho máximo para a esquerda ou direita dos subnós, escolhendo aquele com o maior ganho (Chen & Guestrin, 2016).

4. Experimentos e Resultados

Para testar as capacidades do ROSANA em navegação e reconhecimento, alguns experimentos preliminares foram feitos. Esses resultados podem ser visto no seguinte link: https://youtu.be/AXiI_tEp_cA. O servidor ROSANA nuvem é composto por 2x Intel Xeon E5-2620 6 núcleos 2,5 Ghz 12 threads, 96GB de RAM e 2x Nvidia TITAN X 12GB. O robô que executa o lado móvel ROSANA consiste de um notebook Dell Notebook Vostro 5480 com Intel i7 2,4 GHz, 8 GB de RAM e Nvidia 820M. No primeiro experimento de navegação

(Figura 3-a), o robô deve seguir uma pessoa detectada. A identificação é feita por meio de imagens capturadas por um Microsoft Kinect ONE e usando o serviço de percepção ROSANA nuvem. O robô segue a pessoa com sucesso quando está suficientemente distante e se move para trás quando a pessoa chega muito perto, como ilustrado na Figura 3-a. O segundo experimento apresenta um obstáculo inesperado entre o robô e a pessoa, destacado na Figura 3-b; ele testa se a prevenção de obstáculos por campos potenciais funciona corretamente e se a oclusão parcial de uma pessoa afeta o comportamento. O terceiro experimento testa como o robô se comporta quando mais de uma pessoa é detectada. Duas pessoas se revezam ficando na frente do robô para verificar se o robô altera a pessoa que ele está seguindo, visto na Figura 3-c. Os estudos de caso realizados comprovam que o ROSANA pode concluir atividades complexas em ambientes dinâmicos e não estruturados.



(a) (b) (c)

Figura 3 –Experimentos para os seguintes estudos de caso (a) seguir um humano no ambiente; (b) seguir o humano enquanto evita obstáculos do ambiente e (c) escolher um dos humanos para seguir.

Para avaliar nossa abordagem de estimação de orientação comparamos nosso método com outro abordagens existentes usando o mesmo conjunto de dados, dividido em treinamento/teste e usando métricas de avaliação. Testamos nosso método no dataset TUD Multiview Pedestrians (Andriluka et al., 2010), contendo 4730 imagens para treinamento, 248 para validação e 248 para teste. Comparando nosso resultado com outros resultados da literatura, (apresentado na Tabela I), pode-se ver que nossa abordagem teve melhor desempenho em previsões exatas, isto é, sem considerar intersecção de intervalos. É importante destacar que a previsão exata é o mais utilizável para muitos aplicativos de orientação. Um estudo detalhado sobre esta comparação foi publicado em Paiva *et al.* (2020).

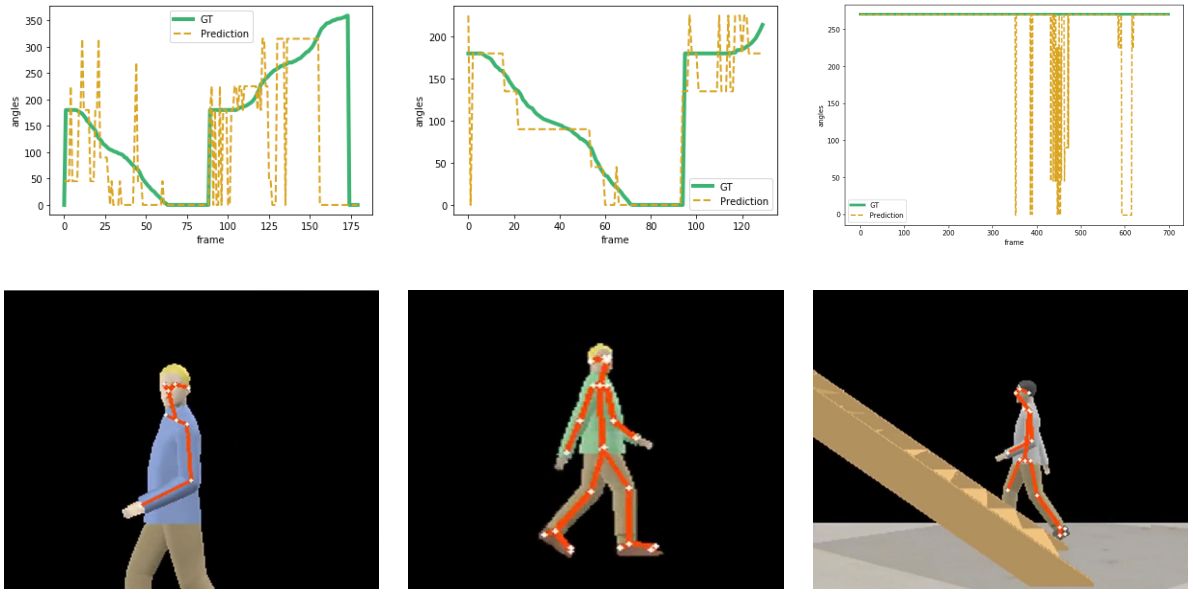
Para demonstrar como a abordagem proposta é aplicável em um robô, experimentos simulados foram conduzidos. Os experimentos foram simulados por dois motivos: a flexibilidade dada pelo ambiente de simulação e devido ao surto COVID-19. Três cenários são apresentados: (i) parte inferior do corpo ocluída, (ii) visão de corpo inteiro, e (iii) oclusão parcial / total corpo. No primeiro e segundo casos, o modelo humano faz uma caminhada de retorno de um ponto fixo e no terceira é um caminho reto com obstáculos. Um vídeo cobrindo o estudo de caso em todos cenários está disponível em: https://vimeo.com/447311206.

