



UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Faculdade de Engenharia Elétrica

Programa de Pós-Graduação em Engenharia Biomédica

**Estratégia de Alcance para Próteses de Membro Superior  
Baseada em Visão Neuromórfica**

Eduardo Borges Gouveia

Uberlândia

2022

Eduardo Borges Gouveia

# **Estratégia de Alcance para Próteses de Membro Superior Baseada em Visão Neuromórfica**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Biomédica da Faculdade de Engenharia Elétrica da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciências.

Área de concentração: Engenharia Biomédica

Linha de Pesquisa: Engenharia de reabilitação e tecnologias assistivas

**Orientador:** Prof. Dr. Alcimar Barbosa Soares.

**Coorientador:** Prof. Dr. Edgard Afonso Lamounier Júnior.

**Coorientador:** Prof. Dr. Márcio José da Cunha.

Uberlândia

2022

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU  
com dados informados pelo(a) próprio(a) autor(a).

G719 2022	<p>Gouveia, Eduardo Borges, 1996- Estratégia de Alcance para Próteses de Membro Superior Baseada em Visão Neuromórfica [recurso eletrônico] / Eduardo Borges Gouveia. - 2022.</p> <p>Orientador: Alcimar Barbosa Soares. Coorientador: Edgard Afonso Lamounier Júnior. Coorientador: Márcio José da Cunha. Dissertação (Mestrado) - Universidade Federal de Uberlândia, Pós-graduação em Engenharia Biomédica. Modo de acesso: Internet. Disponível em: <a href="http://doi.org/10.14393/ufu.di.2022.106">http://doi.org/10.14393/ufu.di.2022.106</a> Inclui bibliografia. Inclui ilustrações.</p> <p>1. Engenharia biomédica. I. Soares, Alcimar Barbosa, 1965-, (Orient.). II. Júnior, Edgard Afonso Lamounier , 1964-, (Coorient.). III. Cunha, Márcio José da, 1978-, (Coorient.). IV. Universidade Federal de Uberlândia. Pós-graduação em Engenharia Biomédica. V. Título.</p> <p style="text-align: right;">CDU: 62:61</p>
--------------	---

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:



**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**  
 Coordenação do Programa de Pós-Graduação em Engenharia Biomédica  
 Av. João Naves de Ávila, 2121, Bloco 3N, Sala 115 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902  
 Telefone: (34) 3239-4761 - www.ppggeb.feelt.ufu.br - ppegb@feelt.ufu.br



### ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Engenharia Biomédica				
Defesa de:	Dissertação de Mestrado Acadêmico, 079, PPGEB				
Data:	vingte e cinco de fevereiro de dois mil e vinte e dois	Hora de início:	09:00	Hora de encerramento:	11:50
Matrícula do Discente:	11922EBI005				
Nome do Discente:	Eduardo Borges Gouveia				
Título do Trabalho:	Estratégia de Alcance para Próteses de Membro Superior Baseada em Visão Neuromórfica				
Área de concentração:	Engenharia Biomédica				
Linha de pesquisa:	Engenharia de Reabilitação e Tecnologias Assistivas				
Projeto de Pesquisa de vinculação:	Neuromorphic Feedback. A Strategy to Enhance Prosthesis Embodiment and Performance				

Reuniu-se em web conferência pela plataforma Conferência Web em conformidade com a PORTARIA Nº36, DE 19 DE MARÇO DE 2020 da COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR - CAPES, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Engenharia Biomédica, assim composta: Professores Doutores: Edgard Afonso Lamounier Júnior - PPGEB/UFU; Márcio José da Cunha - FEELT/UFU; Ailton Luiz Dias Siqueira Júnior - IFTM; Alcimar Barbosa Soares - PPGEB/UFU orientador(a) do(a) candidato(a).

Iniciando os trabalhos o presidente da mesa, Dr. Alcimar Barbosa Soares, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Edgard Afonso Lamounier Junior, Professor(a) do Magistério Superior**, em 25/02/2022, às 11:53, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Alcimar Barbosa Soares, Professor(a) do Magistério Superior**, em 25/02/2022, às 11:53, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Ailton Luiz Dias Siqueira Junior, Usuário Externo**, em 25/02/2022, às 11:53, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Marcio José da Cunha, Professor(a) do Magistério Superior**, em 25/02/2022, às 11:53, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://www.sei.ufu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **3403326** e o código CRC **403A7659**.

## Agradecimentos

*“Every path is the right path. Everything could have been anything else and it would have just as much meaning”*, foi o que o dramaturgo, conhecido como Tennessee Williams, disse acerca da busca pelas escolhas significativas. A finitude do ser, a unidirecionalidade do tempo e a imprevisibilidade do futuro contribuem para que todo caminho seja o caminho certo. *“Every path is the right path”*. Um apelo para que a jornada seja apreciada em detrimento da preocupação com o destino, ou a chegada.

Quando penso na minha trajetória até chegar à entrega dessa dissertação penso que fazer o mestrado significa mais sobre se tornar mestre do que o produto desse processo. Trilhar o caminho do que chegar ao destino. Nesse sentido, não poderia deixar de reconhecer os aspectos dessa trajetória que fizeram o caminho valer a pena.

Uma constante, sempre presente, com influência majoritária, não apenas durante o mestrado, mas durante toda a minha vida, tem sido a minha mãe Vanda e, para ela, mulher resiliente e doada, eu só poderia deixar os mais especiais agradecimentos. Por todos os caminhos que a senhora deixou de percorrer para trilhar o caminho cujo produto sou eu, minha mais sincera e profunda gratidão. É desejo do meu coração que meu caminho de vida produza tanto amor quanto o que a senhora sempre dispôs na sua trajetória.

Ao professor Alcimar, por todas as vezes que o senhor viu em mim muito mais do que eu jamais vi. Agradeço por todos os momentos, mesmo antes de ingressar no mestrado, que tive a oportunidade de trabalhar com o senhor e pude aprender contigo. Se “é junto dos bão que nois fica mió” eu tive o prazer e privilégio de estar próximo de uns dos melhores, você.

Aos professores Edgard e Márcio que, de um jeito leve e sempre descontraído, me coorientaram. Agradeço cada ensinamento e conversa que nem sempre se referia ao trabalho desenvolvido nessa dissertação, mas sempre era carregado de profundos ensinamentos.

Aos meus companheiros de laboratório que, se um dia foram apenas colegas ou algo menos que queridos amigos, não me recordo: Andrei, Amanda, Eber, Vinícius e Ronaldo que sempre estiveram presentes no caminho. Sei que vocês, em algum momento, foram para outras partes do mundo conquistar seus respectivos lugares ao sol e seguir cada qual com sua própria trilha,

mas saibam que, se os afetos se assemelham a uma harpa, minhas cordas continuarão a vibrar por vocês, mesmo que o dia a dia, que serviu de dedilhar para essas cordas, já seja coisa do passado.

À Caroline, que se fez exemplo de dedicação e comprometimento. Com seu incentivo, influência e afeto o caminho se tornou mais agradável e o destino, qualquer que seja, se tornou símbolo de felicidade. Se somos a soma das pessoas que escolhemos para estar ao nosso lado, eu fico grato, por ter um pouco de você em mim. Agradeço por você e agradeço por ter vindo do passado, para voltar a existir no meu presente.

Por fim, agradeço à Universidade Federal de Uberlândia, a todos os servidores e professores da Faculdade de Engenharia Biomédica e Elétrica por toda a contribuição de forma direta ou indireta no caminho percorrido. Meus mais sinceros agradecimentos.

## Resumo

A perda de um membro do corpo humano tem um impacto devastador na autonomia do indivíduo de realizar tarefas. Os impactos sociais, emocionais e físicos relacionados a perda de um membro são extremamente abrangentes, principalmente em casos de perda de membros superiores. A mão humana é interface de utilização de ferramentas, e permite interações sociais e afetivas. O meio usual de recobrar uma parte das funcionalidades perdidas com a amputação é a utilização de próteses. Entretanto, as próteses de membro superior comerciais mais sofisticadas apresentam empecilhos de usabilidade, seja por dificuldade de controlar todos os graus de liberdade ou mesmo a falta de graus de liberdade em modelos simples. Esses dispositivos nem sempre atendem as expectativas dos usuários de recobrar sua autonomia. Diversas abordagens visam desenvolver mecanismos e sistemas que possuem o intuito de aprimorar a usabilidade e devolver a funcionalidade do membro perdido, melhorando as interfaces de controle. O trabalho desenvolvido por essa dissertação propõe um modelo semiautônomo de alcance para próteses de membro superior, baseado em visão neuromórfica, para indivíduos que sofreram amputações acima do nível do cotovelo. A utilização de sistemas neuromórficos de visão é uma abordagem interessante por ser uma ferramenta com alta eficiência energética e computacional, porém seu funcionamento baseado em eventos pressupõe novas abordagens computacionais. Esse trabalho contribui para o desenvolvimento de novas abordagens e algoritmos para sistemas neuromórficos de visão computacional, revelando-se uma interface interessante para a implementação em dispositivos embarcados onde poder computacional e disponibilidade energética são escassos. O modelo desenvolvido por esse trabalho foi validado utilizando uma câmera baseada em eventos (Dynamic Vision Sensor 128) e um braço robótico (WidowX) que, aliados a um sistema de detecção e rastreamento de objetos, desempenha sua tarefa de alcance em cinco experimentos que apresentam ambientes desafiadores para um sistema com esse intuito. Os resultados obtidos utilizando o sistema semiautônomo de alcance para próteses de membro superior, sugerem que sistemas neuromórficos podem guiar o desenvolvimento de dispositivos inteligentes de controle com baixo consumo de energia e alto desempenho computacional.

**Palavras-chave:** Sistema de controle semiautônomo, controle de próteses, visão neuromórfica.

## **Abstract**

The loss of a limb has a devastating impact on the individual's autonomy to perform day-to-day tasks. The social, emotional, and physical impacts related to it are extremely hard, especially in cases of loss of upper limbs. The human hand is an interface for using tools and allows social and affective interactions. The usual way of trying to recover some of the functionality lost with amputation is the use of prostheses. However, even the most sophisticated commercial upper limb prostheses present usability obstacles for users, either because of the difficulty of controlling all the degrees of freedom present in sophisticated models or the lack of them in simpler models. The prostheses do not always meet the expectations of users to regain their autonomy. In this way, several approaches aim to develop mechanisms and systems that improve usability and try to give back the functionality of the lost limb by improving the control interfaces of prosthetic arms. The work developed in this dissertation proposes a reach model for upper limb prostheses based on neuromorphic vision, intending to implement a semi-autonomous model for prosthetic devices for individuals who have suffered amputations above the elbow level. The use of neuromorphic vision systems is a good match to embedded systems because of their power and computational efficiency. However, an event-based system requires new computational approach to deal with event-data. Therefore, contributing to the development of new approaches and algorithms for neuromorphic computer vision systems and proving to be an interesting interface for implementation in embedded devices whose parameters of computational requirements and available power are scarce. The model developed by this work was validated using an event-based camera (Dynamic Vision Sensor 128) and a commercial robotic arm (Widow X) which, combined with object detection and tracking system, performs reaching objects task in five experiments that present challenging environments for a system with this purpose. The results obtained using the semi-autonomous model for upper limb prostheses suggest that neuromorphic systems can guide the development of intelligent control devices with low energy consumption and high computational performance.

**Keywords:** Semi-autonomous control system, control of prostheses, neuromorphic vision.

## Sumário

1	Introdução .....	14
1.1	Justificativa .....	18
1.2	Objetivos .....	19
1.3	Estrutura da dissertação .....	19
2	Fundamentação Teórica.....	21
2.1	Próteses de membros superiores .....	21
2.1.1	Sistemas de automatização de próteses .....	23
2.2	Sistemas neuromórficos .....	24
2.2.1	Câmera baseada em eventos .....	26
2.3	Visão Neuromórfica .....	29
3	Metodologia.....	32
3.1	Proposta do modelo de alcance para próteses de membro superior .....	32
3.2	Modelo de detecção de objetos .....	35
3.2.1	Conversão de informações esparsas em informações densas.....	35
3.2.2	Arquitetura do modelo .....	36
3.2.3	Conjunto de dados .....	39
3.3	Sistema de rastreamento em tempo real .....	41
3.4	Procedimentos experimentais de avaliação e validação .....	42
3.4.1	Experimento 1 .....	44
3.4.2	Experimento 2 .....	45
3.4.3	Experimento 3 .....	46
3.4.4	Experimento 4 .....	47
3.4.5	Experimento 5 .....	48
3.4.6	Análise de dados.....	49
4	Resultados e discussão .....	53
4.1	Conjunto de dados resultante.....	53
4.2	Modelo de detecção .....	57
4.3	Análise do desempenho do modelo de rastreamento em tempo real.....	63
4.4	Considerações finais .....	73
5	Conclusões.....	75
6	Apêndice .....	77
6.1	Lista de artigos publicados.....	77
7	Referências.....	78

## Lista de Figuras

Figura 1: <b>Representação dos eventos de uma gravação utilizando uma câmera baseada em eventos.</b>	16
Figura 2: <b>Representação do desafio de associação de dados.</b>	17
Figura 3: <b>Representação hierárquica das classificações dos dispositivos protéticos comerciais.</b>	22
Figura 4: <b>Imagem dos três modelos comerciais de próteses de mão, tecnologicamente avançadas.</b>	23
Figura 5: <b>Representação da resposta elétrica da membrana de uma célula excitável neuronal.</b>	26
Figura 6: <b>Representação da diferença de saída entre uma câmera tradicional (baseada em frames) e uma câmera neuromórfica (baseada em eventos).</b>	28
Figura 7: <b>Ilustração do processo de conversão da luminosidade que atinge cada fotodiodo (<i>pixel</i>) em eventos de variação de luminosidade.</b>	29
Figura 8: <b>Modelo de alcance para próteses de membro superior para amputações em nível superior ao nível do cotovelo.</b>	33
Figura 9: <b>Ilustração genérica do funcionamento da proposta de sistema semiautônomo de alcance para próteses de membro superior proposto.</b>	34
Figura 10: <b>Demonstração do ambiente e disposição dos componentes utilizados para gravação do conjunto de dados.</b>	39
Figura 11: <b>Objetos utilizados para compor o conjunto de dados desenvolvido no trabalho.</b>	39
Figura 12: <b>Representação do movimento do Widow X ao longo do tempo para registro das diferentes perspectivas dos objetos utilizados para compor o conjunto de dados.</b>	40
Figura 13: <b>Malha de controle proporcional desenvolvida para concretização do sistema de rastreamento em tempo real baseado no modelo de detecção de objetos para alcance.</b>	42
Figura 14: <b>Processo de movimentação do WidowX em direção ao objeto de interesse.</b>	43
Figura 15: <b>Braço robótico (<i>WidowX Robot Arm Mark II</i>).</b>	44
Figura 16: <b>Arranjo experimental para a realização do Experimento 1.</b>	45
Figura 17: <b>Arranjo experimental para a realização do Experimento 2.</b>	46
Figura 18: <b>Arranjo experimental proposto para realização do Experimento 3.</b>	47
Figura 19: <b>Arranjo experimental utilizado para o Experimento 4 de validação do modelo de alcance para próteses de membro superior.</b>	48
Figura 20: <b>Disposição do objeto para a realização do Experimento 5.</b>	49
Figura 21: <b>Representação da superfície de eventos de cada um dos objetos gravados para o conjunto de dados.</b>	54
Figura 22: <b>Representação das superfícies de eventos em diferentes momentos da gravação do conjunto de dados de acordo com o movimento do braço robótico.</b>	55
Figura 23: <b>Resultado do processo de conversão das informações neuromórficas em uma superfície de eventos e do processo de geração dos rótulos da posição de cada classe em sua respectiva gravação.</b>	57
Figura 24: <b>Matriz de Confusão do modelo de detecção de objetos gravados pela câmera baseada em eventos (DVS128).</b>	58
Figura 25: <b>Comparação entre o tamanho e formato da classe Chave com a classe Caneca.</b>	59
Figura 26: <b>Comparação entre a morfologia dos objetos contidos na classe Garfo e Faca.</b>	59

Figura 27: Apresentação da Curva-P do modelo de detecção de objetos.....	60
Figura 28: Apresentação da Curva-R do modelo de detecção de objetos.....	61
Figura 29: Apresentação da Curva-PR do modelo de detecção de objetos.....	62
Figura 30: Apresentação da curva do <i>F1-Score</i> do modelo de detecção de objetos.....	63
Figura 31: Imagem comparando o padrão de eventos gerados no Experimento 1 e 3. ....	64
Figura 32: Imagem comparando o padrão de eventos gerados no Experimento 2 e 4.....	64
Figura 33: Imagem comparando o padrão de eventos gerados no Experimento 1 e 2.....	65
Figura 34: Imagem comparando o padrão de eventos gerados no Experimento 3 e 4.....	65
Figura 35: Imagem do processo de movimentação do objeto durante o Experimento 5. ....	66
Figura 36: Acurácia geral do sistema de rastreo em tempo real, em cada experimento realizado. .....	67
Figura 37: Gráfico de acurácia do sistema de rastreo ao longo do tempo, discriminado por experimento.....	69
Figura 38: Relação entre a proximidade do objeto de interesse em relação ao campo de visão da DVS128. ....	70
Figura 39: Apresentação da média de distribuição de resultados do sistema de rastreo em tempo real durante o movimento de alcance para cada experimento realizado.....	71
Figura 40: Tempo médio gasto realizando movimento sacádico até gerar eventos suficientes para que o modelo de detecção de objetos realizasse uma inferência. ....	72
Figura 41: Tempo médio necessário para o modelo realizar o movimento de alcance do objeto de interesse.....	73

## Lista de Abreviaturas e Siglas

DI	Detecção Inválida
DoF	<i>Degrees of freedom</i> - Graus de Liberdade
DV	Detecção Válida
DVS128	Dynamic Vision Sensor 128
EMG	Eletromiografia
GPU	<i>Graphic Processing Unit</i> - Unidade de Processamento Gráfico
HDR	<i>High Dynamic Range</i> – Alto Range Dinâmico
IoU	<i>Intersection over Union</i> – Interseção sobre União
MS	Movimento Sacádico
SIHSUS	Sistema de Informações Hospitalares do Sistema Único de Saúde
SNC	Sistema Nervoso Central
YOLO	<i>You Only Look Once</i>

---

# Capítulo 1

---

## Introdução

Diminuição de autonomia, limitação na capacidade de realização de atividades laborais e do dia a dia, impactos nas interações físicas e sociais são algumas das consequências devastadoras da perda de um membro do corpo humano para o indivíduo. A mão, por sua vez, permite aos seres humanos realizar diversas tarefas no dia a dia, além de ser interface de interações sociais e afetivas [1], [2]. A destreza manual permitiu, e continua permitindo, que o ser humano performe atividades que necessitam de força e precisão, garantindo que a espécie humana pudesse fabricar ferramentas e desenvolvesse atividades como agricultura e caça, promovendo a perpetuação e evolução da espécie [2].

Por desempenhar papel tão importante na vida humana, aqueles que sofrem amputações de membro superior são profundamente impactados com as consequências desse trauma, afetando sua comunicação e capacidade de realizar tarefas do dia a dia [3], [4]. Além de prejuízos às condições motoras e sociais, amputações podem causar danos psicológicos no amputado e também em seus familiares [5].

Segundo dados disponibilizados pelo Sistema de Informações Hospitalares do Sistema Único de Saúde (SIHSUS - [6]), somente no Brasil, a cada ano, têm aproximadamente 1480 novos casos de amputações de membros superiores, ou seja, a cada ano é possível encontrar um número significativo de pessoas que necessitam de amparo e assistência. Uma das abordagens para recobrar algumas funcionalidades do membro perdido tem sido a utilização de próteses mioelétricas. Entretanto, a falta de graus de liberdade e dificuldades relacionadas a usabilidade das próteses tem sido motivos de rejeição das mesmas pelos usuários [7], [8]. Dessa forma, iniciativas que reduzem a rejeição e aumentam a usabilidades das próteses a partir de novas interfaces tecnológicas e sensoriais têm se tornado cada dia mais relevantes [9]–[11].

Para enfrentar diversos desafios da robótica, uma nova metodologia tecnológica tem ganhado espaço, alinhando as perspectivas biológicas e de engenharia, os chamados

dispositivos neuromórficos utilizam de abordagens bioinspiradas para desenvolver sistemas com alto poder computacional e baixo consumo de energia [9], [12]–[17].

As câmeras neuromórficas (ou câmeras baseadas em eventos) foram apresentadas ao mundo pela primeira vez na capa da *Scientific American* em 1991 com o título “*Silicon Retina*” [18]. Esses dispositivos bioinspirados funcionam de forma assíncrona capturando eventos de mudanças de luminosidade em uma cena, similarmente aos *spikes* (ou eventos) nas vias nervosas do sistema visual biológico [13], [14], [19]. Após isso alguns trabalhos apresentaram modelos aprimorados do dispositivo até chegarmos aos modelos comerciais atuais [13], [14].

As informações adquiridas pelas câmeras baseadas em eventos são diferentes das imagens de intensidade luminosa padrão. Enquanto as câmeras tradicionais adquirem a intensidade da luz de uma cena de forma periódica, nas câmeras baseadas em eventos cada pixel funciona de forma assíncrona monitorando a variação da luminosidade (Figura 1). Essas características garantem vantagens que consistem em um alto alcance dinâmico, alta resolução temporal, baixa latência e baixo consumo de energia [12]–[14], [19].

Por causa da natureza esparsa dos dados adquiridos pelas câmeras baseada em eventos, novas abordagens computacionais devem ser desenvolvidas para resolver tarefas tradicionais de Visão Computacional. Cada evento adquirido por esses dispositivos pode ser definido como

$$e_i = [S_i(x, y), t_i, P_i], \quad i \in N^* \quad (1),$$

onde,  $e_i$  é o  $i^o$  evento na sequência de eventos e contém três informações básicas de cada evento: a localização espacial  $S_i(x, y)$  onde a variação de luminosidade ocorreu, o tempo  $t_i$  em que o evento ocorreu e a polaridade  $P_i$ , com  $P_i \in \{-1, 1\}$ , onde  $-1$  e  $1$  representam os eventos de diminuição e aumento de luminosidade, respectivamente.

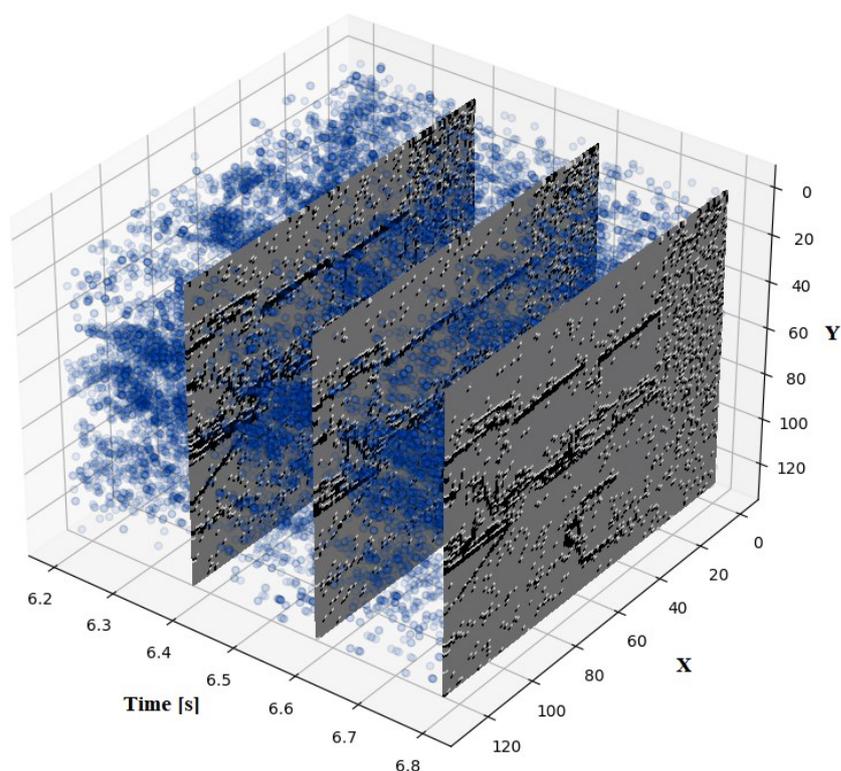


Figura 1: **Representação dos eventos de uma gravação utilizando uma câmera baseada em eventos.** Os eventos são apresentados em azul, considerando a informação temporal e espacial do evento, embora não seja apresentada a distinção de polaridade dos eventos na imagem. Em tons de cinza é possível observar o resultado de um processo de integração dos eventos em uma janela temporal ( $\Delta t$ ), responsável por gerar um processo de visualização dos eventos de uma maneira familiar (produzida pelas câmeras tradicionais).

Para usar técnicas convencionais de aprendizado profundo em tais dados esparsos das câmeras baseadas em eventos, alguns trabalhos têm se concentrado em converter os eventos em uma representação densa [20]–[22] dos mesmos. Porém, as informações registradas pelas câmeras baseadas em eventos, constantemente enfrentam o desafio de associação de dados gerando uma mudança na aparência da cena dependendo da direção do movimento do sensor (Figura 2).

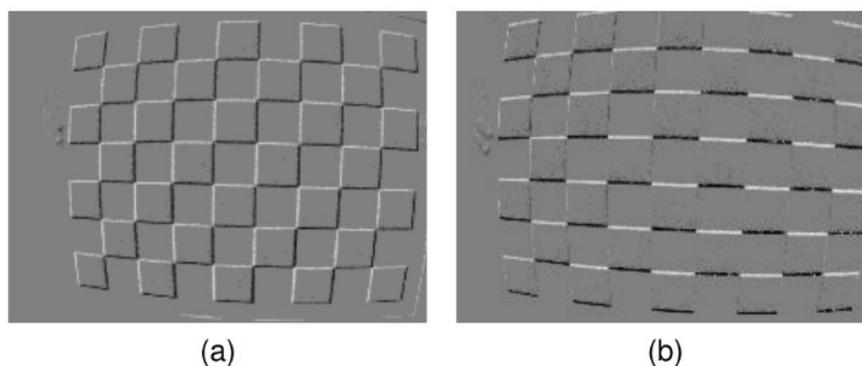


Figura 2: **Representação do desafio de associação de dados.** Ambas as representações (a) e (b) são da mesma cena, no entanto, os movimentos do sensor baseado em eventos são diferentes. Em (a) o movimento é na diagonal e em (b) o movimento é vertical. Essas imagens são geradas integrando eventos em um intervalo de tempo, onde em cinza são representados os pixels onde não há alteração de intensidade, e os pixels onde há alteração de intensidade positiva e negativa são marcados como branco e preto, respectivamente. Imagem adaptada de [23].