Método	Acurácia	
	0° (exato)	±45° (não-exato)
Abordagem proposta	**82.6%**	**83.9%**
MoAWG	67.4%	*
PLS-RF	66.3%	*
HOG+LRC	57.9%	83.7%
HOG+SVM+PCA	53.2%	78.8%
ERT+MoAWG	53.0%	81.5%

*valores não disponíveis nos trabalhos de referência

Tabela 1 – Comparativo entre a acurácia média alcançado e demais métodos na literatura.

O caso (i) reflete situações onde o robô está perto do humano, o que geralmente compromete a percepção visual de partes inferiores do corpo. Neste cenário, apresentado na Figura 4-a, nosso método sofre com a falta de informações do esqueleto. No entanto, mesmo com a indisponibilidade de dados, nossa abordagem pode prever a tendência de variação, conforme visto no gráfico. Contrastante com o Caso (i), o Caso (ii) apresenta um cenário ótimo onde o corpo inteiro é visível. Neste caso, nossos arquivos de método um bom desempenho como visto em 4-b. Finalmente, o Caso (iii) simula uma caminhada lado a lado com oclusão ocasional ocorrendo. A abordagem proposta foi capaz de lidar com desafios obstrução, conforme mostrado no vídeo vinculado. Neste caso, 4-c prova que na maioria das vezes, os valores preditos e os reais.



(a) (b) (c)

Figura 4 – Comparativo de performance entre valores preditos e padrão-ouro em ambiente simulados: (a) membros inferiores ocluídos, (b) corpo totalmente visível e (c) obstruções parciais.

5. Conclusão

Neste artigo, uma visão geral do robô social ROSANA é apresentado, mostrando sua arquitetura, componentes de nuvem, e estratégias de percepção baseado em visão computacional. Algumas provas de conceito foram conduzidas e discutidas a fim de avaliar as capacidades reais da ROSANA e estado de desenvolvimento. Experimentos demonstraram a validade de nossa técnica de estimação de orientação e seu melhor desempenho em comparação com outros métodos da literatura. Como trabalhos futuro, pretendemos fazer um ciclo completo de interação social da ROSANA com uma pessoa, incluindo interação verbal com um agente virtual. Também há trabalhos em andamento na arquitetura da nuvem para alcançar resultados mais rápidos e em reconhecer informações não verbais. Os métodos e a arquitetura proposta durante este período inicial do programa PCI foram detalhados em 2 publicações aceitas na conferência internacional IEEE LARS 2020 - 17th Latin American Robotics Symposium.

Referências

ANDRILUKA, Mykhaylo; ROTH, Stefan; SCHIELE, Bernt. *Monocular 3d pose estimation and tracking by detection.* In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010. p. 623-630.

BRAY, J. Roger; CURTIS, John T. *An ordination of the upland forest communities of southern Wisconsin.* Ecological monographs, v. 27, n. 4, p. 326-349, 1957.

CAO, Zhe et al. *Realtime multi-person 2d pose estimation using part affinity fields*. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 7291-7299.

CHEN, Tianqi; GUESTRIN, Carlos. *Xgboost: A scalable tree boosting system*. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016. p. 785-794.

CHEN, Yinong; DU, Zhihui; GARCÍA-ACOSTA, Marcos. *Robot as a service in cloud computing*. In: 2010 Fifth IEEE International Symposium on Service Oriented System Engineering. IEEE, 2010. p. 151-158.

CHOI, Ki-Bong; KIM, Jong-Hyune; YEO, In-Dong. *Apparatus and method for managing delay for TCP/IP communications in a mobile communication system.* U.S. Patent n. 8,027,307, 27 set. 2011.

FONG, Terrence; NOURBAKHSH, Illah; DAUTENHAHN, Kerstin. *A survey of socially interactive robots.* Robotics and autonomous systems, v. 42, n. 3-4, p. 143-166, 2003.

GOODRICH, Michael A. et al. *Human–robot interaction: a survey*. Foundations and Trends® in Human–Computer Interaction, v. 1, n. 3, p. 203-275, 2008.

KUFFNER, James. *Cloud-enabled humanoid robots*. In: Humanoid Robots (Humanoids), 2010 10th IEEE-RAS International Conference on, Nashville TN, United States, Dec. 2010.

PAIVA, Pedro V. V.; BATISTA, Murillo R.; RAMOS, Josué J. G. *Estimating human body orientation using skeletons and Extreme Gradient Boosting.* In: 2020 Latin American Robotics Symposium (LARS). IEEE, 2020.

REDMON, Joseph et al. *You only look once: Unified, real-time object detection.* In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 779-788.

RUSSELL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 2002.

SAHA, Olimpiya; DASGUPTA, Prithviraj. *A comprehensive survey of recent trends in cloud robotics architectures and applications*. Robotics, v. 7, n. 3, p. 47, 2018.

SHERIDAN, Thomas B. *Human–robot interaction: status and challenges*. Human factors, v. 58, n. 4, p. 525-532, 2016.

TIAN, Nan et al. *A cloud-based robust semaphore mirroring system for social robots*. In: 2018 IEEE 14th International Conference on Automation Science and Engineering (CASE). IEEE, 2018. p. 1351-1358.

ZWILLINGER, Daniel. *CRC standard mathematical tables and formulae*. CRC press, 2002.