O desafio de associação de dados se manifesta quando o intuito é discriminar padrões de eventos relacionados a uma mesma fonte de informação. Ou seja, um objeto simples como um copo gera diversos padrões de eventos dependendo da direção do movimento da câmera neuromórfica. Dessa maneira, é necessário que a detecção de objetos, através das informações geradas por esses dispositivos, seja robusta à multiplicidade de representações que um mesmo objeto pode possuir.

Por possuir baixa dimensionalidade de dados, alto desempenho computacional e baixo consumo de energia, os dispositivos neuromórficos são adequados para aplicações vestíveis [24], [25], aplicações portáteis onde a alta eficiência energética é conveniente [26], [27] e em aplicações onde o alto desempenho computacional é requerido [28], [29]. Essas características permitem que esses dispositivos sejam muito adequados para aplicações robóticas [9], [30]–[33].

Nesse contexto, a primeira hipótese desse trabalho pressupõe um alinhamento entre sistemas neuromórficos e sistemas de aprimoramento de usabilidade e controle para próteses e se baseia na premissa de que um sistema semiautônomo para atividades complexas de alcance de objetos, utilizando visão neuromórfica, pode ser apropriado para próteses de membro superior, em amputações acima do nível do cotovelo. Por sua vez, dispositivos de sensoriamento neuromórfico apresentam contribuições significativas para o campo da robótica [14], [34]–[36]. Dessa maneira, a segunda hipótese é que sistemas neuromórficos podem guiar

o desenvolvimento de dispositivos inteligentes de controle com baixo consumo de energia e alto desempenho computacional.

Para tanto, essa dissertação apresenta contribuições no campo do desenvolvimento de sistemas semiautônomos para próteses, ao propor um sistema de alcance de objetos para próteses de membro superior, utilizando uma câmera baseada em eventos (Dynamic Vision Sensor - DVS128), além de contribuir para o desenvolvimento de novas abordagens e algoritmos para o desenvolvimento de sistemas neuromórficos de visão computacional.

## 1.1 Justificativa

Projeções mostram que apenas nos Estados Unidos o número de amputados pode chegar a 3,6 milhões de pessoas até 2050, 30% desses são amputações de membro superior [7]. No Brasil, segundo o SIHSUS, de 2014 a 2019 foram registrados um total de 7521 casos de amputações de membro superior. As amputações de membro superior acometem principalmente a população mais jovem (até 40 anos) e, portanto, economicamente ativa [6].

Dessa maneira, é importante ressaltar que há um impacto social resultante dos casos de amputações. O processo de recuperação e reabilitação são de suma importância para reinserção dos indivíduos na sociedade. Segundo o trabalho realizado em [37], cerca de 73,2% dos amputados voltam a trabalhar e 66% deles trocam de emprego. Já o trabalho realizado em [38], ressalta que cerca de 80% dos amputados utilizam uma prótese com o intuito recobrar as funcionalidades do membro perdido, entretanto a prótese só era utilizada em metade das tarefas do dia a dia.

O custo cognitivo aplicado no controle e utilização desses dispositivos protéticos é tão alto quanto os graus de liberdade que ele possui [11], [39], [40], dessa maneira, quanto mais alto o nível da amputação, maior a quantidade de graus de liberdade que a prótese possuirá e mais difícil o controle por parte do usuário. Nesse sentido, alguns trabalhos apontam que próteses mioelétricas com maior usabilidade e mais graus de liberdade seriam mais eficientes em recobrar as funcionalidades do membro perdido [5]. Logo, emerge a necessidade em desenvolver estratégias para melhorar a usabilidade e controle de todo o aparato protético utilizado por esses indivíduos.

Portanto, justifica-se esse estudo pela sua contribuição tanto para a academia quanto para a sociedade. Para a academia, por se tratar de uma pesquisa acerca dos caminhos possíveis para os avanços tecnológicos de dispositivos protéticos com baixo consumo de energia e alto poder computacional. Ademais, a contribuição do presente projeto para a sociedade ocorre através do auxílio no aprimoramento de mecanismos de amenização dos impactos sociais e pessoais na parcela da população que sofre amputações.

## 1.2 Objetivos

Essa dissertação tem como objetivo geral desenvolver um sistema de controle semiautônomo de alcance para futura aplicação em próteses de membro superior a partir do sensoriamento com dispositivos neuromórficos, além de propor um caminho possível para o desenvolvimento de sistemas inteligentes com baixo consumo de energia.

Ademais, diante do objetivo geral da pesquisa, é possível pormenorizar os seguintes objetivos específicos:

- i. Desenvolver um sistema de visão neuromórfica com o potencial de possibilitar maior autonomia de próteses de membros superiores para alcance de objetos;
- ii. Propor um modelo, para detecção de posição, rastreamento e identificação de objetos, em tempo real, utilizando sistemas de visão neuromórfica;
- iii. Disponibilizar um conjunto de dados de gravações neuromórficas de objetos em cenários complexos (com eventos provenientes da presença de plano de fundo) e em cenários simples (sem informações provenientes da presença de plano de fundo).

## 1.3 Estrutura da dissertação

Esta dissertação está estruturada da seguinte forma:

- Capítulo 2: Fundamentação Teórica

Este capítulo apresenta um panorama geral acerca dos tópicos essenciais para a compreensão dos assuntos abordados nessa dissertação. Sendo assim, são abordados o funcionamento de próteses de membro superior, e o respectivo

processo de evolução ao longo dos anos, além de perspectivas atuais e futuras de sensoriamento e controle na construção de próteses inteligentes. Nesse capítulo, também é abordado os dispositivos desenvolvidos utilizando a filosofia neuromórfica e o funcionamento das câmeras baseadas em eventos (ou câmeras neuromórficas), além de discorrer sobre o panorama atual dos algoritmos desenvolvidos para trabalhar com as informações espaço-temporais desses dispositivos.

- **Capítulo 3: Metodologia**

Trata-se da proposta do sistema de alcance para próteses de membro superior. Nesse cenário, são descritos o modelo utilizado para detecção de objetos, o sistema de rastreamento em tempo real, os procedimentos experimentais de avaliação e validação, além do detalhamento da análise de dados.

- **Capítulo 4: Resultados e Discussão**

São apresentados os resultados obtidos a partir dos modelos e métricas apresentados no Capítulo 3. Além de trazer uma discussão acerca dos resultados e, portanto, um posicionamento do trabalho atual no estado da arte.

- **Capítulo 5: Conclusões**

Nesse capítulo, são apontados e reforçados os objetivos iniciais do sistema de alcance utilizando câmeras baseadas em eventos, as possibilidades em aberto para melhorias em sistemas de alcance para próteses e as perspectivas de trabalhos futuros na linha de utilização de sistemas neuromórficos para dispositivos inteligentes.

---

## Capítulo 2

---

### Fundamentação Teórica

No capítulo 2, a seguir, são tratados os conceitos necessários para o entendimento e compreensão dos tópicos discutidos ao longo do restante do trabalho. Nele será apresentado uma série de trabalhos que fundamentam os assuntos abordados e discutidos adiante.

#### 2.1 Próteses de membros superiores

Próteses funcionam como uma alternativa sintética para a substituição de um membro perdido. Esses dispositivos visam recobrar algum aspecto do membro ausente, seja ele estético, funcional ou, idealmente, ambos. A perda de um membro superior pode ocorrer em diferentes níveis, podendo ser desde uma amputação a níveis dos ossos carpos da mão (transcarpal) até a remoção do membro a partir da escápula (interscapulotorácica).

Dessa maneira, as próteses de membro superior precisam abranger uma vasta gama de possibilidades de amputações. Os dispositivos protéticos para membros superiores são classificados em dois grandes grupos (baseado em suas funcionalidades, conforme ilustrado no diagrama da Figura 3): próteses passivas e ativas. As próteses passivas, por sua vez, são subdivididas em dois outros grupos: próteses funcionais e próteses cosméticas. Enquanto as próteses ativas se dividem em: próteses movidas por energia corporal (próteses por tirantes) e próteses movidas por fonte de energia externa (baterias).

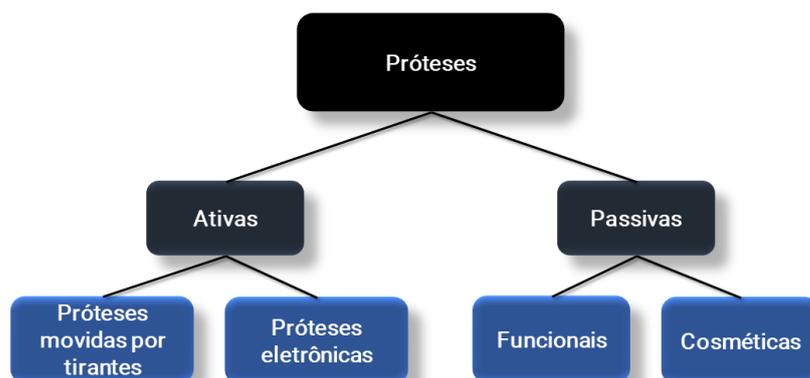


Figura 3: **Representação hierárquica das classificações dos dispositivos protéticos comerciais.**

Enquanto próteses cosméticas se restringem apenas a substituição do membro perdido em termos estéticos, as próteses funcionais se baseiam em dispositivos com finalidades muito específicas como as utilizadas em esportes e trabalho. As próteses por tirante se baseiam em sistemas onde o usuário pode realizar movimentos a partir do tensionamento e relaxamento de cabos presos ao seu membro sadio e, portanto, são dispositivos onde a força motriz de trabalho é proporcionada pelo próprio usuário. Por fim, as próteses ativas e mais sofisticadas são as alimentadas por fonte externa ao usuário, como baterias, e usualmente apresentam diversos motores, sensores e controle baseado em eletromiografia (Tabela 1).

Tabela 1: **Comparação de três tipos de próteses ativas comerciais.** Adaptado de [41]

	Prótese		
	i-Limb	Bebionic	Michelangelo
<b>Peso</b>	443 - 515 g	550 - 598 g	420 g
<b>Nº de atuadores</b>	6 motores DC	5 motores DC	2 motores DC
<b>Nº de Graus de Liberdade (DOF)</b>	6	6	2
<b>Nº de tipos de pegada</b>	5	5	3
<b>Máxima força aplicada</b>	100-136 N	140 N	70 N

Os desafios no desenvolvimento de próteses de membro superior se imiscuam em criar dispositivos com diversos graus de liberdade, sensores, atuadores e componentes eletrônicos,

se mantendo fiel ao peso e tamanho do membro a ser substituído e em promover estratégias de controle para aprimorar as interfaces com o usuário, possibilitando a maior usabilidade e o aumento da funcionalidade desses dispositivos artificiais.

### 2.1.1 Sistemas de automatização de próteses

Os dispositivos protéticos de membro superior disponíveis no mercado como a i-Limb, a Michelangelo e a BeBionic (Figura 4) apresentam uma interface de controle através de eletromiografia de superfície (EMG), caracterizando-se, portanto, como próteses do grupo ativo, movidas por fonte de energia externa [42]. Entretanto, geralmente, sistemas de controle baseados em EMG possuem resposta de ativação/desativação controlando apenas a abertura e o fechamento das próteses [43] ou adotam um sistema de controle proporcional à intensidade da ativação muscular monitorada pelo sinal EMG [44], sendo essas abordagens as mais adotadas em sistemas comerciais para próteses de membro superior [45]. Entretanto, ao utilizar essas estratégias, os graus de liberdade que o usuário é capaz de utilizar é proporcional a quantidade de canais de eletromiografia e músculos utilizados.



Figura 4: **Imagem dos três modelos comerciais de próteses de mão, tecnologicamente avançadas.** Em (a) é apresentada a i-Limb [46], um modelo comercial produzido pela Touch Bionics. Em (b) está a BeBionic, uma prótese de mão produzida pela Ottobock [47], enquanto em (c) se encontra o modelo Michelangelo, também produzida pela Ottobock [48].

Visando a otimização entre os graus de liberdade e a quantidade de músculos necessários para controlá-los, alguns trabalhos visam a utilização de técnicas de reconhecimento de padrões

de atividade muscular para gerar sistemas de controle que permitam maior controle de todos os graus de liberdade possíveis em próteses de membro superior. Entretanto, mesmo essas técnicas encontram saturações na quantidade de canais de EMG necessários para controlar todos os graus de liberdade possíveis para uma prótese de membro superior, além de serem necessárias exaustivas horas de treinamento para que o usuário consiga aprender a controlar esses dispositivos [45], o que limita a aplicabilidade clínica dessas metodologias.

Uma das limitações das próteses de membro superior, constitui-se na falta de controle intuitivo e uma interface confiável capaz de discriminar a intenção do usuário. Dessa maneira, o aprimoramento de dispositivos protéticos se fundamenta em três pilares: a) o desenvolvimento de interfaces naturais e intuitivas de controle para os diversos graus de liberdade presentes nesses dispositivos; b) interface sensorial desses dispositivos com o Sistema Nervoso Periférico; e c) diminuição do tempo de treinamento necessário para utilização dos dispositivos protéticos [8], [49].

Dessa maneira, uma das abordagens utilizadas para controle das próteses constitui-se em dar certa autonomia para esses dispositivos protéticos, de maneira que eles sejam capazes de discriminar a intenção do usuário e promover tarefas como alcançar, manipular, controlar a força de pegada e a estratégia de pegada que melhor se adapta ao objeto desejado [9], [10], [50]–[54] sem que o usuário necessite de atuar ativamente em todas as etapas de execução desses tipos de tarefas. Esses processos de automação de membros protéticos se baseiam no sensoriamento e atuação em malhas de controle embarcadas (presentes na eletrônica das próteses), permitindo que o dispositivo possua certa autonomia na execução de tarefas, diminuindo o esforço do usuário, aumentando a usabilidade desses dispositivos e promovendo a diminuição do tempo de treinamento.

## **2.2 Sistemas neuromórficos**

Ao integrar engenharia e biologia para resolução de problemas, a Engenharia Neuromórfica promove o desenvolvimento de sistemas computacionais que se assemelham ao funcionamento de sistemas biológicos. Esses dispositivos transmitem informações de maneira assíncrona e baseada em eventos, ou seja, o processo de transmissão de informações é semelhante aos

potenciais de ação em sistemas biológicos (*spikes*).

As premissas do desenvolvimento de sistemas neuromórficos consistem em eficiência energética e computacional, baixa redundância de informações, como os sistemas biológicos. Esses dispositivos têm aberto caminho para a nova geração de Redes Neurais, permitindo aplicações mais responsivas e energeticamente eficientes [35], [55], [56]. Empresas como a IBM (*IBM Research in Zurich*) e a Intel (*Intel Neuromorphic Research Community*) possuem laboratórios de pesquisa em tecnologia bioinspirada, algoritmos e aplicações utilizando os conceitos da computação neuromórfica [17], [57], [58].

Sistemas neuromórficos podem se dividir em dois grupos: I) modelos computacionais neuromórficos que tratam do desenvolvimento de algoritmos e abordagens computacionais bioinspiradas; e II) sensores neuromórficos que por sua vez apresentam um processo de codificação da informação baseada em eventos (*spikes*) e, portanto, trabalham de maneira assíncrona e bioinspirada registrando o instante de tempo em que ocorreu o evento, a polaridade e a localização, evento que, por sua vez, são gerados a partir de algum estímulo monitorado.

O processo de transformação de informação contínua no tempo em eventos se dá por um sistema inspirado no disparo de neurônios do Sistema Nervoso Central (SNC). Dessa maneira, em sistemas biológicos, para que ocorra um evento (ou Potencial de Ação) é necessário que o potencial de membrana de um neurônio atinja um limiar de ativação ( $\sim -55\text{mV}$ ) para que desencadeie um processo de despolarização e repolarização da membrana celular [59]. Logo, é necessário que haja um estímulo (e esse estímulo seja grande o suficiente) para elevar o potencial de membrana para o limiar de disparo do Potencial de Ação (Figura 5).

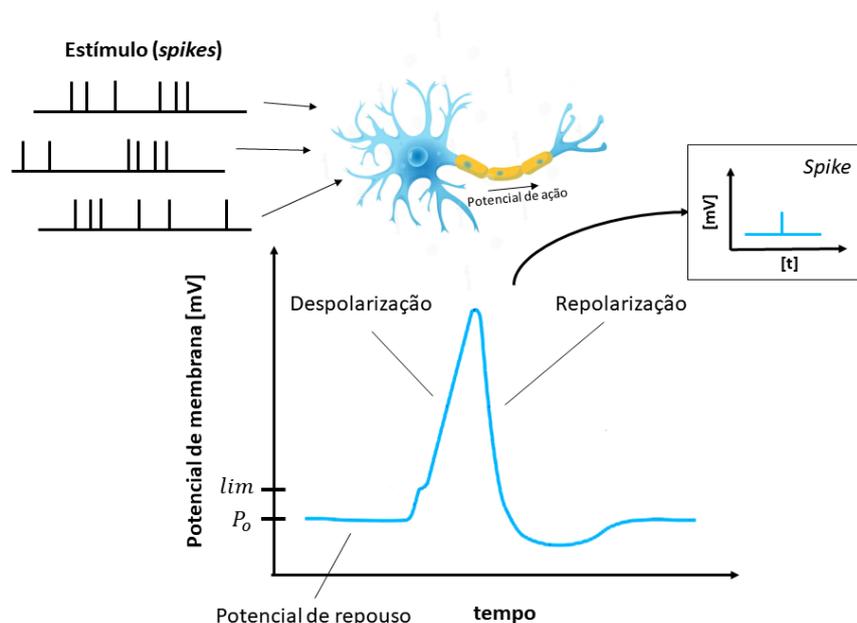


Figura 5: **Representação da resposta elétrica da membrana de uma célula excitável neuronal.** Na parte superior da imagem é apresentada uma série de estímulos sendo aplicados nos dendritos de um neurônio, provocando o disparo de um potencial de ação no corpo celular. Na parte inferior da imagem é apresentado o potencial elétrico da membrana discretizado no tempo, demonstrando o potencial de repouso da membrana, o aumento do potencial de membrana devido aos estímulos recebidos e o momento que ele atinge o limiar de ativação, onde ocorre o processo de abertura dos canais iônicos voltagem dependente do corpo celular e o processo de despolarização ocorre permitindo a propagação da informação [59].

De maneira similar o processo de geração de eventos em sistemas neuromórficos segue o mesmo padrão, ou seja, a partir de um fenômeno físico observável o valor médio basal é determinado como o Potencial de Repouso (como em uma célula excitável no sistema biológico), e se a variação desse sinal for suficiente para atingir o limiar de ativação do sistema, um evento é gerado de acordo com a polaridade do estímulo que lhe deu causa, uma polaridade positiva é referente a um estímulo que representa uma variação positiva do fenômeno, enquanto a polaridade negativa é referente a uma variação negativa do fenômeno.

### 2.2.1 Câmera baseada em eventos

Câmeras neuromórficas (baseadas em eventos) apresentam diversas características que as tornam diferentes das câmeras tradicionais (baseada em frames). A Tabela 2 sumariza as

grandes diferenças entre ambos os sistemas e são as diferenças fundamentais dos dois dispositivos que garantem, como consequência, características práticas que permitem que a câmera baseada em eventos possua vantagens em relação as câmeras tradicionais. Em câmeras baseadas em eventos, as informações são esparsas no tempo e espaço, adquiridas de maneira assíncrona, enquanto em câmeras baseadas em frames a codificação das informações é realizada de maneira síncrona e baseada em intervalos fixos de tempo.

Tabela 2: **Comparação entre uma câmera tradicional e uma neuromórfica.** A comparação foi realizada utilizando um modelo comercial (puA1920-30uc) de câmera tradicional (baseada em *frames*) produzida pela Basler e um modelo comercial (DVS128 – utilizada nesse trabalho) de câmera neuromórfica (baseada em eventos) produzida pela IniVation Labs.

	Câmera	
	Tradicional <sup>1</sup>	Neuromórfica <sup>2</sup>
<b>Modo de aquisição de informação</b>	Síncrono	Assíncrono
<b>Saída</b>	Quadro ( <i>frame</i> )	evento
<b>Resolução temporal</b>	Determinado pelo Tempo de Exposição ( $\sim 33ms$ )	$\sim 1\mu s$
<b>Resolução [HxV]</b>	1920 px x 1080 px	128 px x 128 px
<b>Range Dinâmico</b>	Baixo	Alto
<b>Potência</b>	1.3 W	23 mW

<sup>1</sup> Informações retiradas com base em um modelo intermediário (puA1920-30uc) de uma câmera baseada em frames da Basler.

<sup>2</sup> Informações referentes a DVS128 [13]

Cada pixel, em câmeras baseadas em frames, armazena um valor de intensidade luminosa no mesmo instante de tempo, diferentemente de câmeras neuromórficas que disparam eventos apenas quando há uma variação de intensidade luminosa relativa ao valor de luminosidade basal. Dessa maneira, as duas grandes diferenças no princípio de funcionamento desses dispositivos se encontram no processo de transmissão da informação e na caracterização de qual informação transmitir (Figura 6).

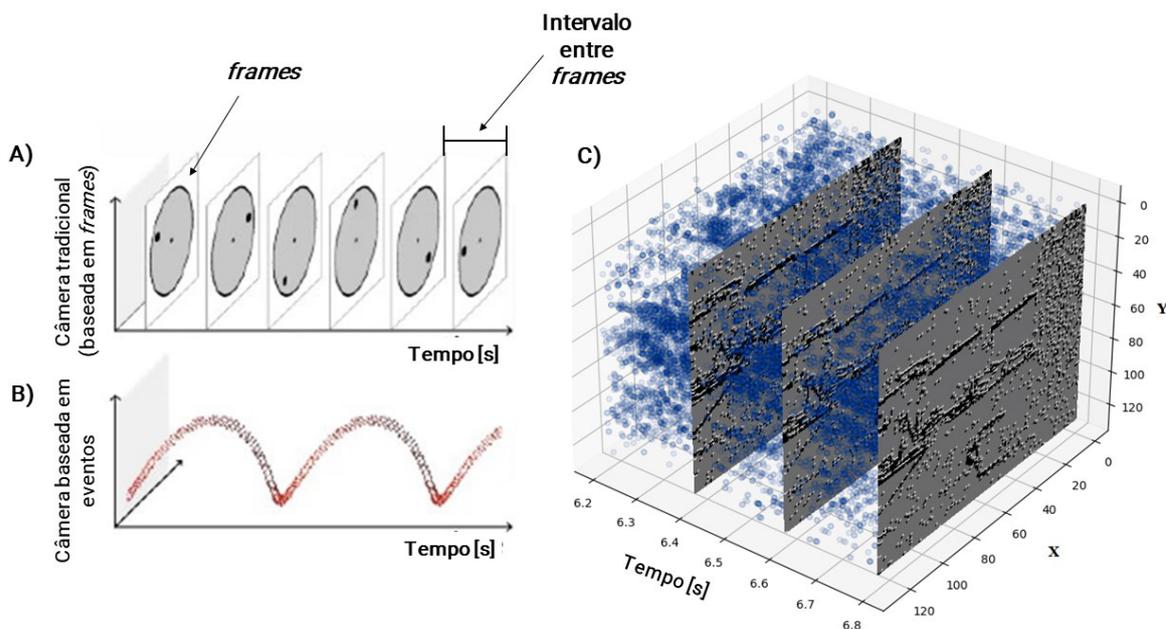


Figura 6: **Representação da diferença de saída entre uma câmera tradicional (baseada em frames) e uma câmera neuromórfica (baseada em eventos).** Em (a) tem-se a saída de uma câmera tradicional, onde a informação é adquirida de forma síncrona e temporalmente espaçada, conforme o Tempo de Exposição dos sensores desses dispositivos. Caso algo aconteça entre o tempo de aquisição dos *frames* essa informação é perdida, entretanto as câmeras baseadas em eventos possuem alta resolução temporal (b) e, portanto, apresentam uma maior capacidade de discriminação de informações no tempo (adaptado de [60]). Em (c) é apresentada uma mesclagem das duas abordagens mostrando os eventos que ocorrem entre consecutivos frames de uma gravação. Uma versão animada de (a) e (b) pode ser encontrada em: <https://youtu.be/LauQ6LWTkxM>.

Devido a seu princípio de funcionamento baseado em eventos, as câmeras neuromórficas apresentam vantagens frente às câmeras baseadas em frames, sendo elas: alta resolução temporal ( $\sim 1 \mu\text{s}$ ), baixa latência, alto alcance dinâmico (HDR) e baixo consumo de energia [13], [14]. Dessa maneira, esses dispositivos apresentam grande aplicabilidade na robótica e em dispositivos vestíveis para funcionamento em aplicações em alta velocidade, baixa latência e baixo consumo de energia. Entretanto, as informações das câmeras baseadas em eventos consistem em dados assíncronos de informações no espaço-tempo e, portanto, novos paradigmas de processamento precisam ser desenvolvidos para tratar esses dados.

Para gerar eventos, as câmeras neuromórficas realizam um processo de conversão da intensidade luminosa, contínua no tempo, em informações discretas de aumento ou diminuição de luminosidade. Dessa maneira, na Figura 7, é apresentado uma ilustração acerca do processo

de geração de eventos a partir da variação de intensidade luminosa em cada pixel do sistema.

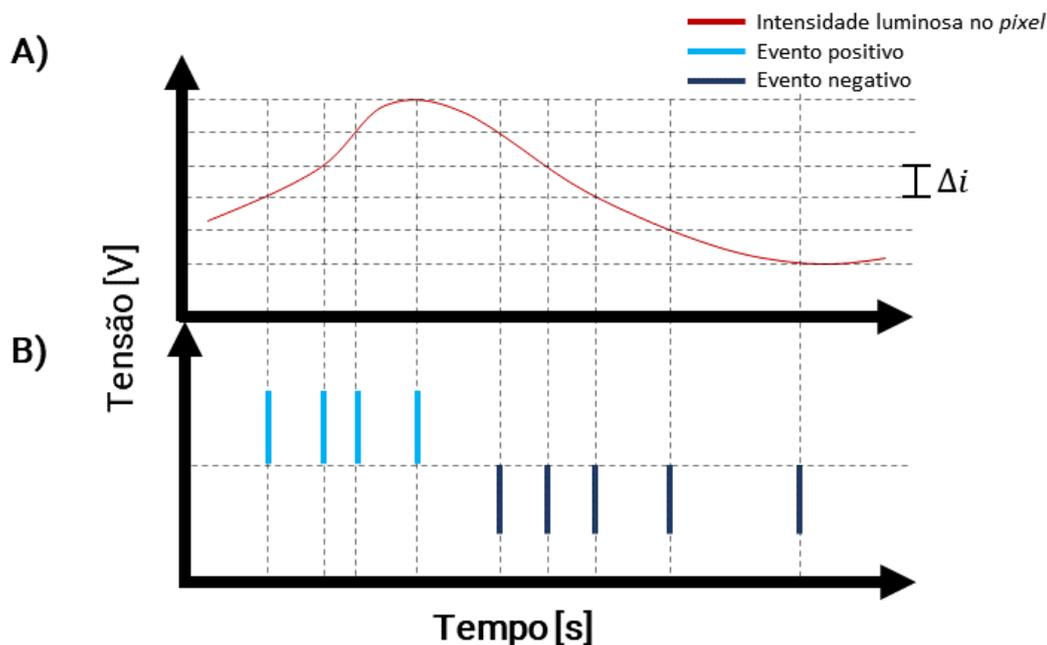


Figura 7: **Ilustração do processo de conversão da luminosidade que atinge cada fotodiodo (*pixel*) em eventos de variação de luminosidade.** Em (a) é possível observar o valor da tensão no fotodiodo (resultado de uma variação diretamente proporcional a intensidade luminosa que atinge o *pixel*) e o limiar de variação de luminosidade ( $\Delta i$ ) necessário para gerar um evento (representado pelas linhas pontilhadas horizontais). As linhas pontilhadas verticais representam o instante temporal em que houve uma variação de luminosidade suficiente para atingir o limiar de variação e, portanto, gerar um evento (b). Caso a variação de luminosidade seja positiva o evento tem polaridade positiva (+1), caso contrário o evento recebe polaridade negativa (-1).

### 2.3 Visão Neuromórfica

Um grande desafio de Visão Computacional se encontra em classificar, detectar e dar sentido ao conteúdo das imagens registradas com câmeras baseadas em frames. Porém, os bons resultados registrados na literatura, só puderam atingir o nível de maturidade atual com o advento da colaboração internacional na criação de grandes conjuntos de dados e da utilização de Unidades de Processamento Gráfico (GPU), que permitiram um alto poder computacional dedicado no tratamento desses grandes conjuntos de dados por meio de técnicas de Aprendizagem Profunda [61]. Diversos algoritmos sofisticados, utilizando informações de câmeras tradicionais, têm explorado essas ferramentas ao longo dos últimos 10 anos [62]–[66].

Entretanto, o campo de Visão Neuromórfica ainda é um tanto quanto imaturo e pouco

colaborativo. Esse baixo nível de maturidade é consequência de fatores como a baixa disponibilidade comercial de câmeras baseadas em eventos e de conjuntos de dados neuromórficos. Todavia, esse cenário tem se modificado ao longo dos anos, pois a presença de grandes empresas, como a Sony e a Samsung, no ramo de visão baseada em eventos, tem impulsionado o desenvolvimento tecnológico das câmeras (com o aumento de resolução das matrizes de fotodiodos), a disponibilidade comercial (através da fabricação e eventual comercialização em larga escala desses dispositivos) e no estabelecimento de uma comunidade de desenvolvedores de novas abordagens computacionais e novos algoritmos capazes de processar as informações esparsas das câmeras baseadas em eventos [67], [68].

Por sua natureza baseada em eventos, novos algoritmos e abordagens de processamento das informações desses dispositivos precisam ser desenvolvidos [12]. Entretanto, vários são os desafios para se alcançar o mesmo desempenho dos sistemas de Visão Computacional tradicionais.

Os modelos de extração de informações, a partir de imagens de sistemas de visão tradicionais, podem se basear em características como cor, morfologia, histogramas, contraste, entre outras que só podem ser adquiridas pela natureza de intensidade de pixel e cor RGB de câmeras baseadas em frames. Entretanto, como dispositivos neuromórficos atuais possuem baixa resolução espacial e possuem funcionamento baseado em eventos, cuja única informação acerca da intensidade luminosa é proveniente do valor da polaridade do evento, essas características se encontram imiscuídas na relação espaço-temporal dos eventos. Portanto, novas metodologias [21], [34], [69] precisam ser desenvolvidas para extrair características desses eventos ou metodologias atuais necessitam de adaptações [20], [70], [71].

Uma abordagem muito eficiente em Visão Computacional para extrair características de imagens e dar significado a elas são os modelos de Aprendizado Profundo (ou *Deep Learning*) [61], [67], [68]. Modelos de Aprendizado Profundo necessitam de grandes conjuntos de dados para que sua capacidade de distinção e generalização possa ser alcançado com certa robustez e as informações possam ser processadas e inferências realizadas com a devida acurácia que se deseja

Dessa maneira, para utilizar tais abordagens em sistemas de Visão Neuromórfica, é necessário criar grandes conjuntos de dados para promover o treinamento desses modelos,

porém, ainda são poucos os conjuntos de dados neuromórficos disponíveis. Contudo, alguns trabalhos tem focado em pavimentar o caminho para a geração de grandes conjuntos de dados neuromórficos [72]. Uma das abordagens muito comuns na literatura é o processo de conversão de conjuntos de dados tradicionais em informações neuromórficas [73]–[76]. Essa abordagem é bastante atrativa, tendo em vista que existem diversos conjuntos de dados de câmeras tradicionais disponíveis na literatura e que ajudaram a impulsionar o desenvolvimento de diversos modelos de Aprendizagem Profunda [77]–[80]. Entretanto, ainda é necessário que o modelo aprenda a realizar aferições em informações verdadeiramente neuromórficas, o que pode ser um problema para modelos treinados em conjuntos de dados neuromórficos sintéticos [73].

Nesse contexto, diferentes abordagens são utilizadas para propor sistemas de rastreo utilizando câmeras baseadas em eventos. Em [81] os autores apresentam um método de detecção e rastreamento utilizando uma abordagem baseada em “partes mínimas” de objetos, onde o sistema possui núcleos com formas geométricas simples e assume que um objeto complexo é composto por várias “partes mínimas”; o trabalho em [82] cria uma representação espacial de eventos em uma janela de tempo e realiza a detecção de cantos (com base na metodologia de *Harris Corners* [83]). O trabalho apresentado em [21] também utiliza o mesmo conceito proposto por [81], porém a metodologia implementada é inteiramente baseada em eventos. Em [23] os autores usam uma abordagem de integração entre uma câmera baseada em frames e uma baseada em eventos para propor um sistema de rastreamento de cantos mais robusto e persistente [83]. Entretanto, sistemas de detecção de cantos e de partes genéricas não garantem a discriminação de características específicas que são necessários para sistemas de rastreo onde objetos precisam ser distinguidos entre diversas outras informações na cena.

---

## Capítulo 3

---

### Metodologia

Apresenta a metodologia desenvolvida para a estratégia de alcance para próteses de membro superior baseada em visão neuromórfica. A primeira parte do capítulo discorre acerca da proposta do trabalho e, posteriormente, é discutido o processo de avaliação do modelo proposto e as métricas de validação dos experimentos realizados.

#### **3.1 Proposta do modelo de alcance para próteses de membro superior**

Com o intuito de criar um mecanismo semiautônomo de alcance para próteses de membro superior utilizando sistemas neuromórficos, a proposta desse trabalho (Figura 8) consiste em utilizar uma câmera baseada em eventos acoplada a palma da mão de uma prótese de membro superior, para guiar o braço robótico enquanto o mesmo realiza o movimento de alcance até o objeto de interesse do usuário.

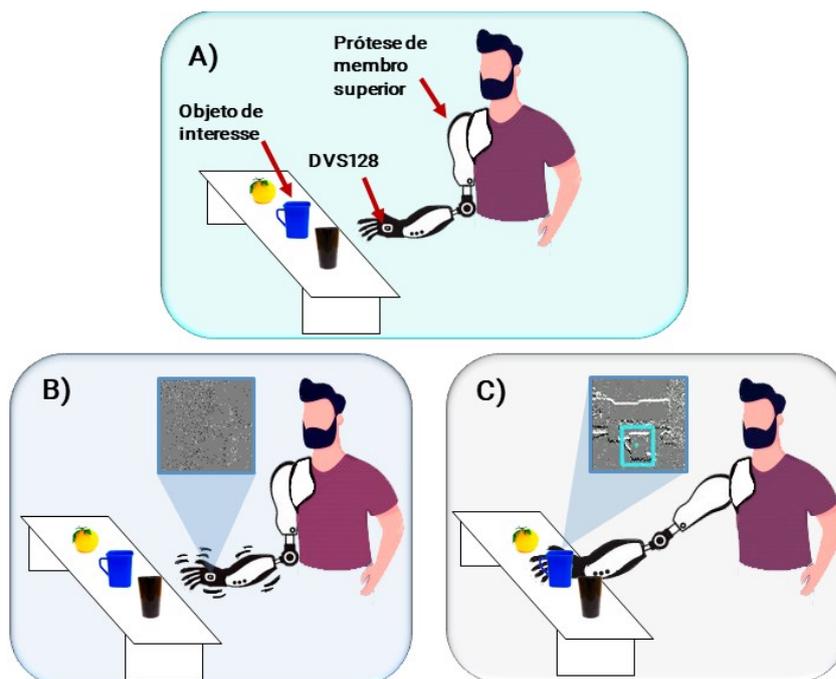


Figura 8: **Modelo de alcance para próteses de membro superior para amputações acima do nível do cotovelo.** Em (a) o usuário encontra-se na intenção de manipular um objeto específico (caneca) em uma mesa com diversos objetos. As informações da cena são captadas pela câmera baseada em eventos acopladas a palma da mão da prótese de membro superior. As informações captadas são processadas e, caso não haja objeto sendo visualizado pela câmera baseada em eventos (b), o braço protético inicia o movimento sacádico (responsável por gerar movimento relativo entre a prótese e o objeto para que ocorra variação de luminosidade e, conseqüentemente, geração de eventos). Quando as informações da câmera baseada em eventos passam pelo modelo de detecção e há a inferência do objeto (c) de interesse presente na cena (objeto mais próximo ao centro da cena), o rastreamento desse objeto se inicia e o braço protético se move em direção ao objeto, adaptando sua posição de maneira que o objeto de interesse se mantenha no centro da imagem adquirida pela câmera baseada em eventos.

Utilizando o sistema proposto, o usuário deve posicionar o membro protético de maneira que o dispositivo de visão neuromórfica (DVS128) consiga visualizar o objeto de interesse (Figura 8.A). Uma vez que o usuário realizou o primeiro passo e determinou o início do processo de alcance, o sistema inicia a captura das informações da câmera e envia o pacote temporal de eventos para o modelo de detecção de objetos (Figura 9).

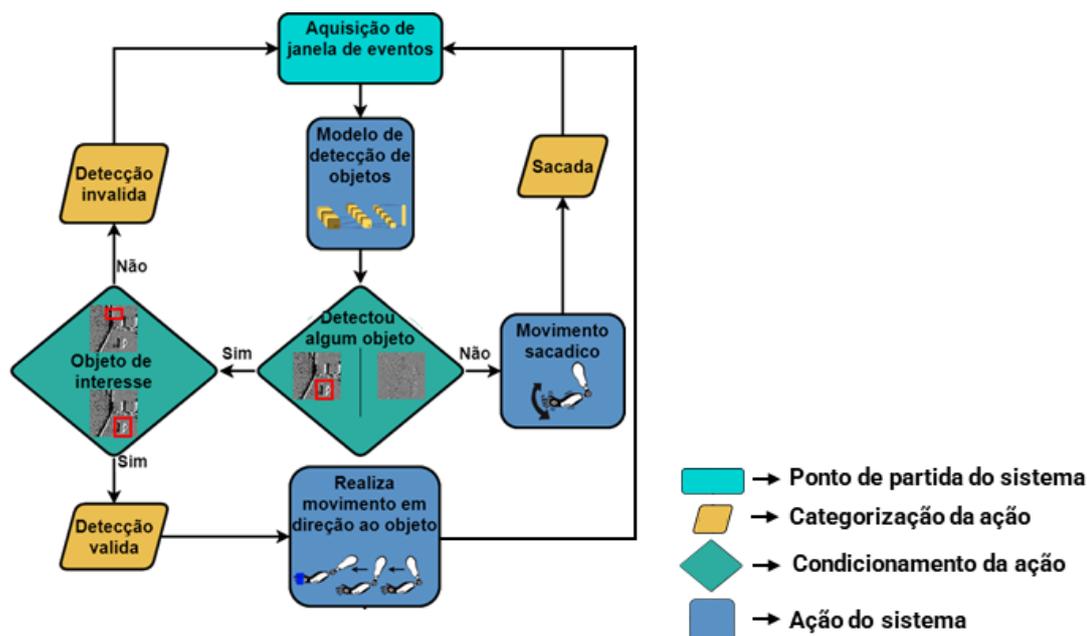


Figura 9: **Ilustração genérica do funcionamento da proposta de sistema semiautônomo de alcance para próteses de membro superior proposto.** Inicialmente o sistema aquisição os eventos provenientes da câmera baseada em eventos, e o modelo de detecção de objetos realiza uma inferência acerca dos objetos em cena. Caso não existam objetos em cena, o modelo realiza um movimento sacádico para gerar eventos a partir do movimento relativo entre a cena e a câmera e caracteriza a interação como uma ação de movimento sacádico (MS). Entretanto, se houver uma detecção, o modelo realiza uma averiguação para confirmar se a detecção é o objeto de interesse, caso o modelo considere a detecção como válida (DV) é realizado um movimento em direção ao objeto, Caso contrário, a interação é classificada como detecção inválida (DI) e nenhum movimento é realizado.

O modelo, por sua vez, pode não detectar objetos (Figura 8.B) devido à falta de um objeto, ou mesmo pela falta de eventos representativos de um objeto, visto a necessidade de variação de luminosidade (ou movimento relativo) necessária para que as câmeras baseadas em eventos gerem informações do ambiente. Nesses casos, o sistema proposto realiza um movimento vertical rápido (semelhante ao movimento sacádico do sistema visual biológico [84]–[86]) para estimular a geração de eventos.

Quando o modelo detecta objetos na cena (Figura 8.C), é realizada uma aferição da posição da detecção atual em relação a detecção anterior. Nesse momento, é suposto que o objeto de interesse não sofrerá uma variação instantânea de posição e, portanto, o modelo assume que uma Detecção Válida (DV) é uma detecção que ocorre abaixo de um limiar de distância (18px)

quando considerado a distância euclidiana entre o centroide da última detecção válida em relação a detecção atual. Caso a detecção realizada não esteja dentro do limiar de distância determinado para caracterização de uma detecção válida, a detecção é classificada como inválida (DI).

Por fim, para determinar o objeto de interesse, na primeira detecção, o modelo leva em consideração a detecção do objeto que está mais próximo do centro da cena. Dessa maneira, as próximas detecções só serão consideradas válidas se estiverem de acordo com a regra de distância do sistema, e o braço robótico, por sua vez, realiza o movimento de alcance se adequando a posição do objeto na cena conforme se aproxima do mesmo.

## 3.2 Modelo de detecção de objetos

A detecção é uma ferramenta muito importante e útil na robótica. Graças ao uso de técnicas de aprendizagem profunda, a quantidade de dados rotulados presentes em grandes conjuntos de dados de visão computacional disponíveis *online* [77], [79], [80] e o grande empenho da comunidade acadêmica em trabalhos realizados utilizando técnicas de aprendizagem profunda [63], [66], [78], [87], [88], a tarefa de detecção é realizada com sucesso em aplicações de visão computacional.

Para alcançar uma alta velocidade de detecção sem comprometer a resolução temporal da câmera baseada em eventos, adaptamos um modelo de aprendizagem profunda baseado no trabalho produzido pelo Joseph Redmon em *You Only Look Once: Unified, real-time object detection (YOLO)* para detectar objetos em ambientes complexos e em diferentes perspectivas [87]. O modelo desenvolvido para esse trabalho recebe como entrada da rede um pacote de informações baseada em eventos agrupados no tempo, de maneira a converter as informações esparsas das câmeras baseadas em eventos em representações densas.

### 3.2.1 Conversão de informações esparsas em informações densas

O processo de conversão das informações esparsas (assíncronas no tempo e espaço) em superfícies de eventos (representação síncrona dos eventos em janelas de tempo) é uma

abordagem utilizada para adequar a natureza assíncrona dos eventos a um algoritmo síncrono, como é o caso do modelo utilizado para detecção de objetos.

Nesse processo, os eventos, esparsos no tempo, são contabilizados e agrupados em janelas temporais de informações, formando uma espécie de *cluster* temporal de eventos. Na literatura é possível encontrar diversos trabalhos acerca de maneiras para realizar o processo de conversão com a menor perda de informação possível [20], [21], [82], [89]–[91]. Para o modelo utilizado nesse trabalho foi adotado a metodologia de conversão que integra uma janela temporal de eventos ( $W_k$ ) onde a intensidade ( $I$ ) em um pixel na posição  $S(x, y)$  é resultado do processo de integração da polaridade  $P \in \{-1, 1\}$  naquela posição espacial. A equação que rege o processo de geração dessa Superfície de Eventos é dada pela Equação 2:

$$I_k = \sum_{e_j \in W_k} \delta(I_{S(x,y)} - P_{S(x,y)j}) \quad (2)$$

### 3.2.2 Arquitetura do modelo

Como modelo para detecção de objetos a partir das informações dos eventos gravados pela DVS128, foi utilizado um modelo baseado no trabalho *You Only Look Once (yolo): Unified, real-time object detection* [87] e adaptado para trabalhar com as informações densas geradas pelo processo de conversão descrito na seção 3.2.1. A arquitetura presente na

Tabela 3 representa o modelo utilizado nesse trabalho. O modelo consiste em uma rede convolutiva treinada para determinar a posição e a região que o objeto ocupa na cena. O processo de treinamento foi realizado com uma matriz de entrada de 128x128 e treinamos nosso modelo usando 1000 épocas e um tamanho de *batch* de 16.

Tabela 3: **Arquitetura do modelo de aprendizagem profunda para detecção de objetos** (baseado em [87]). O modelo desenvolvido é utilizado para construir o identificador de objetos baseado em informações gravadas com a DVS128.

Índice	Quantidade	Nome	Origem	Parâmetros
0	1	Focus	-	[64, 3]
1	1	Conv	0	[128, 3, 2]
2	3	BottleneckCSP	1	[128]
3	1	Conv	2	[256, 3, 2]
4	9	BottleneckCSP	3	[256]
5	1	Conv	4	[512, 3, 2]
6	9	BottleneckCSP	5	[512]
7	1	Conv	6	[1024, 3, 2]
8	1	SPP	7	[1024, [5, 9, 13]]
9	3	BottleneckCSP	8	[1024, False]
10	1	Conv	9	[512, 1, 1]
11	1	Upsample	10	[None, 2, 'nearest']
12	1	Concat	11,6	[1]
13	3	BottleneckCSP	12	[512, False]
14	1	Conv	13	[256, 1, 1]
15	1	Upsample	14	[None, 2, 'nearest']
16	1	Concat	15,4	[1]
17	3	BottleneckCSP	16	[256, False]
18	1	Conv	17	[256, 3, 2]
19	1	Concat	18,14	[1]
20	3	BottleneckCSP	19	[512, False]
21	1	Conv	20	[512, 3, 2]
22	1	Concat	21,10	[1]
23	3	BottleneckCSP	22	[1024, False]

A arquitetura do modelo utilizado consiste em seis diferentes processos, descritos abaixo:

1. **Focus:** O elemento de entrada do modelo de detecção de objetos é um modelo de transformação de formato da informação. Ele consiste em um processo conhecido como *SpaceToDepth* [92] e transforma a informação de entrada de três camadas para 64 camadas comutando a informação espacial em camadas organizadas (responsável por diminuir o nível de processamento do modelo).

2. **Conv:** A camada Conv é uma abreviação para Convolução. As camadas convolutivas de um modelo de aprendizagem profunda são responsáveis por extrair características da imagem através de *kernels* (matrizes com valores que representam os pesos de um atributo e são modificados através da técnica de *backpropagation* durante o treinamento).
3. **BottleneckCSP:** Essa camada é referente a um processo de distribuição computacional para modelos de aprendizagem profunda que processam imagens. Essa camada é uma abreviação para *Cross Stage Partial Network* [93] e é responsável por dividir o montante de informação e dividir o processamento delas nas diversas unidades de processamento de uma CPU ou GPU. Essa etapa, além de acelerar o processamento das informações, ainda apresenta um aumento da acurácia do modelo por concatenar atributos extraídos em diferentes estágios da rede.
4. **SPP:** A camada é uma abreviatura para *Spatial Pyramid Pooling* [94] e é uma camada responsável por realizar uma transformação no formato dos dados de maneira que eles fiquem menos susceptíveis a variação no tamanho e escala dos objetos. Inicialmente, essa camada foi proposta para criar um mecanismo de adequação para que qualquer tamanho de imagem pudesse ser utilizado na entrada de redes convolucionais, sem comprometer a integridade da informação ao realizar operações de recorte/alteração de tamanho na imagem de entrada. Entretanto, no modelo utilizado, ela é responsável por conferir certa robustez a variação de escala aos atributos extraídos pelas camadas convolucionais.
5. **Upsample:** Essa camada é responsável por criar cópias redimensionadas dos atributos aprendidos pela rede de maneira a aumentar a variação da informação processada e melhorar a acurácia da rede.
6. **Concat:** Essa camada é uma abreviação para “concatenar” que, por sua vez, realiza a concatenação das informações.

### 3.2.3 Conjunto de dados

O conjunto de dados utilizado para treinar o modelo de detecção foi gravado utilizando uma DVS128, e é constituído de sete diferentes objetos gravados a partir de várias perspectivas, gerando diferentes padrões de geração de eventos para o mesmo objeto. O arranjo experimental para gerar o conjunto de dados, conta com a DVS128 acoplada na extremidade de um braço robótico comercial - Widow X (Figura 10) e um cenário complexo de gravações, ou seja, com diversas informações provenientes de móveis e equipamentos presentes ao fundo de cada gravação.

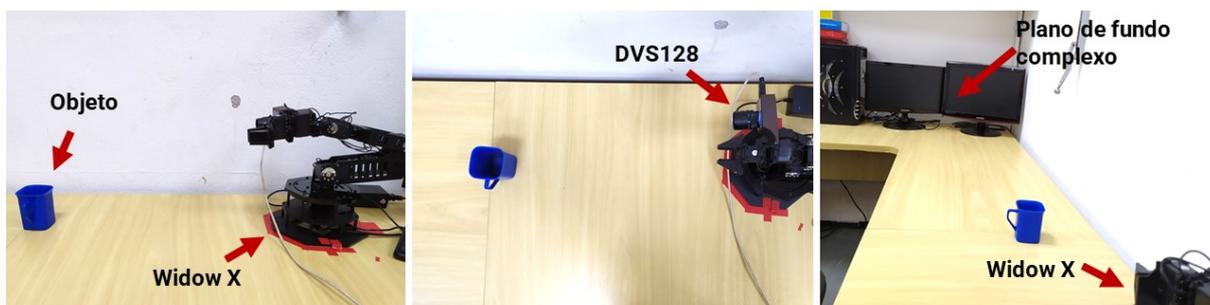


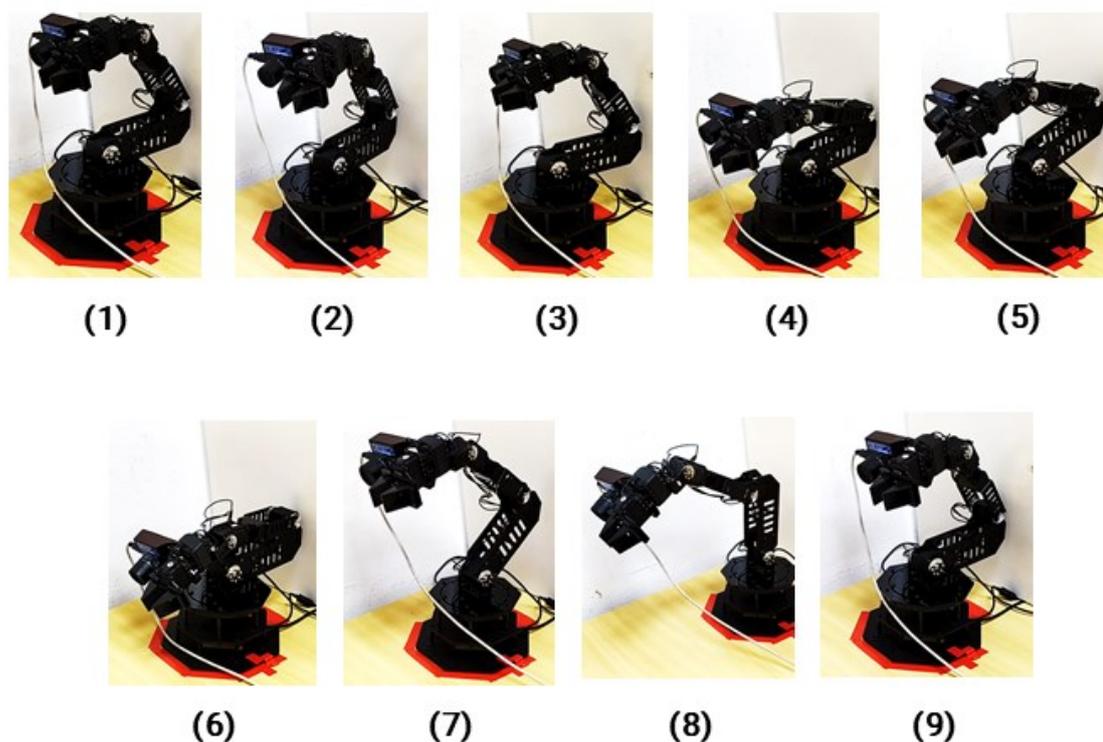
Figura 10: **Demonstração do ambiente e disposição dos componentes utilizados para gravação do conjunto de dados.** As imagens ilustram o braço robótico (Widow X) utilizado para realizar o movimento pela cena, a disposição dos objetos de interesse em relação ao Widow X e a localização da câmera baseada em eventos (DVS128).

Os objetos selecionados para contemplar o conjunto de dados (Figura 11) desenvolvido nesse trabalho foram escolhidos devido sua presença no cotidiano, sendo eles: a) Banana, b) Copo, c) Garfo, d) Chave, e) Faca, f) Caneca e g) Laranja



Figura 11: **Objetos utilizados para compor o conjunto de dados desenvolvido no trabalho.** A imagem consiste em todos os 7 objetos utilizados, apresentando a imagem do objeto adquirida com uma câmera tradicional.

Para gerar as diferentes perspectivas de visão do objeto, o WidowX foi programado para se mover em um percurso determinado, de maneira que o objeto não saísse do campo de visão da câmera e os padrões de geração de eventos para o mesmo objeto fosse variado ao longo do percurso (Figura 12). O rótulo de cada um dos objetos foi gerado individual e manualmente conforme sua classe e posição do objeto na tela delimitado por um retângulo na cena (caixa delimitadora, em inglês *Bounding Box*).



**Figura 12: Representação do movimento do Widow X ao longo do tempo para registro das diferentes perspectivas dos objetos utilizados para compor o conjunto de dados.** A sequência numérica de (1) a (9), representa respectivamente os movimentos realizados pelo Widow X durante a gravação dos eventos.

Cada gravação realizada foi submetida a um processo de conversão das informações esparsas (assíncronas no tempo e espaço) em Superfícies de Eventos (representação síncrona dos eventos em janelas de tempo). Esse processo se faz necessário, haja vista que o parâmetro utilizado de entrada do modelo de aprendizagem profunda requer informações densas.

### 3.3 Sistema de rastreo em tempo real

O sistema de rastreo em tempo real, consiste em utilizar as informações do modelo de detecção de objetos para promover a movimentação do braço robótico em direção ao objeto de interesse. Dessa maneira, a malha de controle de acionamento e posição dos motores do braço robótico é apresentada na Figura 13 e consiste em uma malha de controle proporcional que utiliza como referência a posição central da cena vista pelo sistema de Visão Neuromórfico e como variável de processo a posição do centroide da detecção realizada pelo modelo baseado no *yolo*.

Logo, o sinal de erro é considerado como diferença entre a posição dos centroides e o centro do campo de visão da DVS128. O valor do erro é então multiplicado pela constante  $K_p$  ( $K_{px} = 9.54$  e  $K_{py} = 22.8$ ) e somado ao valor da posição dos motores, e o braço robótico é atuado de maneira que a posição relativa entre o objeto de interesse a câmera é modificada, realimentado o sistema.

O sistema de controle leva em consideração a caracterização da saída do modelo de detecção, possibilitando três cenários diferentes: Detecção Válida (DV), Detecção Inválida (DI) ou Movimento Sacádico (MS). Para ser considerado uma DV o modelo precisa realizar uma comparação entre a posição do centroide da detecção anterior e a posição do centroide da detecção atual ( $f(s)$  na Figura 13). Nesse caso, é realizado o cálculo da distância euclidiana entre esses dois pontos e é verificado se a distância entre eles é menor que o limiar de distância máxima (parâmetro de entrada do sistema = 18px). Caso contrário, a detecção é caracterizada como DI. No caso de o modelo não encontrar nenhuma detecção, o sistema inicia o modo de MS que consiste em movimentos verticais rápidos e de baixa amplitude para gerar movimento relativo entre a câmera e o objeto [84]–[86].

Em cada caso, a malha de controle recebe (como variável de processo) um valor de centroide de uma detecção, que por sua vez é responsável pela movimentação do braço robótico nos eixos x e y, enquanto o eixo de avanço, em direção ao objeto de interesse, é incremental ( $c = 2$ ). Quando a detecção é válida, o sistema realiza a comparação da referência com a detecção atual e é realizado o incremento no eixo z para que o braço se aproxime do objeto de interesse. Entretanto, quando a detecção é inválida ou é realizado o movimento sacádico, o valor que é

comparado com a referência é ele mesmo e, portanto, não há sinal de controle, nem avanço. A regra descrita é apresentada em  $D(s)$  na Figura 13.

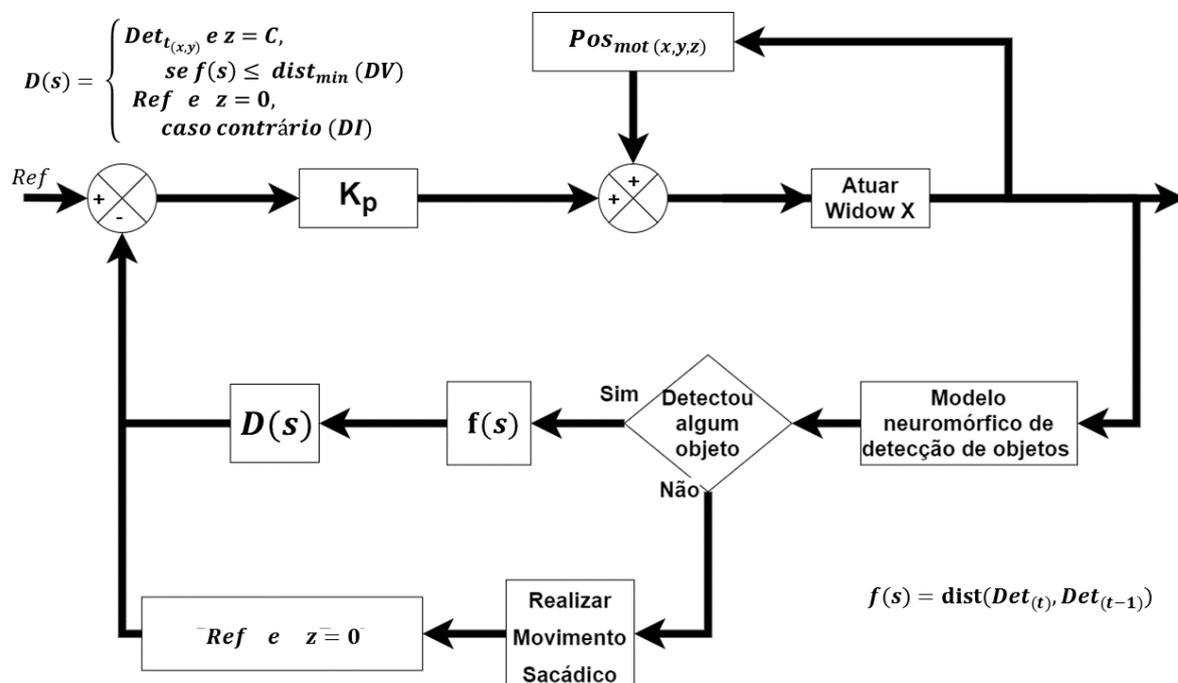


Figura 13: **Malha de controle proporcional desenvolvida para concretização do sistema de rastreamento em tempo real baseado no modelo de detecção de objetos para alcance.** O modelo utiliza como valor de referência a posição do centro da imagem, posteriormente é utilizado o centroide da detecção válida para mensurar o valor do erro entre o valor de referência e o valor da posição do objeto na tela. A posição dos motores é atualizada com base no valor do erro multiplicado por um fator  $K_p$  e acrescido na posição atual dos motores para as posições  $x$  e  $y$  (horizontal e vertical, respectivamente). Para o movimento no eixo  $z$  (longitudinal/avanço) o valor de  $z$  é acrescido de uma constante a cada detecção realizada ( $c = 2$ ). As detecções realizadas pelo modelo se distinguem em Detecções Válidas (DV) e Inválidas (DI). As detecções válidas são detecções que possuem distância euclidiana entre a detecção atual e a anterior menor que um limiar (comparação realizada a partir da segunda detecção do modelo), enquanto as detecções inválidas são as detecções que superam esse limiar de distância. Quando não há detecção, o sistema não atua sobre a posição dos motores e um Movimento Sacádico (MS) é utilizado para gerar eventos que são utilizados para fomentar o modelo de detecção de objetos.

### 3.4 Procedimentos experimentais de avaliação e validação

Para avaliar e validar o modelo proposto, foi utilizado uma câmera baseada em eventos

(Dynamic Vision Sensor – DVS128) e um braço robótico comercial (*WidowX Robot Arm Mark II*). O sistema proposto conta com a câmera baseada em eventos acoplada na lateral da garra do WidowX. Dessa maneira, as informações lidas pela câmera são enviadas via comunicação USB 2.0 para um computador que processa essas informações através do modelo discutido nas sessões 3.2 e 3.3, desse capítulo, e atua sobre a posição dos motores do WidowX promovendo o movimento do mesmo em direção ao objeto de interesse (Figura 14).

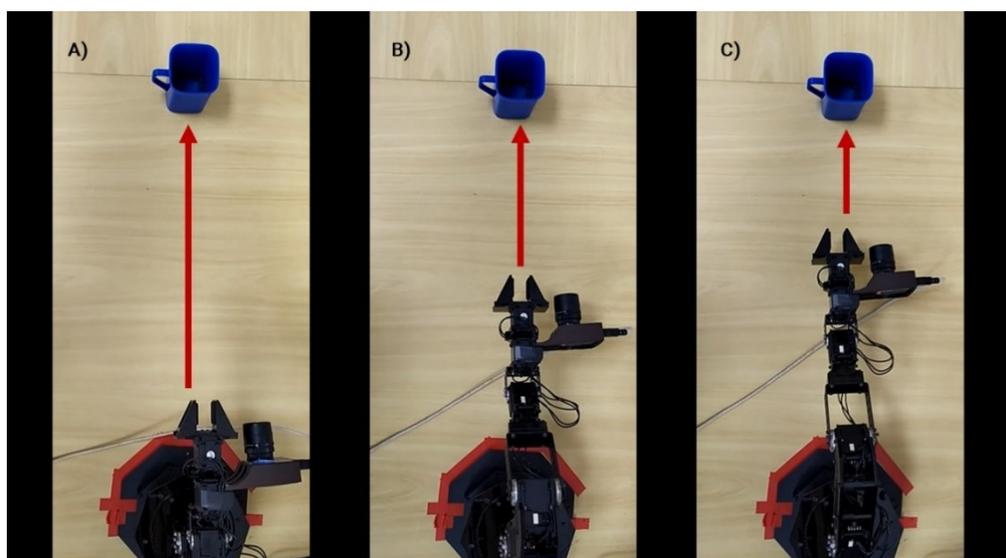


Figura 14: **Processo de movimentação do WidowX em direção ao objeto de interesse.** No exemplo da imagem o WidowX se move na direção do eixo z de maneira a se aproximar do objeto.

O computador utilizado para aquisição e processamento das informações do sistema de rastreamento em tempo real possui um processador Intel® Core™ i7-8700K, 16 GB de memória RAM e uma placa de vídeo NVIDIA GeForce RTX 2070. Todos os algoritmos desenvolvidos para a realização desse trabalho foram desenvolvidos utilizando python 2.7. Em tempo de execução, os processadores eram responsáveis pela aquisição de dados da câmera baseada em eventos, a comunicação serial com o WidowX e o execução do sistema de rastreamento, em tempo real.

O WidowX (Figura 15) é um braço robótico com seis graus de liberdade, entre eles uma base que gira 360° em torno de si própria. Seu alcance é de 45 cm do centro da base (sendo esse o comprimento máximo do braço robótico). Uma vez retraído, o WidowX não ocupa mais do

que o tamanho de sua base (Figura 15. B) que, por sua vez, possui 14 cm de diâmetro.

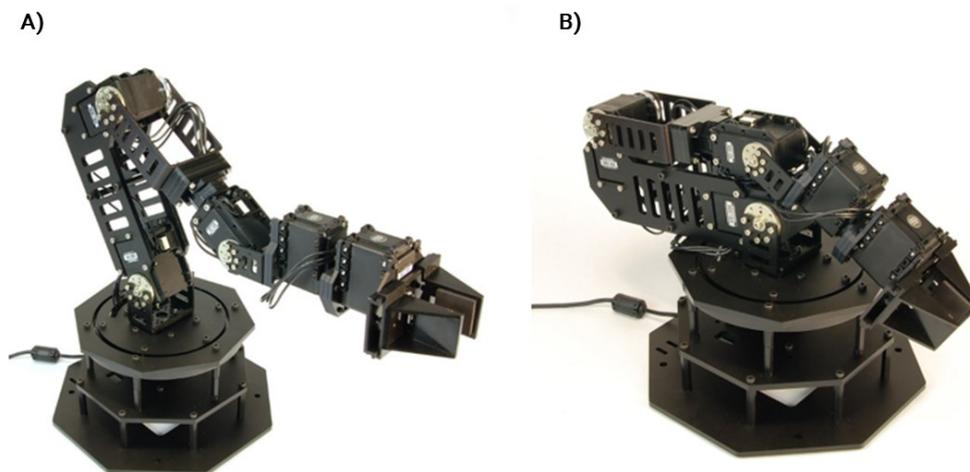


Figura 15: **Braço robótico (*WidowX Robot Arm Mark II*)**. Desenvolvido pela Trossen Robotics e utilizado nos experimentos para avaliação e validação do sistema semiautônomo de alcance para próteses de membro superior.

Por fim, para avaliar e validar o sistema de alcance proposto foi realizado cinco experimentos, que são descritos a seguir:

### 3.4.1 Experimento 1

Em uma cena simples (sem eventos originários da variação de luminosidade de ambiente promovido pelo movimento relativo da câmera), o modelo precisa identificar o objeto em diferentes padrões de movimento gerando uma condição simples para o desafio de associação de dados. Nessas condições, o padrão de eventos do objeto é diferente, dada a direção do movimento da câmera, porém o modelo não precisa distinguir os eventos do objeto em relação aos eventos do ambiente ao redor visto que não há informações provenientes do plano de fundo da cena.

No primeiro experimento (Figura 16), o objetivo é alcançar um único objeto sobre a mesa. Dessa maneira, o objeto de interesse (chave) foi posicionado a uma distância fixa (60 cm da base). Esse experimento foi desenhado para analisar a capacidade de alcance e rastreamento do

modelo com apenas um objeto em uma cena simples (sem informações de plano de fundo).

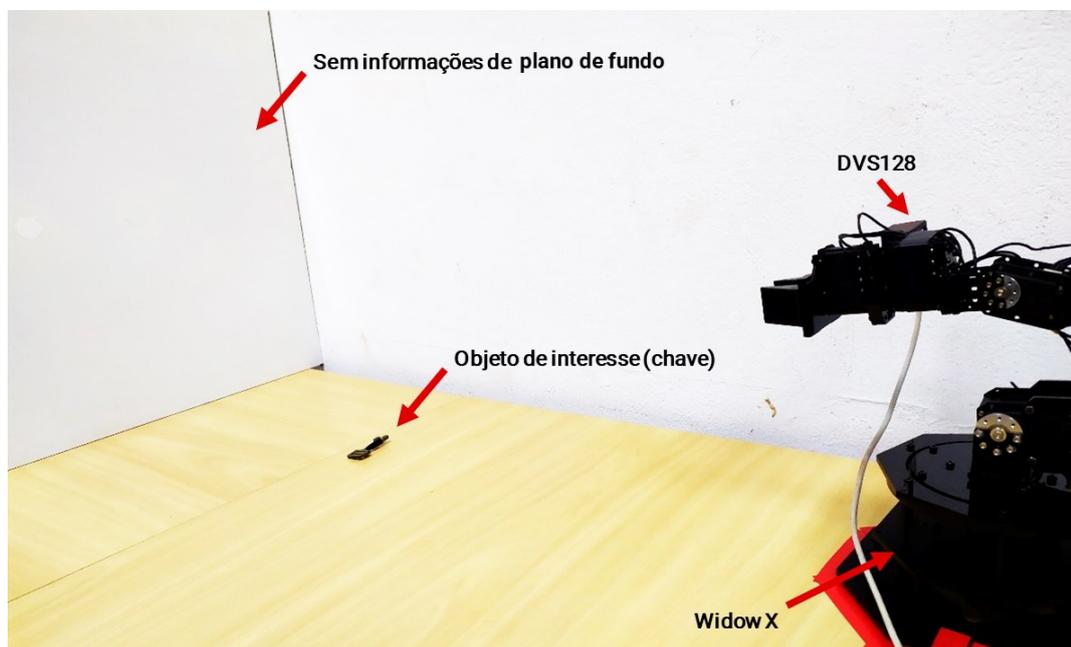


Figura 16: **Arranjo experimental para a realização do Experimento 1.** O objeto de interesse (chave), posicionada a frente do Widow X (que por sua vez possui a DVS128 acoplada a lateral de sua extremidade), com um fundo branco para retirar as informações de plano de fundo do cenário e promover um ambiente com pouco ruído fundo.

### 3.4.2 Experimento 2

Para o segundo experimento (Figura 17), manteve-se o ambiente, sem informações de plano de fundo (conforme o descrito no Experimento 1), porém com o adicional de um novo distrator para o modelo. Nesse experimento, será analisada a capacidade do modelo de distinguir entre três objetos distintos na cena e realizar o alcance do objeto de interesse (localizado no centro da cena). Os objetos utilizados para esse experimento foram a chave, a faca e a laranja, sendo que o objeto de interesse do modelo era a chave.

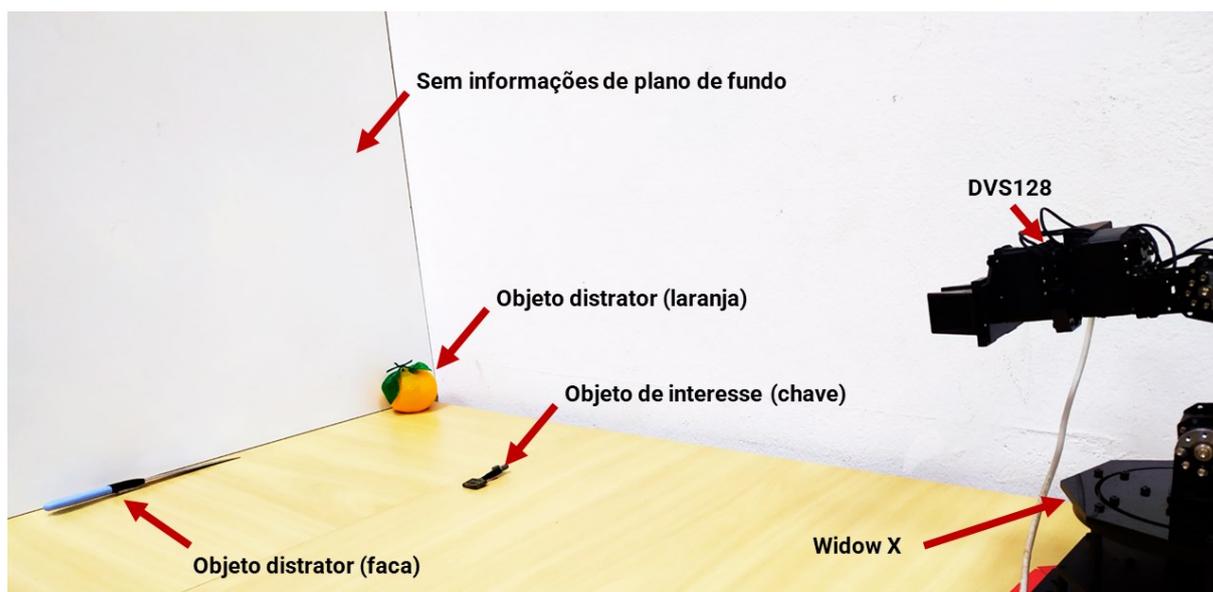


Figura 17: **Arranjo experimental para a realização do Experimento 2.** Os objetos presentes na cena são a chave (objeto de interesse), a faca e a laranja (objetos distratores). O WidowX em posição inicial, com a DVS128 acoplada a sua garra, e o fundo branco para diminuir as informações de plano de fundo presentes nas gravações.

### 3.4.3 Experimento 3

O terceiro experimento (Figura 18), consiste em adicionar as informações de plano de fundo ao Experimento 1. Dessa maneira, tem-se por objetivo testar a robustez do modelo frente a um cenário complexo, onde o modelo deve ser capaz de manter o rastreamento dos eventos originários do objeto de interesse enquanto o braço navega em direção ao mesmo. Para o Experimento 3, a caneca foi posicionada a 60 cm da base do WidowX e o plano de fundo será mantido inalterado, conforme uma mesa de escritório convencional.

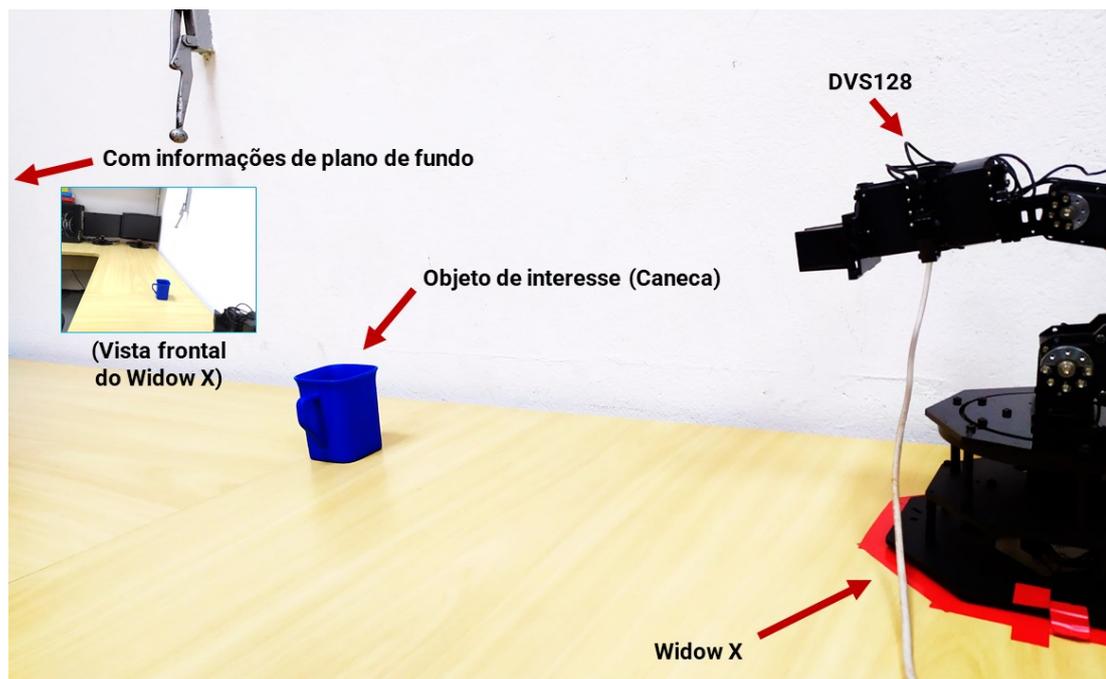


Figura 18: **Arranjo experimental proposto para realização do Experimento 3.** O objeto de interesse (caneca) é posicionado a 60 cm da base do WidowX e o fundo do ambiente é preservado com os móveis e equipamentos naturais de um laboratório.

#### 3.4.4 Experimento 4

O experimento quatro (Figura 19) se baseia na verificação da capacidade do modelo de distinguir os eventos de um objeto de interesse em um cenário com outros dois objetos e eventos originários das informações de plano de fundo. Nesse experimento o objeto de interesse será o objeto central na disposição dos objetos na mesa e o modelo deverá ser capaz de associar os eventos desse objeto e manter o rastreamento em detrimento dos eventos dos demais objetos e do plano de fundo. Sendo assim, constitui-se o cenário mais desafiador dos experimentos aqui listados.

O objeto utilizado nesse experimento será a caneca que estará a uma distância de 60 cm da base do WidowX, enquanto os outros objetos (copo e laranja) estarão dispostos na lateral da cena.

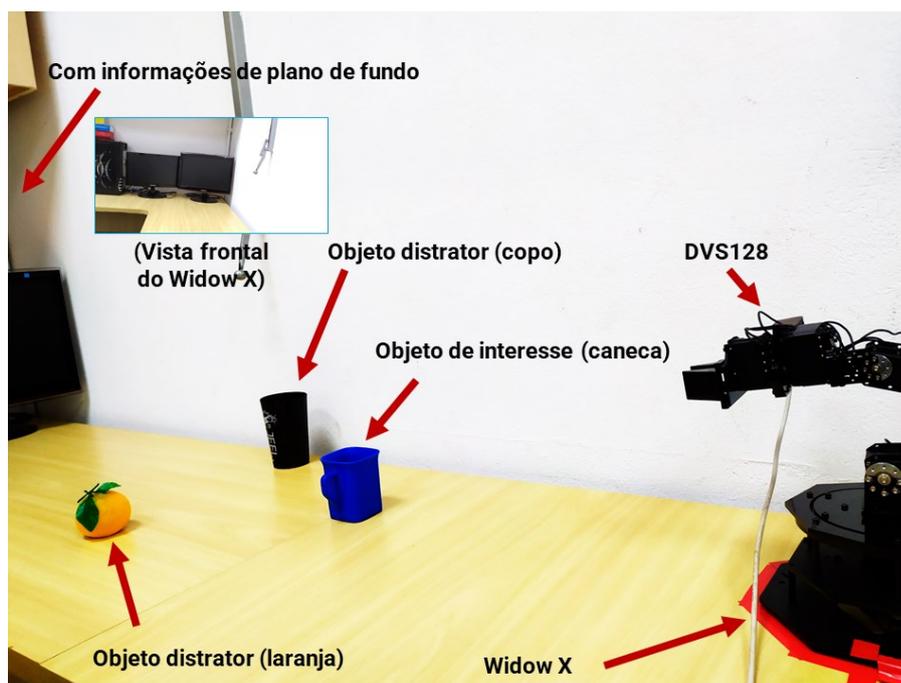


Figura 19: Arranjo experimental utilizado para o Experimento 4 de validação do modelo de alcance para próteses de membro superior. Os objetos em cena são: caneca (objeto de interesse), laranja e copo (objetos distratores). As informações de plano de fundo foram mantidas sem intervenção.

### 3.4.5 Experimento 5

O último experimento consiste em analisar a capacidade de adaptação do modelo frente a movimentação inesperada de um objeto em uma cena complexa. Para realizar esse experimento o objeto de interesse (caneca) estará posicionado em uma região central do campo de visão da câmera neuromórfica e, dois segundos após o início do movimento, a localidade do objeto de interesse (Posição A) será alterada para um extremo do campo de visão (Posição B), ao longo do processo de alcance e o modelo deverá se adaptar a essa alteração de posição (conforme ilustrado na Figura 20).

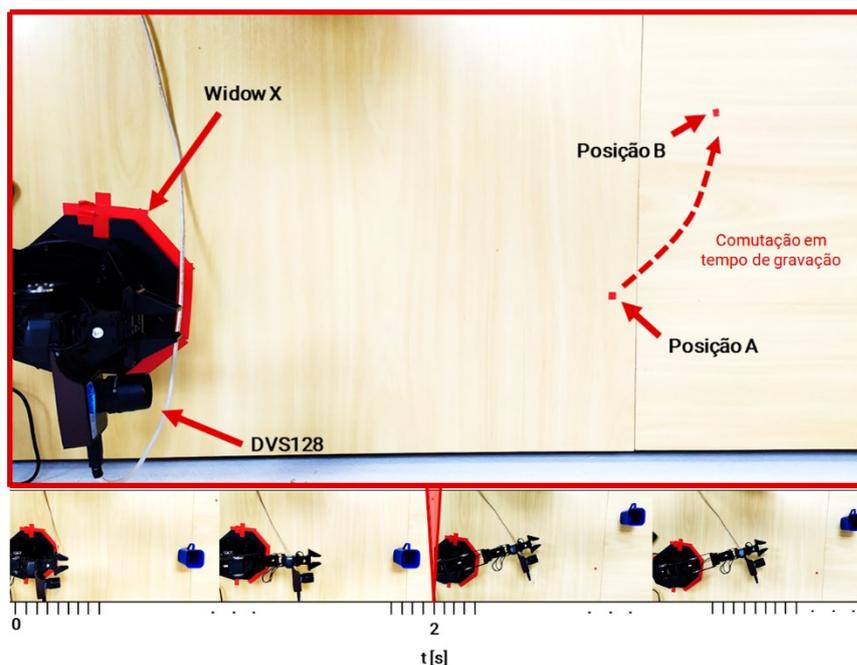


Figura 20: **Disposição do objeto para a realização do Experimento 5.** O objeto de interesses (caneca) é posicionado previamente em uma região central do campo de visão da câmera baseada em eventos e posteriormente (2 segundos após o início do alcance) é movido para a região da extrema direita do campo de visão da câmera.

### 3.4.6 Análise de dados

O modelo utilizado para detecção de objetos foi treinado utilizando 80% do conjunto de dados e validado com 20% do montante total de informações adquiridas. Dessa maneira, a análise do desempenho do modelo de detecção se baseia na Matriz de Confusão, onde é possível observar o desempenho do sistema de detecção classe a classe. Nela é possível verificar se o sistema foi capaz de determinar a classe do objeto corretamente e qual a influência das demais classes sobre cada classe individualmente, além de proporcionar uma visão ampla do desempenho geral do modelo.

Para determinar como o modelo desempenha em precisão (Equação 3) em relação a confiança do modelo na predição realizada, foi utilizada a curva-P, métrica responsável por ilustrar a performance do modelo em termos de precisão versus confiança da predição.

$$P = \frac{V_p}{T_p} \quad (3),$$

onde  $P$  é a precisão do modelo, enquanto  $V_p$  é a quantidade de Verdadeiros Positivos, ou seja, as detecções realizadas que foram corretas e o valor de  $T_p$  é o total de predições realizadas pelo

modelo.

A métrica de sensibilidade do modelo (Equação 4) é a relação entre a quantidade de predições Verdadeiras Positivas ( $V_p$ ) e a quantidade de elementos da mesma classe de análise presentes na cena (Verdadeiros Positivos + Falsos Negativos).

$$S = \frac{V_p}{(V_p + F_N)} \quad (4),$$

onde,  $S$  é a sensibilidade do modelo à classe de análise,  $V_p$  é a quantidade de detecções Verdadeiro Positivo e  $F_N$  é a quantidade de detecções Falso Positivo.

O desempenho em termos de acurácia do modelo foi determinado a partir da análise do *F1-Score* (Equação 5), métrica que determina uma correlação entre a precisão e a sensibilidade do modelo às classes.

$$F_1 = \frac{2}{\left(\frac{1}{S} + \frac{1}{P}\right)} = \frac{V_p}{\left(V_p + \frac{1}{2}(F_p + F_n)\right)} \quad (5),$$

A curva de Precisão Média (mAP) do modelo foi utilizada para mensurar o desempenho do modelo em detectar objetos em cena e determinar a localização e os limites que o mesmo ocupa em cena. Para tanto, foi considerado como uma detecção válida, aferições que possuíam uma área de interseção de 50% (0,5 IoU), ou seja, quando o modelo prediz uma região delimitadora cuja área intersecta a área rotulada em 50%.

A análise do desempenho do sistema de rastreo, frente aos cinco cenários experimentais propostos na seção 3.4, foi realizada considerando os quatro primeiros experimentos como fonte de informação acerca do comportamento do sistema ao longo do movimento de alcance, enquanto os dados extraídos do quinto experimento leva em consideração a capacidade do mesmo de se adaptar frente a perturbações. Cada experimento foi realizado cinco vezes ( $M = 5$ ) e os dados acerca de cada repetição foi coletado para a análise.

Para comparar o desempenho geral do sistema frente aos diferentes cenários propostos nos primeiros quatro experimentos, foi analisada a porcentagem média do percurso em que o sistema foi capaz de determinar o objeto de interesse em detrimento dos demais objetos/informações em cena. Essa métrica de acurácia do sistema de rastreo foi determinada pela razão de detecções válidas pela quantidade de interações necessárias para realizar o alcance

(Equação 6).

$$Acc = \frac{1}{M} \sum_{m=1}^M \frac{1}{N_m} \sum_{n=1}^N C_{mi}, C_{mi} \in D_v \quad (6),$$

onde M é a quantidade de repetições realizadas para cada experimento, N é a quantidade de interações realizadas em cada movimento de alcance realizado e  $C_{mi}$  é a saída determinada pelo sistema a cada interação e  $D_v$  representa o cenário em que ocorreu uma detecção válida.

Dessa maneira é possível analisar o desempenho, em termos de acurácia, durante o movimento de alcance, para estimar como o sistema se comporta ao longo do tempo nos diferentes cenários propostos. Para realizar a análise temporal da acurácia do sistema ao longo do tempo, foi verificada uma taxa média de acertos em janelas de 5 detecções, conforme Equação 7.

$$Tx_{acc} = \frac{K_{dvi}}{N}, K_{dvi} \in \{5, 10, 15 \dots\} \quad (7),$$

onde  $K_{dvi}$  representa uma constante incremental referente a condição de haver cinco detecções válidas e N é a quantidade de interação necessária até que a condição  $K_{dvi}$  seja atendida.

Como o sistema de alcance estima três cenários diferentes a cada interação, podendo ser eles uma detecção válida ( $D_v$ ), uma detecção inválida ( $D_i$ ) ou a não detecção que, por sua vez, resulta em uma ação de movimento sacádico ( $M_s$ ) responsável por gerar eventos, foi mensurada a porcentagem de cada um dos três possíveis cenários, ocorrendo ao longo do movimento de alcance.

O movimento sacádico realizado pelo WidowX (para gerar movimento relativo entre o objeto e a câmera baseada em eventos) representa um intervalo de tempo do tempo total gasto para realizar o alcance. Dessa maneira, foram metrificadas as variáveis temporais que estão relacionadas ao tempo gasto, realizando o movimento sacádico durante o movimento de alcance. Dessa maneira, é analisado, ao longo do movimento de alcance realizado pelo sistema proposto, a quantidade de tempo médio em que o sistema esteve incerto acerca da posição do objeto e realizou o movimento sacádico para avaliar a cena (Equação 8).

$$\tau_{ms} = \frac{1}{M} \sum_{m=1}^M \frac{1}{Z} \sum_{z=1}^Z (t_f - t_i)_z \quad (8),$$

onde M é a quantidade de repetições realizadas para cada experimento, Z é a quantidade de movimento sacádicos realizados em cada experimento,  $t_i$  é o instante de tempo em que o movimento sacádico começou, enquanto o  $t_f$  é o instante de tempo em que o movimento

finalizou.

No quinto experimento realizado, o interesse é determinar a capacidade de adaptação a uma mudança repentina de posição do objeto. Para metrificar o desempenho do sistema frente a alteração de posição do objeto, a abordagem é utilizar uma comparação das mesmas variáveis de desempenho obtidas no Experimento 3 com as obtidas no Experimento 5. Dessa maneira, tem-se uma análise direta entre o cenário de um objeto estático e um objeto dinâmico.

---

## Capítulo 4

---

### Resultados e discussão

No Capítulo 4, a seguir, são apresentados os resultados obtidos durante a realização desse trabalho. Após cada resultado, é realizada uma discussão, posicionando o trabalho no estado da arte ao compará-lo com a literatura existente no assunto e apresenta hipóteses acerca dos resultados obtidos.

#### 4.1 Conjunto de dados resultante

O conjunto de dados criado para execução desse trabalho consiste na gravação de sete objetos distintos utilizados no dia a dia (Figura 21), sendo eles: a) Banana, b) Copo, c) Garfo, d) Chave, e) Faca, f) Caneca e g) Laranja. Essas informações foram adquiridas a partir de uma câmera baseada em eventos (DVS128) e consiste em gravações de 24 segundos de duração de cada um dos objetos.

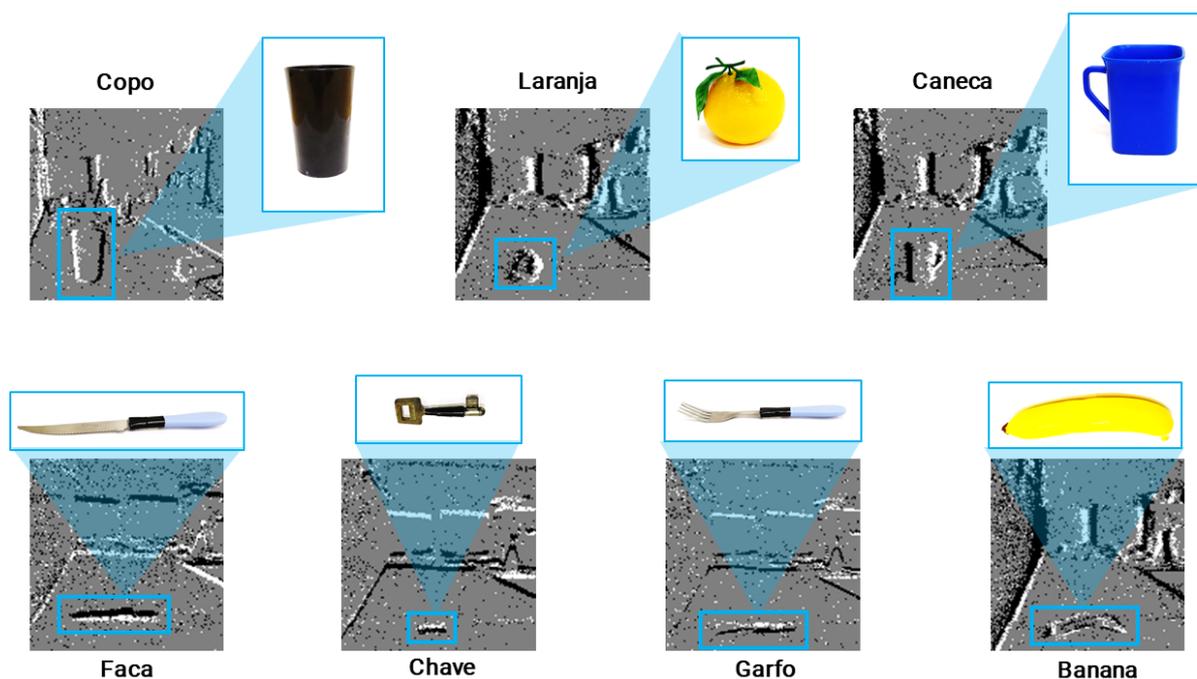


Figura 21: **Representação da superfície de eventos de cada um dos objetos gravados para o conjunto de dados.** Na parte inferior de cada objeto consta a superfície de eventos gerada a partir do processo de conversão de informações esparsas, da câmera baseada em eventos, em informações densas. Acima de cada superfície de eventos é apresentado a imagem do objeto registrado com uma câmera tradicional.

A diversificação e grande variância de representações de um mesmo objeto é de extrema importância para os modelos de aprendizagem profunda [61], [77], [79], [80]. Entretanto, como o campo de visão neuromórfica consiste em um campo recente de pesquisa e não possui conjuntos de dados consolidados, que permitam a pluralidade de representações de diferentes classes, foi necessário o desenvolvimento de um mecanismo de movimentação da câmera em relação ao objeto (Figura 22) para que o conjunto de dados desenvolvido nesse trabalho possua uma maior variedade de perspectivas dos objetos utilizados.

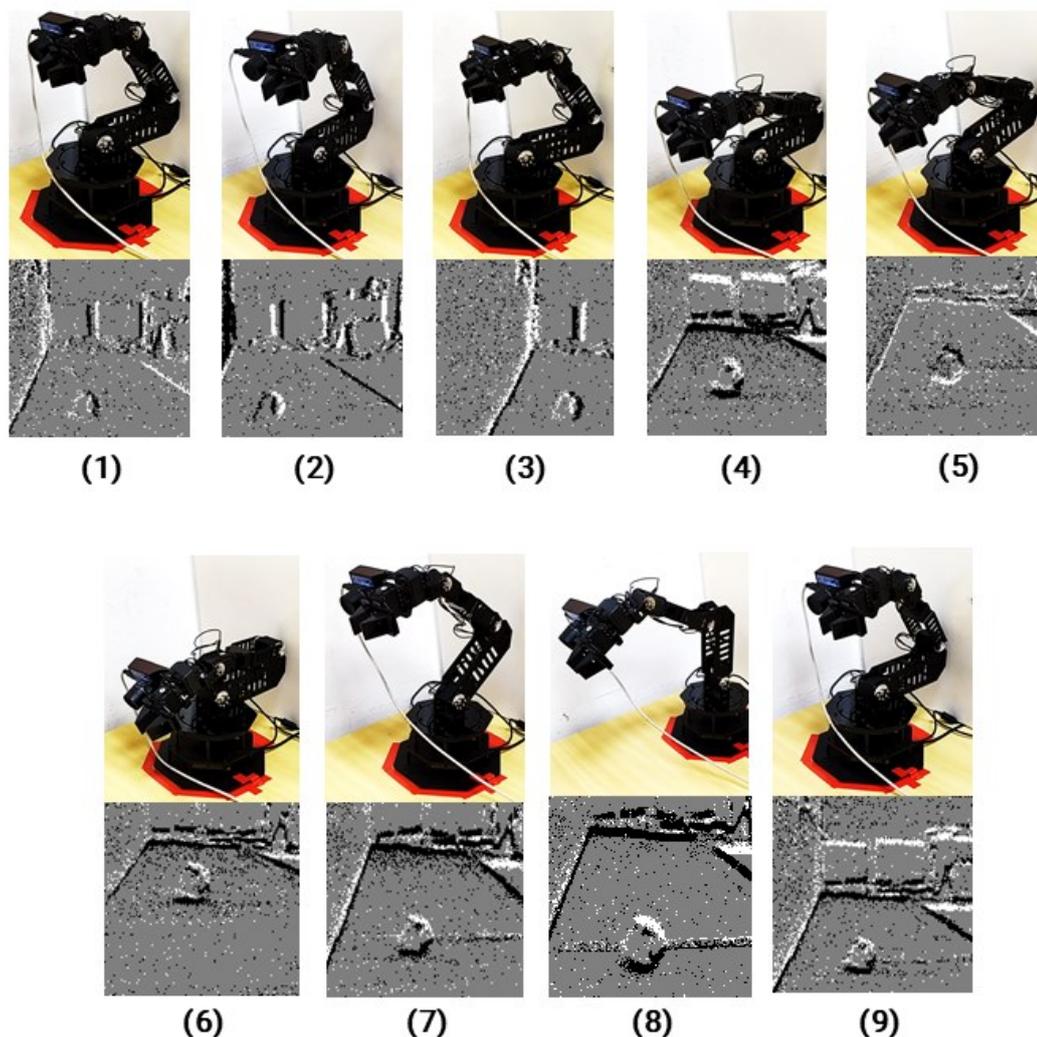


Figura 22: **Representação das superfícies de eventos em diferentes momentos da gravação do conjunto de dados de acordo com o movimento do braço robótico.** Abaixo de cada imagem do WidowX consta a representação dos eventos no momento do movimento realizado demonstrando as diferentes posições/perspectivas dos objetos de interesse.

No início do desenvolvimento dos trabalhos que tentavam realizar inferências em cenas a partir de informações adquiridas utilizando câmeras baseadas em eventos, alguns autores tentaram aproveitar os grandes conjuntos de dados de Visão Computacional [79], [80], [95], [96] que são bem consolidados na literatura para gerar conjuntos de dados pseudo-neuromórficos [72], [76]. Entretanto, os bons resultados no processo de treinamento e teste de modelos de aprendizagem profunda que utilizaram conjuntos de dados pseudo-neuromórficos

são indicativos da boa capacidade dos modelos de aprender a distinguir padrões e não necessariamente do modelo de distinguir informações reais gravadas por câmera baseadas em eventos.

Dessa maneira, para que o modelo performe bem em informações verdadeiramente neuromórficas, é necessário que o conjunto de dados utilizado seja composto por informações gravadas por câmeras baseadas em eventos (como o desenvolvido nesse trabalho). Contudo, quanto maior o conjunto de dados e a variância das representações de uma mesma classe, melhor será a performance dos modelos de aprendizagem profunda. Portanto, são necessários grandes conjuntos de dados verdadeiramente neuromórficos com informações rotuladas, isso pressupõe uma alta carga de trabalho humano para rotular essas informações e alta disponibilidade de câmeras baseadas em eventos para geração das informações.

Todavia, muito tem se discutido sobre maneiras de adequar os modelos de aferição para tirar proveito das informações temporais presentes nas poucas informações verdadeiramente neuromórficas disponíveis [36], [69], [97], além de trabalhos estarem sendo desenvolvidos para que o processo de conversão de informações de Visão Computacional tradicional seja mais fiel a representação real dos eventos de uma câmera neuromórfica [73].

Como o conjunto de dados desenvolvido nesse trabalho foi pensado para a aplicação final de alcance de objetos convencionais do dia a dia por usuários de próteses de membro superior, e não utilizou os métodos de conversão de conjuntos de dados convencionais, foi necessário realizar a rotulação dessas informações (*ground truth*). O processo de determinação da caixa delimitadora (região retangular que confina o objeto e determina sua posição na cena –Figura 23) foi realizado manualmente para cada classe de objetos presentes no conjunto de dados e consiste em cinco informações principais, sendo elas: posição do centroide em x, posição do centroide em y, largura e altura da caixa delimitadora e classe do objeto confinado pela caixa delimitadora.

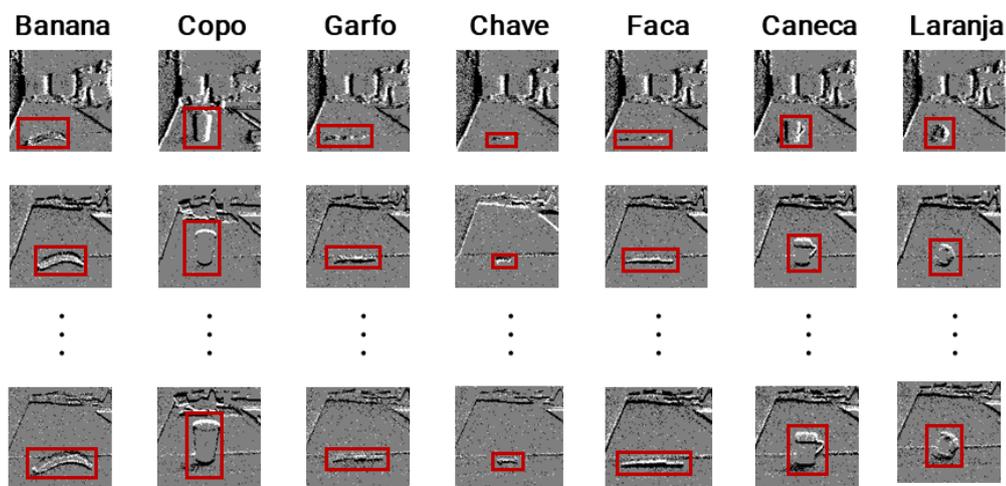


Figura 23: **Resultado do processo de conversão das informações neuromórficas em uma superfície de eventos e do processo de geração dos rótulos da posição de cada classe em sua respectiva gravação.** Cada classe é representada pelo nome acima das superfícies e em vermelho consiste a demarcação da região delimitadora relacionado a posição do objeto na cena em dado instante de tempo.

O intervalo de tempo utilizado no processo de conversão das informações esparsas espaço-temporalmente para informações densas no processo de adequar o conjunto de dados para a entrada do modelo de aprendizagem profunda foi de 100ms, resultando em 208 superfícies de eventos para cada objeto e 1456 superfícies de eventos no total. Os arquivos originais contendo os eventos gerados em cada gravação, juntamente com os arquivos convertidos para representações densa e seu respectivos rótulos foram disponibilizados para a comunidade de pesquisa em Visão Neuromórfica (<https://git.io/JRWw5>).

## 4.2 Modelo de detecção

O modelo de detecção de objetos utilizado nesse trabalho consistiu em um modelo baseado no atual estado da arte em detecção de objetos [87] adaptado para interpretar as informações gravadas pela câmera baseada em eventos (superfície de eventos) e treinado em um conjunto de dados descrito na seção 3.2.3.

A Figura 24 apresenta a Matriz de Confusão do modelo, onde é possível averiguar o desempenho geral do modelo em classificar objetos entre as sete classes presentes no conjunto

de dados. O resultado de todas as classes é superior a 0.8 (80%), exceto pelas classes Chave e Faca que apresentam valores de 0.66 (66%) e 0.67 (67%) respectivamente.

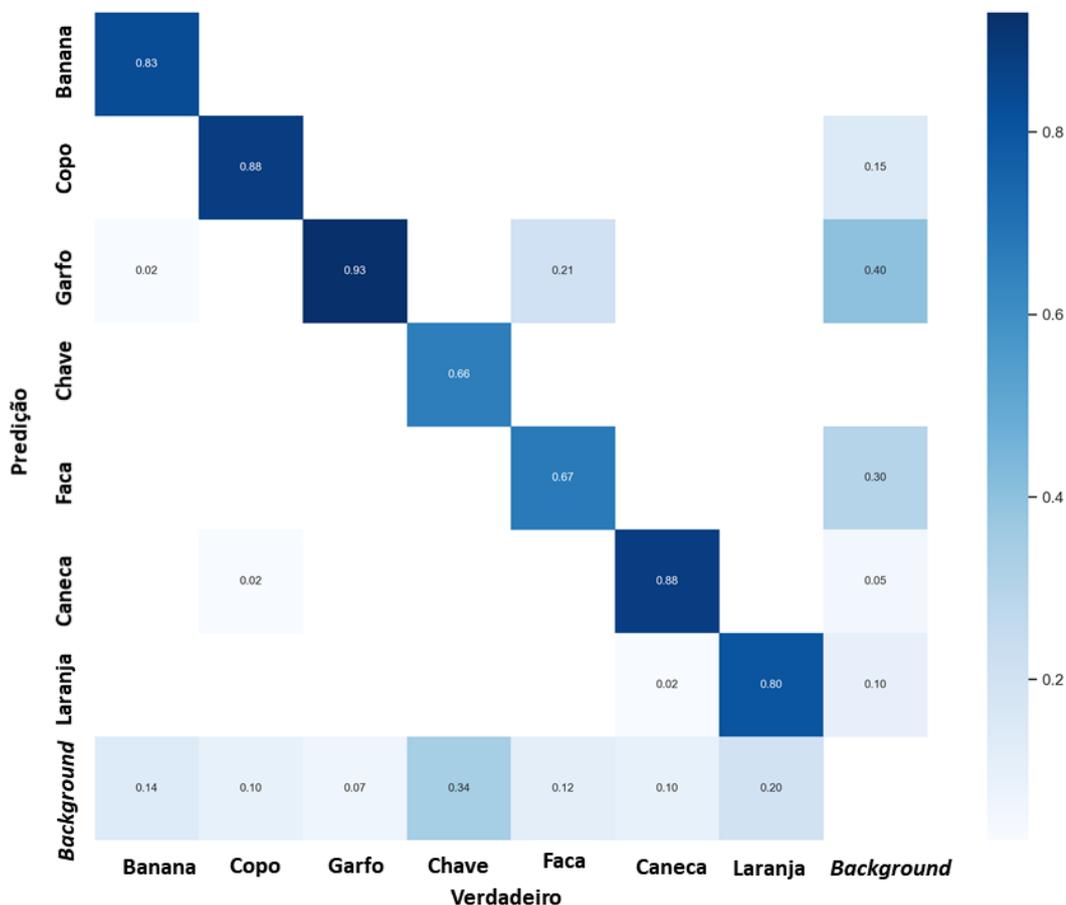


Figura 24: **Matriz de Confusão do modelo de detecção de objetos gravados pela câmera baseada em eventos (DVS128).** A Matriz de Confusão apresenta a estimativa do modelo em relação a um objeto presente na cena (predição) e compara com o valor do rótulo presente no conjunto de dados. A razão entre a quantidade de detecções realizadas com determinada classe em relação ao rótulo é apresentada em escalas de azul e quanto mais próximo de 1 mais escuro é o tom azul (todas as detecções foram corretas) e quanto mais próximo de zero, mais claro o tom de azul (todas as detecções foram erradas).

A Chave consiste em um objeto muito pequeno (Figura 25) e a baixa resolução da DVS128 pode ter contribuído para o, relativo, baixo desempenho de classificação do modelo para essa classe visto que ela é altamente confundida com informação de plano de fundo devido ao fato de que as informações neuromórficas são conhecidas por serem altamente ruidosas [98].

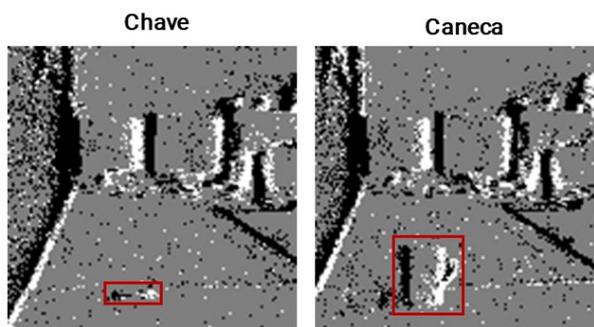


Figura 25: **Comparação entre o tamanho e formato da classe Chave com a classe Caneca.**

A diferença de tamanho entre as duas classes permite com que o modelo apresente uma menor capacidade discriminatória da classe Chave em relação a classe Caneca

A predição de objetos como sendo Garfo ao invés de Faca se baseia no fato de que as características de uma Faca são muito similares ao Garfo o que pode levar o modelo a realizar a classificação de uma Faca como sendo um Garfo com baixa confiança de detecção, entretanto o inverso não ocorre, devido as características específicas de um Garfo que uma Faca não possui como as pontas múltiplas (Figura 26). Entretanto, a performance do modelo e sua capacidade de generalização de classes como Chave e Faca poderia ser melhorado com um conjunto de dados maior, mais diversificado [61].

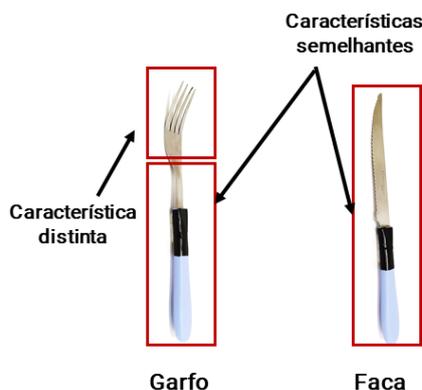


Figura 26: **Comparação entre a morfologia dos objetos contidos na classe Garfo e Faca.**

A Faca possui um corpo prolongado como uma haste simples, enquanto o garfo, além de possuir as mesmas características ainda possui uma região de múltiplas hastes em uma de suas extremidades e, dentro de um contexto de informações baseadas em eventos onde não há alta resolução espacial e distinção de cor, o modelo fica sujeito a pouca capacidade discriminatória da Faca em relação ao Garfo.

O desempenho do modelo no processo de treinamento utilizando o mencionado conjunto de dados neuromórfico, também se apoia em duas principais características do modelo, sendo

elas precisão e sensibilidade. A precisão consiste na razão entre as detecções corretas e as detecções totais do modelo, já a sensibilidade é a razão entre as detecções corretas em relação a quantidade de detecções que deveriam ser estimadas pelo modelo em uma cena.

A Figura 27 apresenta o valor da precisão ao longo de diferentes níveis de confiança do modelo em suas predições. A confiança consiste no valor de certeza que o modelo atribui a uma inferência. É possível observar o desempenho de cada classe do modelo frente a análise da precisão e, em azul, é mostrado o valor médio do mesmo. Em média, a partir de 0.2 (20%) de confiança, o modelo tende a apresentar baixa variação no nível de precisão em suas detecções.

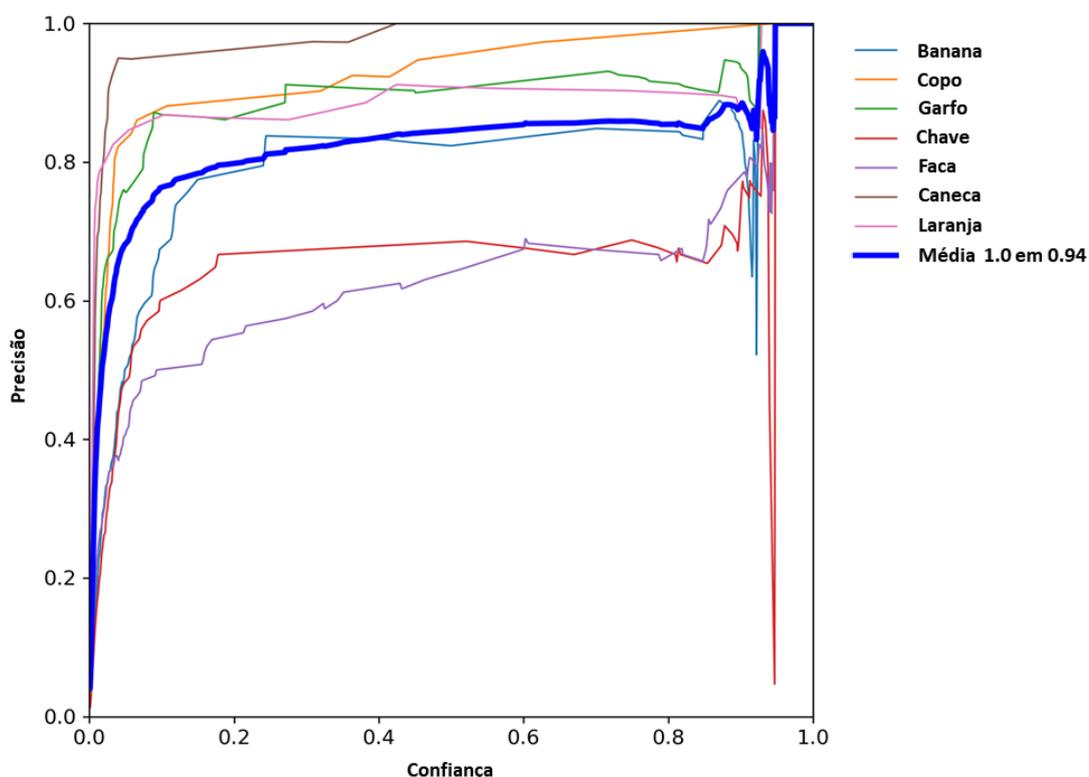


Figura 27: **Apresentação da Curva-P do modelo de detecção de objetos.** O gráfico consiste no valor da precisão do modelo para diferentes níveis de confiança em cada predição. A média de precisão do modelo atinge valor máximo quando o limiar de confiança das predições é determinado para classificar objetos apenas com confiança superior a 0.94 (94%).

Todavia, somente a precisão das detecções não é suficiente para ilustrar a performance do modelo, é necessário entender como o modelo performa em termos de sensibilidade. Para tanto, a Figura 28 apresenta a Curva-R responsável por realizar a mesma análise descrita para a Curva-

P, porém em termos de sensibilidade.

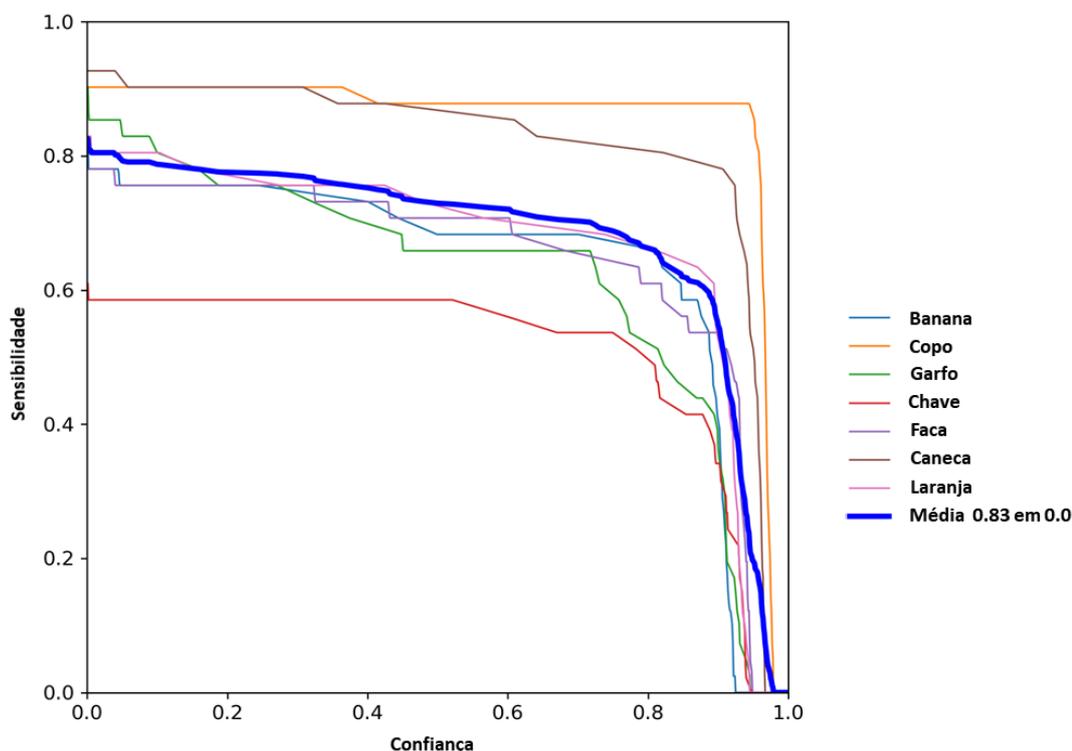


Figura 28: **Apresentação da Curva-R do modelo de detecção de objetos.** O gráfico consiste no valor da sensibilidade do modelo para diferentes níveis de confiança em cada predição. A média de sensibilidade do modelo atinge valor máximo quando o limiar de confiança das predições é determinado para classificar objetos apenas com confiança 0.0 (0%).

Da mesma maneira que é apresentado na Curva-P, o gráfico da Figura 28 apresenta a sensibilidade do modelo para cada classe presente no conjunto de dados e, em azul, apresenta a média geral da sensibilidade do modelo para todas as classes. Em situações em que o limiar de confiança é determinado como zero, o modelo é capaz de estimar detecções mesmo que em um alto nível de incerteza e, portanto, é mais inclinado a encontrar as classes na cena e determinar a classe correta da detecção, mesmo que essa capacidade não seja total (sensibilidade = 1) o valor é bem próximo do ideal, atingindo uma taxa de 0.83 (83%) de sensibilidade. Em ambos os gráficos, a performance da classe Chave se distingue, apresentando tanto precisão quanto sensibilidade abaixo das demais. Novamente, esse fator pode ser explicado pelo tamanho do objeto e a fácil confusão de um pequeno agrupamento de eventos com informações ruidosas.

Para alinhar o resultado da precisão com a sensibilidade e determinar a Precisão Média do modelo ( $mAP$  – *mean Average Precision*) é necessário analisar o gráfico da Figura 29. Nele é possível perceber que o desempenho médio do modelo quando em relação a ambas as características (precisão e sensibilidade) é de 0.73 (73%) para detecções com IoU (*Intersection over Union*) de 0.5 (50%), ou seja, as detecções determinadas válidas consistem em detecções cujo modelo foi capaz de inferir a caixa delimitadora da classe com 50% de área em comum com o rótulo.

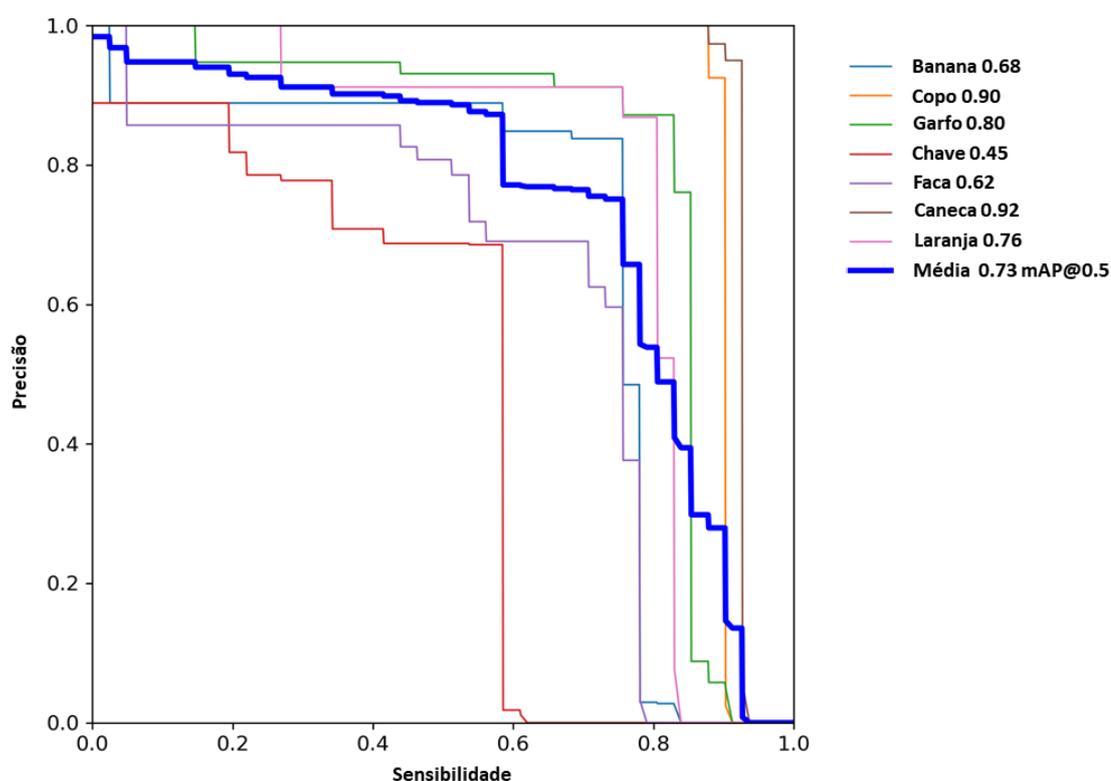


Figura 29: **Apresentação da Curva-PR do modelo de detecção de objetos.** O gráfico apresenta a relação entre a sensibilidade e a precisão do modelo para cada classe presente no conjunto de dados e, em azul, apresenta o valor médio para todas as classes.

Por fim, a métrica de acurácia geral do modelo pode ser analisada através do valor de *F1-score*. O gráfico da Figura 30 apresenta a curva do valor de F1 para diferentes limiares de confiança e apresenta uma acurácia geral (linha azul) de 0.79 (79%) para uma confiança de 0.27 (27%). O valor de F1 médio do modelo não apresenta grande variação entre 0.18 (18%) até 0.8 (80%) de limiar de confiança. Em outras palavras, dentro do intervalo de confiança citado, o

modelo performa de maneira balanceada entre sensibilidade e precisão.

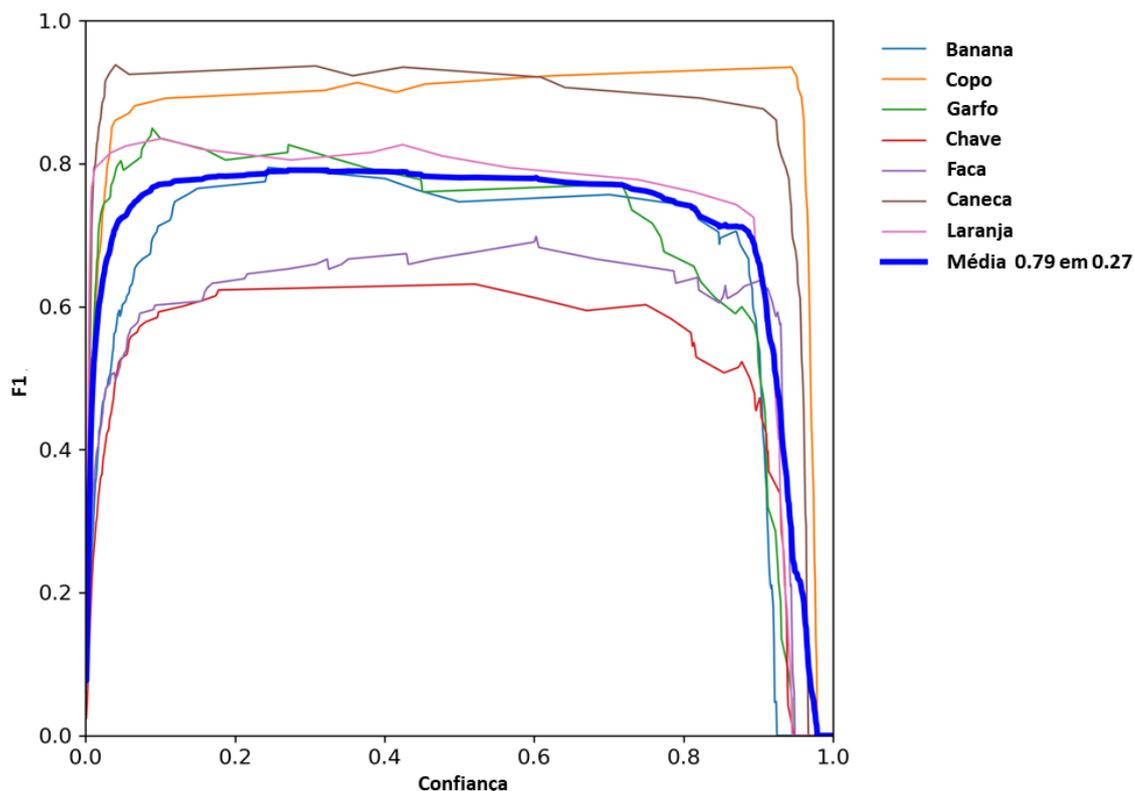


Figura 30: Apresentação da curva do *F1-Score* do modelo de detecção de objetos. O gráfico apresenta a relação entre a sensibilidade e a precisão nos parâmetros do limiar de confiança das predições do modelo para cada classe presente no conjunto de dados e, em azul, apresenta o valor médio para todas as classes.

### 4.3 Análise do desempenho do modelo de rastreamento em tempo real

A seguir, são apresentados e discutidos os resultados dos experimentos realizados para caracterizar o desempenho do sistema semiautônomo de alcance para próteses de membro superior. Os experimentos consistem em testar o sistema de rastreamento em tempo real em cinco situações distintas.

A primeira situação consiste em um ambiente com apenas um objeto em cena com eventos sendo gerados pela presença de um cenário sem plano de fundo (Experimento 1) e com a presença de plano de fundo (Experimento 3). A comparação entre o padrão de eventos originados em cada um dos dois experimentos citados pode ser observada na Figura 31.

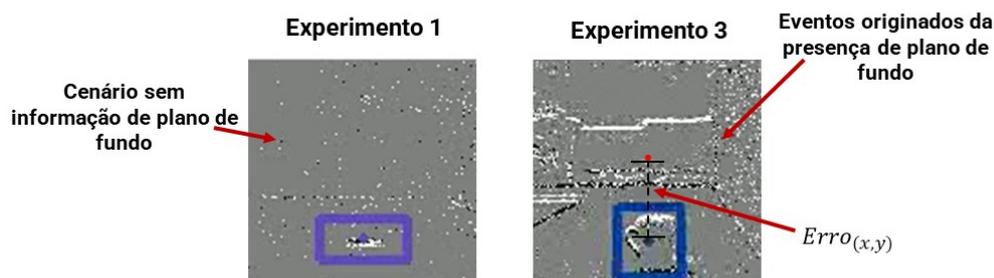


Figura 31: **Imagem comparando o padrão de eventos gerados no Experimento 1 e 3.** A chave é o objeto de interesse utilizado no Experimento 1 enquanto a caneca é o objeto de interesse no Experimento 3. O sinal de erro do modelo de rastreamento em tempo real é apresentado na imagem e se baseia na distância euclidiana entre o centroide do objeto detectado e do centro da cena gravada pela DVS128.

O segundo momento dos experimentos (Figura 32) realizados com o sistema de rastreamento em tempo real tem o intuito de analisar sua robustez frente a presença de informações de plano de fundo em situações com múltiplos objetos em cena, e consiste na inclusão de mais dois objetos que funcionam como distratores em um cenário com informações de plano de fundo (Experimento 4) e sem a presença dos eventos gerados pelo movimento relativo da câmera neuromórfica em relação ao cenário (Experimento 2).

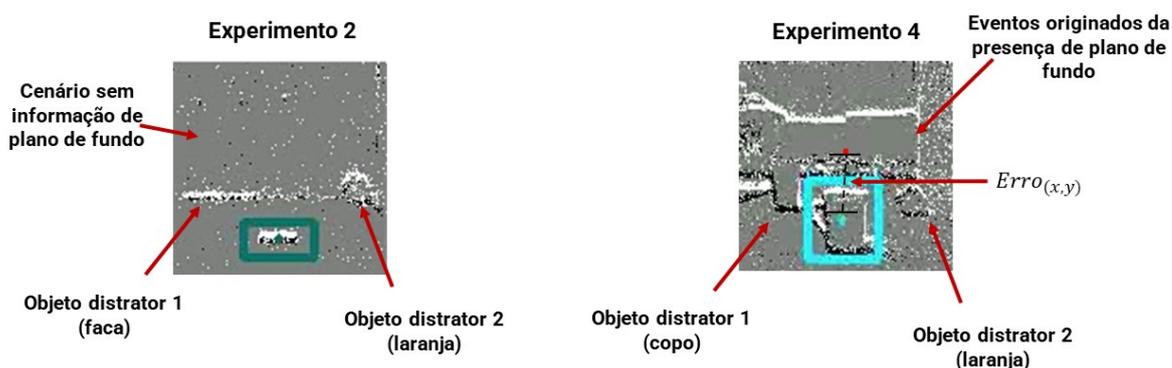


Figura 32: **Imagem comparando o padrão de eventos gerados no Experimento 2 e 4.** A chave é o objeto de interesse, enquanto os objetos distratores são a faca e a laranja no Experimento 2 enquanto a caneca é o objeto de interesse no Experimento 4 e os objetos distratores são o copo e a laranja.

O terceiro e quarto cenário de comparação dos resultados consistem em analisar a influência dos objetos distratores no desempenho do sistema de rastreamento em tempo real (com e

sem a presença do plano de fundo, respectivamente). Portanto é importante analisar o desempenho comparado entre os experimentos 1 e 2 (sem plano de fundo) e experimentos 3 e 4 (com plano de fundo).

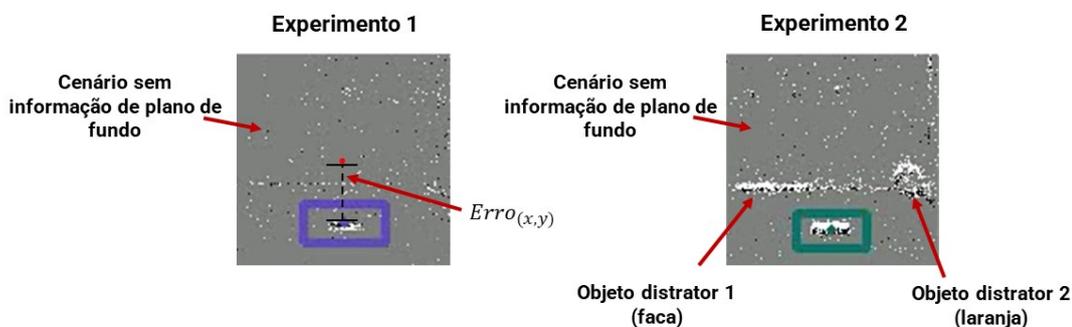


Figura 33: **Imagem comparando o padrão de eventos gerados no Experimento 1 e 2.** A chave é o objeto de interesse, enquanto os objetos distratores são a faca e a laranja no Experimento 2 enquanto, no Experimento 1, a chave é o único objeto em cena. Em ambos os casos não há informação proveniente do plano de fundo.

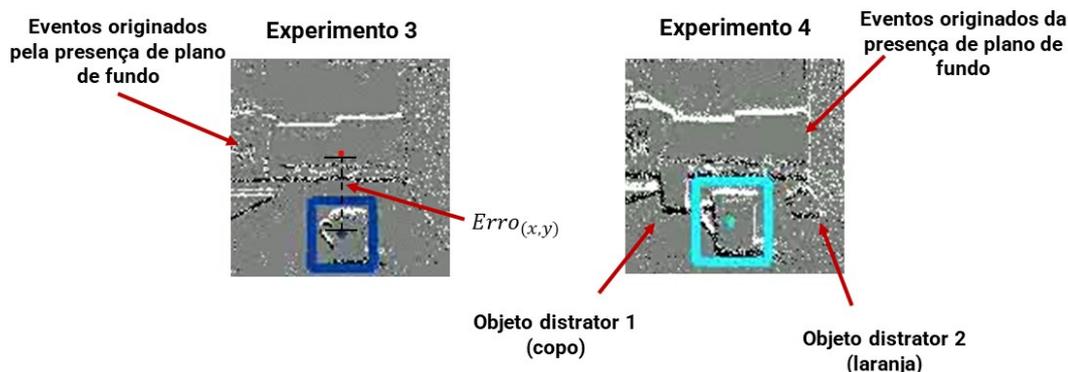


Figura 34: **Imagem comparando o padrão de eventos gerados no Experimento 3 e 4.** A caneca é o objeto de interesse e os objetos distratores são o copo e a laranja no Experimento 4 enquanto, no Experimento 3, a caneca é o único objeto em cena. Em ambos os casos há informação proveniente do plano de fundo.

O último estágio de avaliação do sistema de rastreamento em tempo real (Experimento 5), consiste em analisar a performance em condições de movimentação do objeto na cena durante a realização do alcance. Para simular essas condições, foi realizada uma movimentação do objeto de um ponto A para um ponto B dois segundos após o início do processo de alcance.

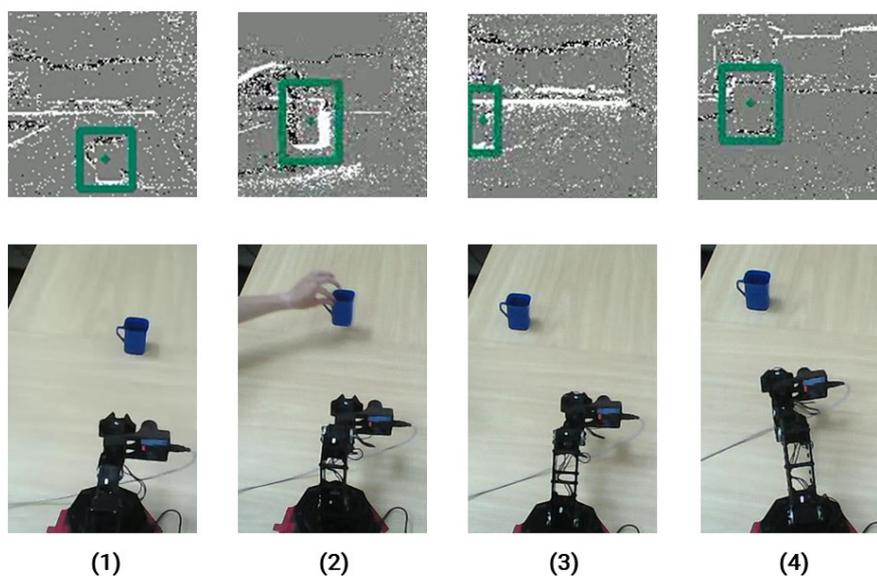


Figura 35: **Imagem do processo de movimentação do objeto durante o Experimento 5.** Os estágios da gravação (de 1 a 5) são sequenciais de acordo com o passar do tempo enquanto o WidowX realiza o movimento de alcance. Em (2) é apresentado o momento em que o objeto sofre a translação em cena. Em (3) o objeto encontra-se transladado para o canto esquerdo da cena, sendo parcialmente ocluído ao campo de visão da DVS128 e em (4) há a início da correção da trajetória do WidowX conforme informações do sistema de alcance para próteses de membro superior.

O desempenho geral do movimento de alcance do braço robótico em direção ao objeto de interesse promovido pelo sistema proposto de alcance para próteses de membro superior, utilizando visão neuromórfica, foi calculado utilizando a Equação 6 e é apresentado na Figura 36.

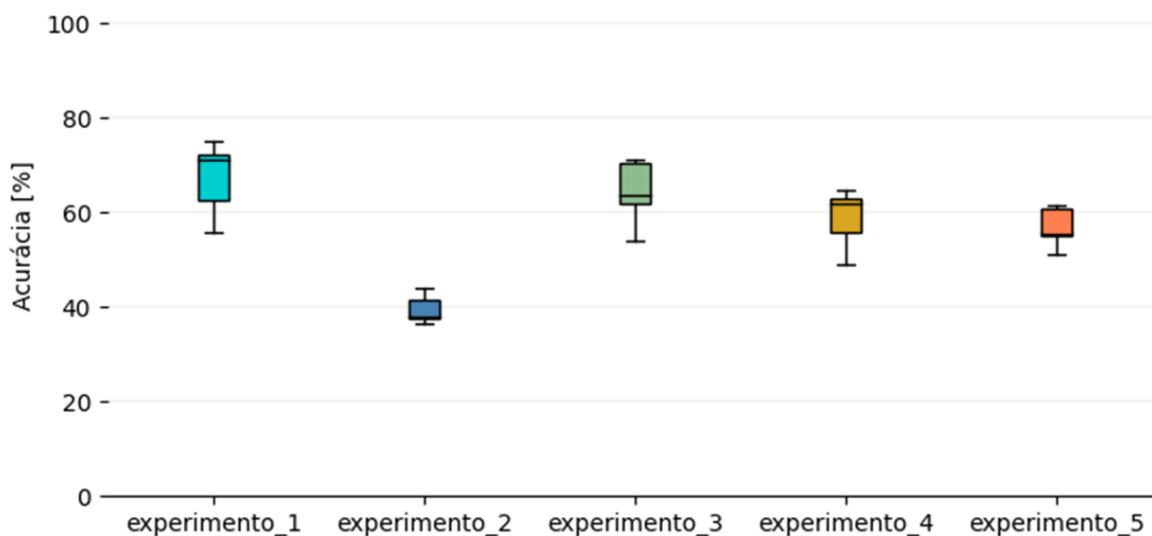


Figura 36: **Acurácia geral do sistema de rastreamento em tempo real, em cada experimento realizado.** Cálculo realizado conforme Equação 6 descrita na seção 3.4.6.

Quando realizada a comparação entre o desempenho do sistema frente a situações distintas, conforme descrito nos parágrafos anteriores, é possível perceber que não há grande variação de resultados entre os cenários em cada experimento, exceto pelo Experimento 2 que apresentou desempenho abaixo dos demais, entretanto o objeto de interesse no caso do Experimento 2 foi a Chave que, por sua vez, apresenta a pior performance do modelo de detecção, tanto em termos de precisão quanto de sensibilidade (Figura 27, Figura 28), comprometendo a performance geral do sistema de rastreamento em tempo real. Contudo, o desempenho de alcance no Experimento 1 (que também utiliza a chave como objeto de interesse), não parece ser prejudicado por esse fato, entretanto a falta de plano de fundo e objetos distratores pode ter contribuído para o bom desempenho nesse experimento, em relação ao experimento 2.

Todavia, a presença de plano de fundo e elementos distratores não parece afetar o desempenho geral do sistema de alcance quando comparamos os experimentos realizados (Experimento 3 e 4) com uma das classes que melhor performou dentro do processo de treinamento do modelo (caneca). Porém, ao analisar a influência do plano de fundo e dos objetos distratores separadamente comparando os experimentos 1 e 3, e os experimentos 2 e 4 é possível notar que entre os experimentos 1 e 3 há baixa diferença entre os valores de performance, já entre os experimentos 2 e 4 há notável diferença, porém é presumível que essa variação seja reflexo da escolha dos objetos para realização dos experimentos.

Uma das hipóteses levantadas para explicar a diferença na performance do alcance entre os experimentos sem plano de fundo (Experimento 1 e 2) e os experimentos com plano de fundo (Experimento 3 e 4), seria o fato de que o conjunto de dados desenvolvido nesse trabalho continha apenas gravações com plano de fundo levando a um sobre ajuste do modelo. Entretanto, quando se contrasta essa hipótese com o resultado obtido no Experimento 5 que, por sua vez, possui o objeto sendo transladado de posição e as gravações incluem a mão do operador na hora de realizar a translação, percebe-se que o modelo de detecção, mesmo em condições diferentes das previstas no conjunto de dados, consegue distinguir as características do objeto de interesse. Portanto, é possível crer que o modelo de detecção conseguiu discriminar características do objeto em detrimento do tipo de plano de fundo.

Essa análise também permite concluir que o modelo de detecção desenvolvido para esse trabalho (aliado com um conjunto de dados verdadeiramente neuromórfico e razoavelmente diversificado), é capaz de performar bem, mesmo em condições que seriam problemáticas quando pensamos no desafio de associação de data [12], [23] ao qual os modelos de classificação de informações gravadas por câmeras baseadas em eventos estão susceptíveis.

Para estimar a acurácia do sistema de rastreo em tempo real ao longo do movimento de alcance, foi realizada uma análise do valor de acurácia do modelo a cada cinco interações do algoritmo (conforme Equação 7). Dessa maneira, o gráfico da Figura 37 apresenta o valor de acurácia ao longo do tempo para cada um dos cinco experimentos conduzidos durante a realização da validação da proposta do sistema de alcance para próteses de membro superior.

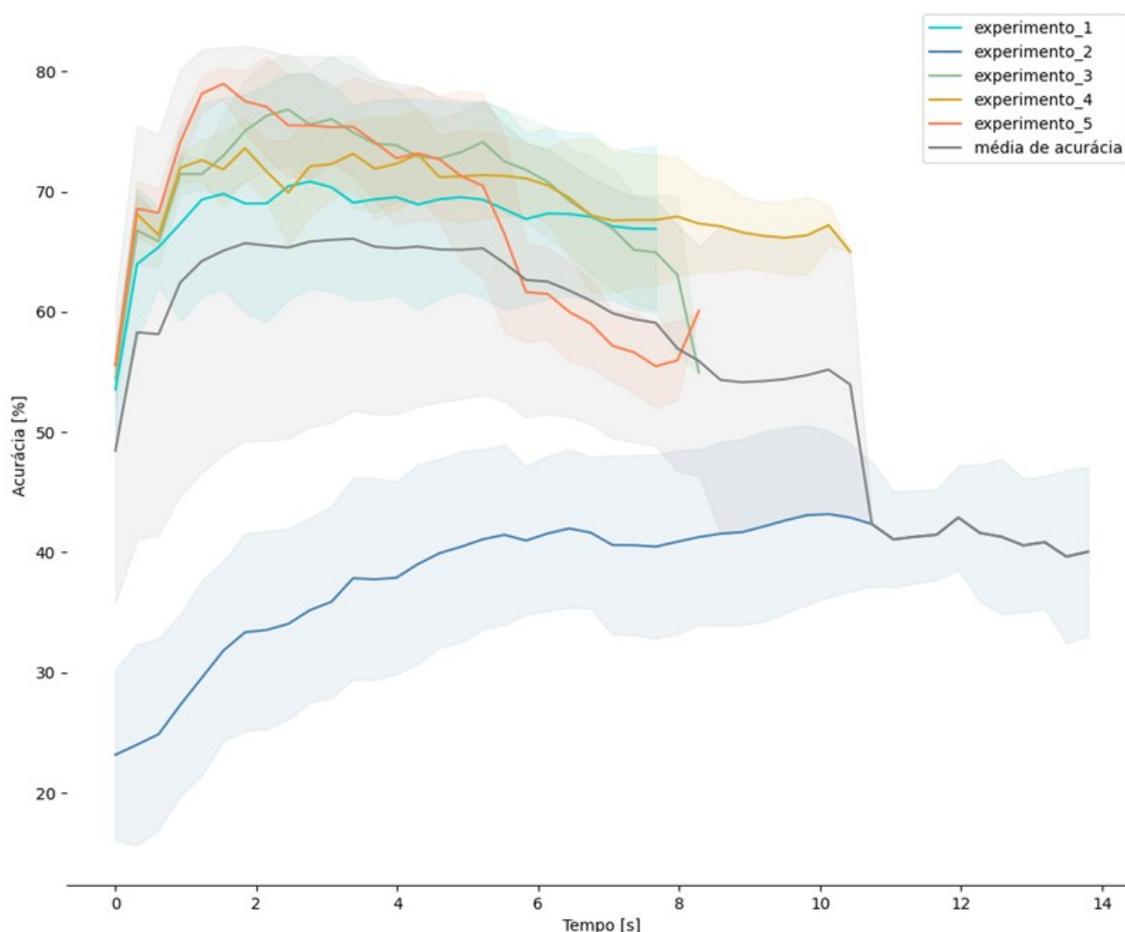


Figura 37: **Gráfico de acurácia do sistema de rastreo ao longo do tempo, discriminado por experimento.** A imagem apresenta a acurácia ao longo do movimento de alcance realizado pelo WidowX de cada experimento realizado e, em cinza, apresenta a acurácia média geral para todos os experimentos.

Ao analisar atentamente as curvas apresentadas na Figura 37, é possível perceber que, ao realizar o cálculo da acurácia, a cada janela de cinco detecções, temos uma acurácia baixa nos primeiros instantes de movimento do braço robótico, uma acurácia que se eleva e, posteriormente, volta a diminuir. Essa variação no valor de acurácia ao longo do tempo é reflexo da própria característica da DVS128 que, por ser baseada em eventos, precisa de movimento relativo entre o sensor e o objeto para que haja eventos para serem discriminados e o movimento de alcance depende da discriminação de objetos na cena. Portanto, no início do processo de alcance, o modelo promove movimentos sacádicos com o intuito de gerar eventos suficientes para discriminação de objetos na cena e, uma vez em movimento, o próprio movimento de alcance permite que os eventos sejam gerados, elevando o valor da acurácia.

Entretanto, ao se aproximar do objeto de interesse, as proporções do objeto crescem em relação ao tamanho da imagem, dificultando a classificação do objeto pelo modelo de detecção. Esse processo ocasiona uma maior incerteza do modelo acerca da posição do objeto, levando ao aumento de interações classificadas como DI e MS e, portanto, diminuindo a acurácia geral do modelo.

Vale ressaltar que os objetos de interesse utilizados nos experimentos foram posicionados a uma distância de 60 cm do centro da base. Essa distância foi escolhida pois o sistema utiliza apenas as informações da câmera para se guiar pelo ambiente, e a Abertura Numérica (NA) da lente utilizada na câmera é o que determina o campo de visão do sistema (Figura 38). Dessa maneira, os objetos eram apenas visualizados por completo a uma distância mínima de 15 cm.

A utilização de lentes de campo amplo de visão (ou objetiva grande-angular) com uma Abertura Numérica maior poderia proporcionar para o sistema de visão acoplado a prótese uma percepção do objeto como um todo, mesmo que muito próximo do objeto de interesse e, portanto, não seria necessária uma alteração no sistema proposto, mas sim na lente do dispositivo de visão utilizado.

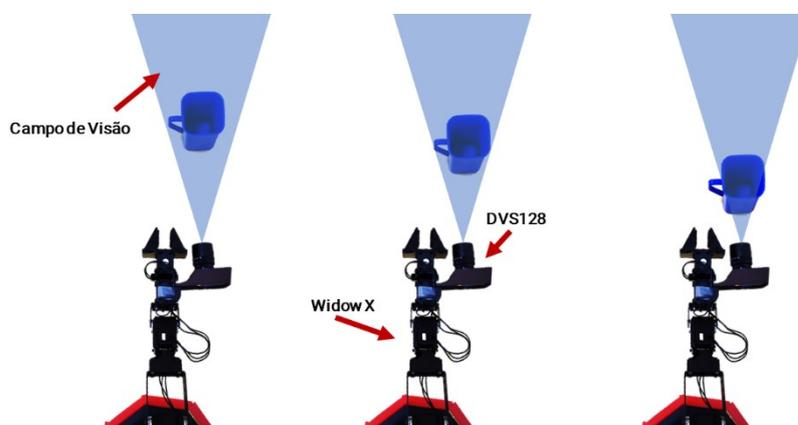


Figura 38: **Relação entre a proximidade do objeto de interesse em relação ao campo de visão da DVS128.** O campo de visão de câmeras é determinado pelas características das lentes utilizadas e, portanto, o modelo pode operar em diferentes distâncias conforme as lentes utilizadas no sistema de visão.

O sistema de rastreamento em tempo real gera como resultado, a cada interação, três diferentes cenários, sendo eles: Detecção Válida (DV), Detecção Inválida (DI) e Movimento Sacádico (MS). Dessa maneira, a Figura 39 apresenta a porcentagem média de cada um dos tipos de resultado gerado pelo sistema, para cada experimento realizado.

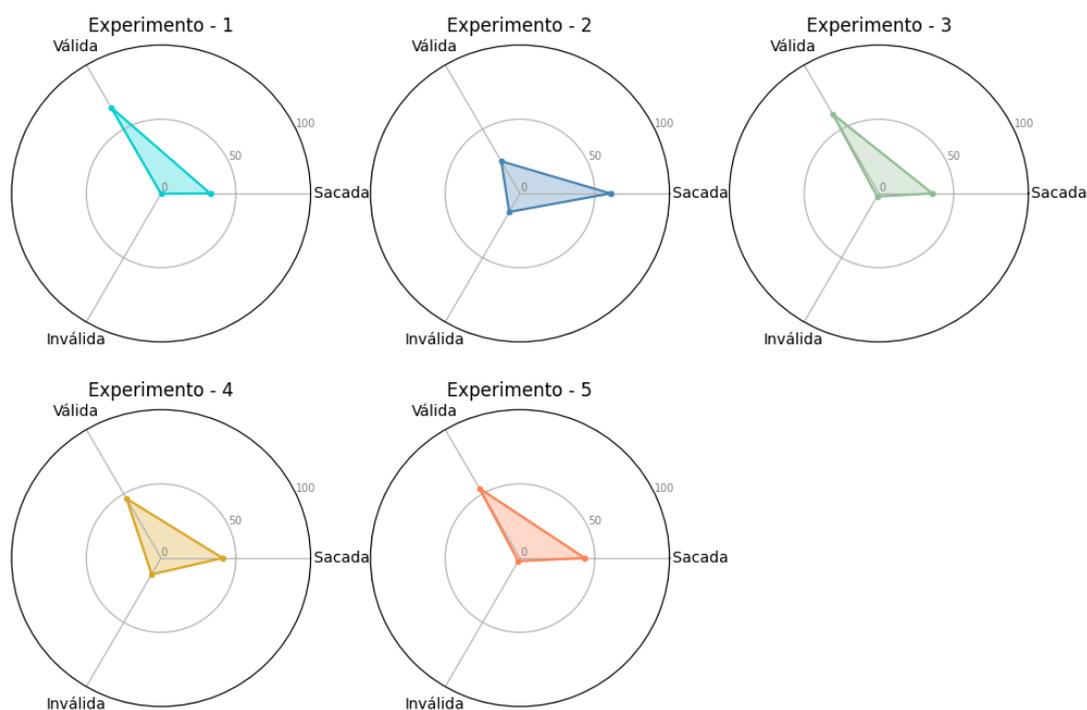


Figura 39: **Apresentação da média de distribuição de resultados do sistema de rastreo em tempo real durante o movimento de alcance para cada experimento realizado.** Cada gráfico representa um experimento e contém a distribuição média percentual do resultado do sistema, em cada experimento.

Quando comparamos os dois experimentos realizados com a presença de objetos distratores (Experimento 2 e 4) nos gráficos da Figura 39, percebe-se que a presença de saídas caracterizados como DI (detecção referente a um objeto que não é o de interesse), é relativamente maior que nos demais experimentos, outra característica desses experimentos é a presença de uma alta taxa de MS. Um dos motivos para que isso ocorra, no Experimento 4, pode ser o fato de que, em determinados momentos do alcance, os objetos distratores se obstruíam (na perspectiva da DVS128) e gerava um padrão de eventos indistinguíveis para o modelo de detecção o que ocasionava uma falta de capacidade do modelo de separar os eventos de cada objeto e, portanto, o modelo de detecção não era capaz de realizar uma inferência na cena. Entretanto, no Experimento 2 o objeto de interesse é a chave e, por ser um objeto de proporções pequenas, esse fenômeno não ocorre e, portanto, a explicação provável é o baixo desempenho do modelo de detecção para essa classe de objeto.

Os Experimentos 1 e 3 apresentam um resultado mais assertivo, gerando uma maior

quantidade de DV, menos MS (em comparação com os experimentos 2 e 4) e quase nenhuma DI, devido à falta de objetos distratores. A falta de DI no Experimento 3 (com plano de fundo) é um bom indicativo de que, mesmo com diversos eventos oriundos do plano de fundo, o modelo foi capaz de distinguir o objeto de interesse das informações do plano de fundo.

O Experimento 5 apresenta uma porcentagem de MS ligeiramente superior que o Experimento 3 (mesmo objeto e condições similares), resultado que demonstra que o sistema de rastreamento em tempo real esteve incerto acerca da posição do objeto de interesse por mais tempo por conta do movimento de translação do objeto em cena. A Figura 40 apresenta um compilado acerca do tempo médio gasto cada vez que o sistema esteve incerto acerca da posição do objeto e precisou realizar movimento sacádico. Em todos os experimentos (exceto Experimento 2) o modelo não ultrapassa 550ms realizando movimento sacádico antes de ser capaz de realizar uma inferência acerca da posição do objeto em cena.

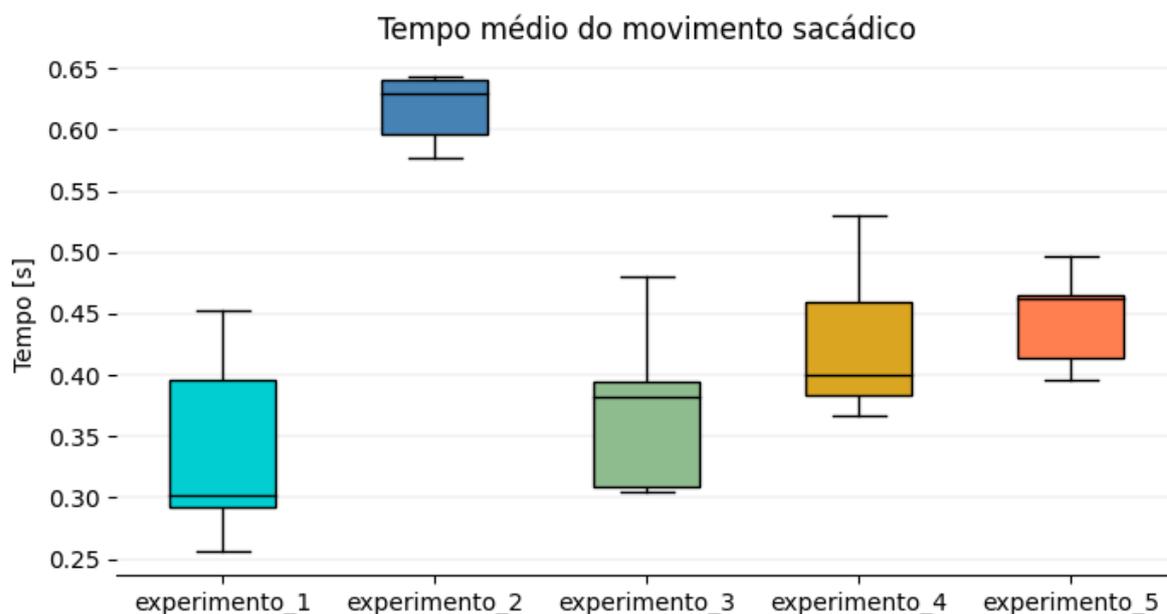


Figura 40: **Tempo médio gasto realizando movimento sacádico até gerar eventos suficientes para que o modelo de detecção de objetos realizasse uma inferência.** O gráfico discrimina os valores de tempo médio do movimento sacádico para cada experimento realizado.

Por fim, a Figura 41 apresenta o tempo médio gasto pelo sistema de rastreamento em cada experimento para realizar todo o movimento de alcance. Além de possuir alta taxa de MS e DI, o Experimento 2 também apresenta um elevado tempo total de realização do alcance do objeto

de interesse, entretanto os demais experimentos não apresentam tempo superior a dez segundos para realizar o alcance do objeto utilizando o sistema semiautônomo de rastreamento em tempo real baseado em visão neuromórfica, proposto para próteses de membro superior.

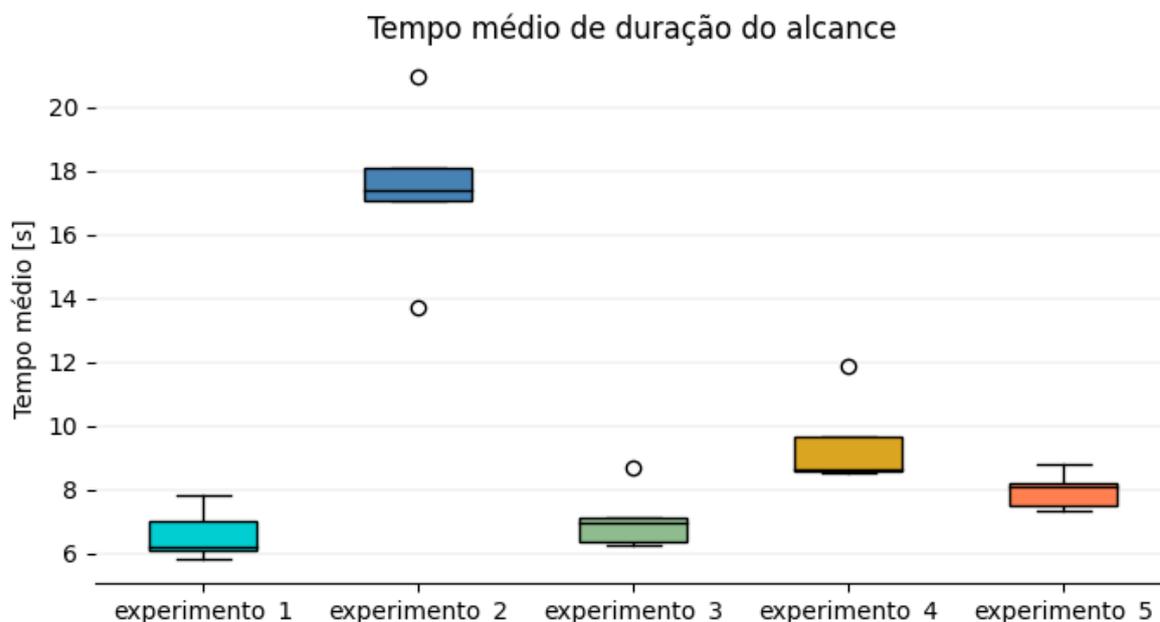


Figura 41: **Tempo médio necessário para o modelo realizar o movimento de alcance do objeto de interesse.** O gráfico apresenta o valor do tempo médio, em segundos, para cada experimento realizado.

#### 4.4 Considerações finais

Esse trabalho se encaixa na linha de desenvolvimento de dispositivos semiautônomos que proporcionam maior usabilidade e diminuição de carga cognitiva para usuários de próteses de membro superior [9]–[11], [50]. Almejando o desenvolvimento de um sistema embarcado de alcance para próteses de membro superior, nós propomos um modelo e realizamos a validação da proposta do sistema de alcance utilizando uma câmera baseada em eventos (DVS128) e um braço robótico comercial (WidowX). Esse trabalho visa pavimentar o caminho para um sistema semiautônomo para usuários de prótese de membro superior com amputações acima do nível do cotovelo. Na literatura é possível encontrar mecanismos robóticos para realização de tarefas de manipulação de objetos [99], entretanto, a necessidade de sistemas protéticos de possuir

baixo consumo de energia, alta eficiência, baixo custo computacional e portabilidade é um impeditivo para a implementação da maioria das técnicas de processamento de imagens em tempo real utilizando câmeras convencionais [63], [87], [100].

Dessa maneira, a utilização de sistemas neuromórficos abre a oportunidade do desenvolvimento de sistemas altamente eficientes, computacional e energeticamente [34], [55], [101]. Porém, por se tratar de um novo paradigma computacional (computação assíncrona e baseada em eventos), novos algoritmos e abordagens computacionais para tratar os dados precisam ser desenvolvidos e desafios precisam ser vencidos [12].

Todavia, o sistema de identificação de objetos proposto nesse trabalho possui potencial para realizar aferições acerca de padrões em superfícies de eventos superando o problema de associação de data [12], [23] ao adaptar o modelo de detecção de objetos estado arte em Visão Computacional [87] para trabalhar com informações de câmeras baseadas em eventos e utilizar o modelo de detecção para propor um sistema de alcance de objetos que é capaz de realizar tarefas de alcance em aproximadamente 10 segundos (conservada a noção de espaço e tamanho do braço robótico utilizados na validação).

Alguns trabalhos presentes na literatura utilizam de câmeras neuromórficas para promover um sistema de automação de próteses de membro superior realizando aferições acerca dos objetos em cena e determinar estratégias de pegada para próteses de mão [10], [102], [103]. Entretanto, o nosso trabalho se difere dos encontrados na literatura ao propor um sistema de alcance baseado em visão neuromórfica, para próteses acima do nível do cotovelo, que se baseia em aferições da posição de objetos gravados por câmeras baseada em eventos para guiar um braço robótico pela cena. Nos trabalhos encontrados na literatura, a prótese é guiada pelo usuário e as informações adquiridas pela câmera são utilizadas para aferir características acerca da rotação e da forma do objeto para estimar a estratégia de pegada ideal para manipulação do objeto.

Por fim, o sistema desenvolvido nesse trabalho realiza aferição de objetos em aproximadamente 7ms utilizando superfícies de eventos com  $\Delta t$  de 30ms gravado por uma câmera baseada em eventos (DVS128) que possui potência de 23mW [13], sugerindo a potencialidade computacional e energética para a realização da implementação do sistema de alcance para próteses de membro superior em sistemas embarcados.

## Capítulo 5

---

### Conclusões

O presente trabalho apresentou a proposta, e validação, de um sistema semiautônomo de alcance baseado em visão neuromórfica. Os resultados sugerem que o sistema proposto pode ser adequado para próteses de membro superior conforme a primeira hipótese e objetivo geral do trabalho. O sistema proposto nesse trabalho apresenta indícios de que tem potencial para permitir que próteses de membro superior tenham maior autonomia na realização de tarefas cotidianas, ao utilizar visão neuromórfica, que possui alta eficiência energética e computacional, corroborando a segunda hipótese e primeiro objetivo específico do trabalho.

O modelo proposto para detecção de objetos gravados por uma câmera baseada em eventos, apresentou capacidade para distinguir informações de objetos distintos além de informações de plano de fundo proveniente do movimento da câmera pelo ambiente, demonstrando que modelos de aprendizagem profunda para detecção e rastreamento de características, aliados a um dataset bem selecionado, pode apresentar bons resultados, completando o segundo objetivo específico.

Ademais, essa dissertação, apresenta uma adequação de um modelo de Visão Computacional, estado da arte em detecção de objetos, para ser utilizado em Visão Neuromórfica que, por sua vez, abre caminhos para implementação da técnica apresentada em diversas aplicações de robótica tradicional que utilizam visão.

O sistema de alcance apresenta bom desempenho ao realizar o rastreamento de características utilizando câmeras baseadas em eventos, além de lidar com falhas de rastreamento ao realizar movimento sacádico para geração de informações neuromórficas quando o modelo não possui informações suficientes para interpretar a cena.

Esse trabalho apresenta para a comunidade de Visão Neuromórfica um conjunto de dados, gravados utilizando uma DVS128, de sete objetos distintos em diferentes perspectivas com informações acerca de sua localização na cena, rotulados manualmente e disponibilizado online

para a realização de quaisquer trabalhos com modelos de aprendizado supervisionado, cumprindo o terceiro objetivo específico dessa dissertação.

Por fim, como trabalho futuro, o conjunto de dados proposto poderia conter uma maior quantidade de objetos, além de conter mais representações de um mesmo objeto, de maneira a possuir maior variedade de posições e perspectivas. Além de que, o sistema proposto nesse trabalho, poderá ser implementado de maneira embarcada em dispositivos *TinyML* para aplicação da metodologia desenvolvida em uma prótese de membro superior comercial, e posterior teste de usabilidade com indivíduos amputados.

## Apêndice

### 6.1 Lista de artigos publicados

- E.B. Gouveia, E.L.S. Gouveia, V.T. Costa, A. Nakagawa-Silva, A.B. Soares, **“Classification of Objects Using Neuromorphic Camera and Convolutional Neural Networks”**, *Congresso Brasileiro de Engenharia Biomédica*, 2020.
- E.B. Gouveia, L.V. Costa, E.L.S. Gouveia, V.T. Costa, A. Nakagawa-Silva, A.B. Soares, **“An Object Tracking Using a Neuromorphic System Based on Standard RGB Cameras”**, *Congresso Brasileiro de Engenharia Biomédica*, 2020.
- E.L.S. Gouveia, E.B. Gouveia, A. Nakagawa-Silva, A.B. Soares, **“Neuromorphic Vision-aided Semi-autonomous System for Prosthesis Control”**, *Congresso Brasileiro de Engenharia Biomédica*, 2020.
- E.B. Gouveia, G.F. Tavares, L.L. Almada, A. Nakagawa-Silva, E.L.S. Gouveia, M.J. Cunha, E.A.L. Junior, A.B. Soares, **“Object detection using sparse data representation with convolutional neural networks for event-based cameras”**, *Simpósio de Engenharia Biomédica*, 2021

## Referências

- [1] H. Y. N. Lindner, B. S. Nätterlund, and L. M. N. Hermansson, “Upper limb prosthetic outcome measures: Review and content comparison based on International Classification of Functioning, Disability and Health,” *Prosthetics and Orthotics International*. 2010, doi: 10.3109/03093641003776976.
- [2] R. MacDougall, “The Significance of the Human Hand in the Evolution of Mind,” *Am. J. Psychol.*, 1905, doi: 10.2307/1412130.
- [3] K. Demet, N. Martinet, F. Guillemin, J. Paysant, and J. M. André, “Health related quality of life and related factors in 539 persons with amputation of upper and lower limb,” *Disabil. Rehabil.*, 2003, doi: 10.1080/0963828031000090434.
- [4] S. I. Woodruff, M. R. Galarneau, D. I. Sack, C. T. McCabe, and J. L. Dye, “Combat amputees’ health-related quality of life and psychological outcomes: A brief report from the wounded warrior recovery project,” *J. Trauma Acute Care Surg.*, 2017, doi: 10.1097/TA.0000000000001348.
- [5] A. D. Roche, B. Lakey, I. Mendez, I. Vujaklija, D. Farina, and O. C. Aszmann, “Clinical Perspectives in Upper Limb Protheses: An Update,” *Current Surgery Reports*. 2019, doi: 10.1007/s40137-019-0227-z.
- [6] DATASUS, “Number of upper-limb amputations registred on SUS,” <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sih/cnv/qiuf.def>, 2020.  
<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sih/cnv/qiuf.def>.
- [7] K. Ziegler-Graham, E. J. MacKenzie, P. L. Ephraim, T. G. Travison, and R. Brookmeyer, “Estimating the Prevalence of Limb Loss in the United States: 2005 to 2050,” *Arch. Phys. Med. Rehabil.*, 2008, doi: 10.1016/j.apmr.2007.11.005.
- [8] F. Cordella *et al.*, “Literature review on needs of upper limb prosthesis users,” *Frontiers in Neuroscience*. 2016, doi: 10.3389/fnins.2016.00209.
- [9] L. E. Osborn *et al.*, “Prosthesis with neuromorphic multilayered e-dermis perceives touch and pain,” *Sci. Robot.*, 2018, doi: 10.1126/scirobotics.aat3818.
- [10] M. Hays, L. Osborn, R. Ghosh, M. Iskarous, C. Hunt, and N. V. Thakor, “Neuromorphic vision and tactile fusion for upper limb prosthesis control,” 2019, doi: 10.1109/NER.2019.8716890.

- [11] L. Resnik, H. H. Huang, A. Winslow, D. L. Crouch, F. Zhang, and N. Wolk, "Evaluation of EMG pattern recognition for upper limb prosthesis control: A case study in comparison with direct myoelectric control," *J. Neuroeng. Rehabil.*, 2018, doi: 10.1186/s12984-018-0361-3.
- [12] G. Gallego *et al.*, "Event-based vision: A survey," *arXiv Prepr. arXiv1904.08405*, pp. 1–25, 2019, doi: 10.1109/tpami.2020.3008413.
- [13] P. Lichtsteiner, C. Posch, and T. Delbruck, "A  $128 \times 128$  120 dB 15  $\mu$ s latency asynchronous temporal contrast vision sensor," *IEEE J. Solid-State Circuits*, 2008, doi: 10.1109/JSSC.2007.914337.
- [14] R. Berner, C. P. Brändli, M. Yang, S.-C. Liu, and T. Delbrück, "A  $240 \times 180$  120 dB 10 mW 12 microsecond - latency sparse output vision sensor for mobile applications," *Int. Image Sens. Work.*, 2013.
- [15] G. Indiveri and T. K. Horiuchi, "Frontiers in neuromorphic engineering," *Frontiers in Neuroscience*. 2011, doi: 10.3389/fnins.2011.00118.
- [16] C. D. Schuman *et al.*, "A Survey of Neuromorphic Computing and Neural Networks in Hardware," *arXiv*. 2017.
- [17] M. Davies *et al.*, "Loihi: A Neuromorphic Manycore Processor with On-Chip Learning," *IEEE Micro*, 2018, doi: 10.1109/MM.2018.112130359.
- [18] M. A. Mahowald and C. Mead, "The silicon retina," *Scientific American*. 1991, doi: 10.1038/scientificamerican0591-76.
- [19] T. Delbrück, B. Linares-Barranco, E. Culurciello, and C. Posch, "Activity-driven, event-based vision sensors," 2010, doi: 10.1109/ISCAS.2010.5537149.
- [20] R. W. Baldwin, R. Liu, M. Almatrafi, V. Asari, and K. Hirakawa, "Time-Ordered Recent Event (TORE) Volumes for Event Cameras," vol. 14, no. 8, pp. 1–14, 2021, [Online]. Available: <http://arxiv.org/abs/2103.06108>.
- [21] J. Manderscheid, A. Sironi, N. Bourdis, D. Migliore, and V. Lepetit, "Speed invariant time surface for learning to detect corner points with event-based cameras," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 10237–10246, 2019, doi: 10.1109/CVPR.2019.01049.
- [22] A. Sironi, M. Brambilla, N. Bourdis, X. Lagorce, and R. Benosman, "HATS: Histograms of Averaged Time Surfaces for Robust Event-Based Object Classification," 2018, doi:

- 10.1109/CVPR.2018.00186.
- [23] D. Gehrig, H. Rebecq, G. Gallego, and D. Scaramuzza, “Asynchronous, photometric feature tracking using events and frames,” *arXiv*. 2018.
- [24] S. Ham *et al.*, “One-dimensional organic artificial multi-synapses enabling electronic textile neural network for wearable neuromorphic applications,” *Sci. Adv.*, 2020, doi: 10.1126/sciadv.aba1178.
- [25] H. E. Lee *et al.*, “Novel Electronics for Flexible and Neuromorphic Computing,” *Advanced Functional Materials*. 2018, doi: 10.1002/adfm.201801690.
- [26] E. Chicca and G. Indiveri, “A recipe for creating ideal hybrid memristive-CMOS neuromorphic processing systems,” *Applied Physics Letters*. 2020, doi: 10.1063/1.5142089.
- [27] L. Q. Guo *et al.*, “Oxide Neuromorphic Transistors Gated by Polyvinyl Alcohol Solid Electrolytes with Ultralow Power Consumption,” *ACS Appl. Mater. Interfaces*, 2019, doi: 10.1021/acsami.9b05717.
- [28] I. Sugiarto, G. Liu, S. Davidson, L. A. Plana, and S. B. Furber, “High performance computing on SpiNNaker neuromorphic platform: A case study for energy efficient image processing,” 2017, doi: 10.1109/PCCC.2016.7820645.
- [29] V. Ntinis, P. Karakolis, G. C. Sirakoulis, and P. Dimitrakis, “Neuromorphic circuits on segmented crossbar architectures with enhanced properties,” 2020, doi: 10.1109/ECCTD49232.2020.9218289.
- [30] H. L. Park, Y. Lee, N. Kim, D. G. Seo, G. T. Go, and T. W. Lee, “Flexible Neuromorphic Electronics for Computing, Soft Robotics, and Neuroprosthetics,” *Adv. Mater.*, 2020, doi: 10.1002/adma.201903558.
- [31] A. Linares-Barranco, F. Perez-Peña, A. Jimenez-Fernandez, and E. Chicca, “ED-BioRob: A Neuromorphic Robotic Arm With FPGA-Based Infrastructure for Bio-Inspired Spiking Motor Controllers,” *Front. Neurobot.*, 2020, doi: 10.3389/fnbot.2020.590163.
- [32] A. Spaeth, M. Tebyani, D. Haussler, and M. Teodorescu, “Neuromorphic Closed-Loop Control of a Flexible Modular Robot by a Simulated Spiking Central Pattern Generator,” 2020, doi: 10.1109/RoboSoft48309.2020.9116007.
- [33] C. Michaelis, A. B. Lehr, and C. Tetzlaff, “Robust Trajectory Generation for Robotic Control on the Neuromorphic Research Chip Loihi,” *Front. Neurobot.*, 2020, doi:

- 10.3389/fnbot.2020.589532.
- [34] M. Hopkins, G. Pineda-Garcia, P. A. Bogdan, and S. B. Furber, “Spiking neural networks for computer vision,” *Interface Focus*, 2018, doi: 10.1098/rsfs.2018.0007.
- [35] N. K. Upadhyay, H. Jiang, Z. Wang, S. Asapu, Q. Xia, and J. Joshua Yang, “Emerging Memory Devices for Neuromorphic Computing,” *Advanced Materials Technologies*. 2019, doi: 10.1002/admt.201800589.
- [36] W. He *et al.*, “Comparing SNNs and RNNs on neuromorphic vision datasets: Similarities and differences,” *Neural Networks*, 2020, doi: 10.1016/j.neunet.2020.08.001.
- [37] D. Datta, K. Selvarajah, and N. Davey, “Functional outcome of patients with proximal upper limb deficiency - Acquired and congenital,” *Clin. Rehabil.*, 2004, doi: 10.1191/0269215504cr716oa.
- [38] K. Ostlie, I. M. Lesjø, R. J. Franklin, B. Garfelt, O. H. Skjeldal, and P. Magnus, “Prosthesis use in adult acquired major upper-limb amputees: Patterns of wear, prosthetic skills and the actual use of prostheses in activities of daily life,” *Disabil. Rehabil. Assist. Technol.*, 2012, doi: 10.3109/17483107.2011.653296.
- [39] P. Kutilek, J. Mares, J. Hybl, V. Socha, J. Schlenker, and A. Stefek, “Myoelectric arm using artificial neural networks to reduce cognitive load of the user,” *Neural Comput. Appl.*, 2017, doi: 10.1007/s00521-015-2074-x.
- [40] W. Zhang *et al.*, “Cognitive workload in conventional direct control vs. pattern recognition control of an upper-limb prosthesis,” 2017, doi: 10.1109/SMC.2016.7844587.
- [41] J. T. Belter, J. L. Segil, A. M. Dollar, and R. F. Weir, “Mechanical design and performance specifications of anthropomorphic prosthetic hands: A review,” *J. Rehabil. Res. Dev.*, 2013, doi: 10.1682/JRRD.2011.10.0188.
- [42] M. Atzori and H. Müller, “Control capabilities of myoelectric robotic prostheses by hand amputees: A scientific research and market overview,” *Front. Syst. Neurosci.*, 2015, doi: 10.3389/fnsys.2015.00162.
- [43] R. N. Scott and P. A. Parker, “Myoelectric prostheses: State of the art,” *J. Med. Eng. Technol.*, 1988, doi: 10.3109/03091908809030173.
- [44] A. Fougner, O. Stavdahl, P. J. Kyberd, Y. G. Losier, and P. A. Parker, “Control of upper limb prostheses: Terminology and proportional myoelectric control review,” *IEEE Trans. Neural Syst.*

- Rehabil. Eng.*, 2012, doi: 10.1109/TNSRE.2012.2196711.
- [45] J. NING and F. Dario, “Myoelectric control of upper limb prosthesis: current status, challenges and recent advances,” *Front. Neuroeng.*, 2014, doi: 10.3389/conf.fneng.2014.11.00004.
- [46] C. Connolly, “Prosthetic hands from Touch Bionics.” pp. 290–293, 2008.
- [47] C. Medynski and B. Rattray, “Bebionic Prosthetic Design,” *MEC 2011 Symp. MyoElectric Control. Prosthetics Symp.*, 2011.
- [48] Ottobock, “Michelangelo prosthetic hand,” *Michelangelo prosthetic hand*, 2015. .
- [49] G. S. Dhillon and K. W. Horch, “Direct neural sensory feedback and control of a prosthetic arm,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, 2005, doi: 10.1109/TNSRE.2005.856072.
- [50] G. Ghazaei, A. Alameer, P. Degenaar, G. Morgan, and K. Nazarpour, “Deep learning-based artificial vision for grasp classification in myoelectric hands,” *J. Neural Eng.*, 2017, doi: 10.1088/1741-2552/aa6802.
- [51] M. Markovic, S. Dosen, C. Cipriani, D. Popovic, and D. Farina, “Stereovision and augmented reality for closed-loop control of grasping in hand prostheses,” *J. Neural Eng.*, 2014, doi: 10.1088/1741-2560/11/4/046001.
- [52] M. Markovic, S. Dosen, D. Popovic, B. Graitmann, and D. Farina, “Sensor fusion and computer vision for context-aware control of a multi degree-of-freedom prosthesis,” *J. Neural Eng.*, 2015, doi: 10.1088/1741-2560/12/6/066022.
- [53] R. A. Romeo *et al.*, “Identification of slippage on naturalistic surfaces via wavelet transform of tactile signals,” *IEEE Sens. J.*, 2019, doi: 10.1109/JSEN.2018.2881831.
- [54] L. Osborn, R. R. Kaliki, A. B. Soares, and N. V. Thakor, “Neuromimetic Event-Based Detection for Closed-Loop Tactile Feedback Control of Upper Limb Prostheses,” *IEEE Trans. Haptics*, 2016, doi: 10.1109/TOH.2016.2564965.
- [55] B. J. Shastri *et al.*, “Photonics for artificial intelligence and neuromorphic computing,” *Nature Photonics*. 2021, doi: 10.1038/s41566-020-00754-y.
- [56] J. Tang *et al.*, “Bridging Biological and Artificial Neural Networks with Emerging Neuromorphic Devices: Fundamentals, Progress, and Challenges,” *Advanced Materials*. 2019, doi: 10.1002/adma.201902761.
- [57] M. V. Debole *et al.*, “TrueNorth: Accelerating From Zero to 64 Million Neurons in 10 Years,” *Computer (Long Beach, Calif.)*, 2019, doi: 10.1109/MC.2019.2903009.

- [58] A. Andreopoulos, H. J. Kashyap, T. K. Nayak, A. Amir, and M. D. Flickner, "A Low Power, High Throughput, Fully Event-Based Stereo System," 2018, doi: 10.1109/CVPR.2018.00786.
- [59] A. Guyton and J. Hall, "Fundamentos de Fisiologia," in *Fundamentos de fisiologia vegetal*, 2012.
- [60] Z. O. Repository and M. Library, "Fast Event-based Corner Detection," no. September, pp. 1–8, 2017.
- [61] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*. 2015, doi: 10.1038/nature14539.
- [62] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," 2012, doi: 10.1061/(ASCE)GT.1943-5606.0001284.
- [63] M. J. Shafiee, B. Chywl, F. Li, and A. Wong, "Fast YOLO: A fast you only look once system for real-Time embedded object detection in video," *arXiv*. 2017, doi: 10.15353/vsnl.v3i1.171.
- [64] S. Jegou, M. Drozdal, D. Vazquez, A. Romero, and Y. Bengio, "The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation," 2017, doi: 10.1109/CVPRW.2017.156.
- [65] J. Liang, L. Jiang, J. C. Niebles, A. Hauptmann, and L. Fei-Fei, "Peeking into the future: Predicting future person activities and locations in videos," 2019, doi: 10.1109/CVPRW.2019.00358.
- [66] S. Yeung *et al.*, "A computer vision system for deep learning-based detection of patient mobilization activities in the ICU," *npj Digit. Med.*, 2019, doi: 10.1038/s41746-019-0087-z.
- [67] R. Sun, D. Shi, Y. Zhang, R. Li, and R. Li, "Data-driven technology in event-based vision," *Complexity*. 2021, doi: 10.1155/2021/6689337.
- [68] J. Cox, A. Ashok, and N. Morley, "An analysis framework for event-based sensor performance," 2020, doi: 10.1117/12.2567620.
- [69] S. R. Kheradpisheh, M. Ganjtabesh, S. J. Thorpe, and T. Masquelier, "STDP-based spiking deep convolutional neural networks for object recognition," *Neural Networks*, 2017, doi: <https://doi.org/10.1016/j.neunet.2017.12.005>.
- [70] X. Lagorce, G. Orchard, F. Galluppi, B. E. Shi, and R. B. Benosman, "HOTS: A Hierarchy of Event-Based Time-Surfaces for Pattern Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, doi: 10.1109/TPAMI.2016.2574707.
- [71] M. Cannici, M. Ciccone, A. Romanoni, and M. Matteucci, "Asynchronous convolutional networks for object detection in neuromorphic cameras," *IEEE Comput. Soc. Conf. Comput. Vis.*

- Pattern Recognit. Work.*, vol. 2019-June, pp. 1656–1665, 2019, doi: 10.1109/CVPRW.2019.00209.
- [72] C. Tan, S. Lallee, and G. Orchard, “Benchmarking neuromorphic vision: Lessons learnt from computer vision,” *Front. Neurosci.*, 2015, doi: 10.3389/fnins.2015.00374.
- [73] Y. Hu, S. C. Liu, and T. Delbruck, “V2e: From video frames to realistic DVS events,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 1312–1321, 2021, doi: 10.1109/CVPRW53098.2021.00144.
- [74] Y. Hu, H. Liu, M. Pfeiffer, and T. Delbruck, “DVS benchmark datasets for object tracking, action recognition, and object recognition,” *Front. Neurosci.*, vol. 10, no. AUG, pp. 1–5, 2016, doi: 10.3389/fnins.2016.00405.
- [75] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, and D. Scaramuzza, “The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM,” *Int. J. Rob. Res.*, 2017, doi: 10.1177/0278364917691115.
- [76] G. Orchard, A. Jayawant, G. K. Cohen, and N. Thakor, “Converting static image datasets to spiking neuromorphic datasets using saccades,” *Front. Neurosci.*, 2015, doi: 10.3389/fnins.2015.00437.
- [77] S. Shao *et al.*, “Objects365: A large-scale, high-quality dataset for object detection,” 2019, doi: 10.1109/ICCV.2019.00852.
- [78] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The Pascal Visual Object Classes Challenge: A Retrospective,” *Int. J. Comput. Vis.*, 2015, doi: 10.1007/s11263-014-0733-5.
- [79] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” 2010, doi: 10.1109/cvpr.2009.5206848.
- [80] T.-Y. Lin *et al.*, “Microsoft COCO: Common Objects in Context,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2015.
- [81] D. R. Valeiras, X. Lagorce, X. Clady, C. Bartolozzi, S. H. Ieng, and R. Benosman, “An asynchronous neuromorphic event-driven visual part-based shape tracking,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 26, no. 12, pp. 3045–3059, 2015, doi: 10.1109/TNNLS.2015.2401834.
- [82] V. Vasco, A. Glover, and C. Bartolozzi, “Fast event-based Harris corner detection exploiting the

- advantages of event-driven cameras,” *IEEE Int. Conf. Intell. Robot. Syst.*, vol. 2016-Novem, pp. 4144–4149, 2016, doi: 10.1109/IROS.2016.7759610.
- [83] C. Harris and M. Stephens, “A Combined Edge and Corner Detector,” 1988.
- [84] J. D. Schall, “Neural Basis of Saccade Target Selection,” *Rev. Neurosci.*, 1995, doi: 10.1515/REVNEURO.1995.6.1.63.
- [85] J. D. Schall and D. P. Hanes, “Neural basis of saccade target selection in frontal eye field during visual search,” *Nature*, 1993, doi: 10.1038/366467a0.
- [86] J. D. Schall and J. Y. Cohen, “The neural basis of saccade target selection,” in *The Oxford Handbook of Eye Movements*, 2012.
- [87] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” 2016, doi: 10.1109/CVPR.2016.91.
- [88] W. Budiharto, E. Irwansyah, J. S. Suroso, and A. A. S. Gunawan, “Design of object tracking for military robot using pid controller and computer vision,” *ICIC Express Lett.*, 2020, doi: 10.24507/icicel.14.03.289.
- [89] X. Lagorce, G. Orchard, F. Galluppi, B. E. Shi, and R. B. Benosman, “HOTS: A Hierarchy of Event-Based Time-Surfaces for Pattern Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1346–1359, 2017, doi: 10.1109/TPAMI.2016.2574707.
- [90] R. Benosman, S. H. Ieng, C. Clercq, C. Bartolozzi, and M. Srinivasan, “Asynchronous frameless event-based optical flow,” *Neural Networks*, 2012, doi: 10.1016/j.neunet.2011.11.001.
- [91] E. Mueggler, C. Bartolozzi, and D. Scaramuzza, “Fast event-based corner detection,” 2017, doi: 10.5244/c.31.33.
- [92] T. Ridnik, H. Lawen, A. Noy, E. Ben, B. G. Sharir, and I. Friedman, “TRResNet: High performance GPU-dedicated architecture,” 2021, doi: 10.1109/WACV48630.2021.00144.
- [93] C.-Y. Wang, H.-Y. M. Liao, I.-H. Yeh, Y.-H. Wu, P.-Y. Chen, and J.-W. Hsieh, “CSPNet: A New Backbone that can Enhance Learning Capability of CNN,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, 2019.
- [94] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2015, doi: 10.1109/TPAMI.2015.2389824.
- [95] L. Fei-Fei, R. Fergus, and P. Perona, “Learning generative visual models from few training

- examples: An incremental bayesian approach tested on 101 object categories,” 2004, doi: 10.1109/CVPR.2004.383.
- [96] Lecun Yann, Cortes Corinna, and Burges Christopher, “THE MNIST DATABASE of Handwritten Digits,” *Courant Inst. Math. Sci.*, 1998.
- [97] L. R. Iyer, Y. Chua, and H. Li, “Is Neuromorphic MNIST Neuromorphic? Analyzing the Discriminative Power of Neuromorphic Datasets in the Time Domain,” *Front. Neurosci.*, 2021, doi: 10.3389/fnins.2021.608567.
- [98] S. H. Ieng, C. Posch, and R. Benosman, “Asynchronous neuromorphic event-driven image filtering,” *Proc. IEEE*, 2014, doi: 10.1109/JPROC.2014.2347355.
- [99] S. H. Bengtson, T. Bak, L. N. S. Andreasen Struijk, and T. B. Moeslund, “A review of computer vision for semi-autonomous control of assistive robotic manipulators (ARMs),” *Disability and Rehabilitation: Assistive Technology*. 2020, doi: 10.1080/17483107.2019.1615998.
- [100] M. Rastegari, V. Ordonez, J. Redmon, and A. Farhadi, “XNOR-net: Imagenet classification using binary convolutional neural networks,” 2016, doi: 10.1007/978-3-319-46493-0\_32.
- [101] T. Delbruck and M. Lang, “Robotic goalie with 3 ms reaction time at 4% CPU load using event-based dynamic vision sensor,” *Front. Neurosci.*, 2013, doi: 10.3389/fnins.2013.00223.
- [102] S. Tang, R. Ghosh, N. V. Thakor, and S. L. Kukreja, “Orientation estimation and grasp type detection of household objects for upper limb prostheses with dynamic vision sensor,” *Proc. - 2016 IEEE Biomed. Circuits Syst. Conf. BioCAS 2016*, vol. 1, no. c, pp. 99–102, 2017, doi: 10.1109/BioCAS.2016.7833734.
- [103] R. Ghosh, A. Mishra, G. Orchard, and N. V. Thakor, “Real-time object recognition and orientation estimation using an event-based camera and CNN,” *IEEE 2014 Biomed. Circuits Syst. Conf. BioCAS 2014 - Proc.*, pp. 544–547, 2014, doi: 10.1109/BioCAS.2014.6981783.