UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

MODELO E FERRAMENTA PARA RECONHECIMENTO E CLASSIFICAÇÃO DE GESTOS DO CORPO

GUSTAVO JORDAN CASTRO BRASIL

ORIENTADOR: PROF. DR. LUIS CARLOS TREVELIN

São Carlos – SP

Agosto/2017

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

MODELO E FERRAMENTA PARA RECONHECIMENTO E CLASSIFICAÇÃO DE GESTOS DO CORPO

GUSTAVO JORDAN CASTRO BRASIL

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação, área de concentração: Sistemas Distribuídos e Redes de Computadores
Orientador: Prof. Dr. Luis Carlos Trevelin

São Carlos – SP

Agosto/2017

Brasil, Gustavo Jordan Castro

Modelo e Ferramenta para Reconhecimento e Classificação de Gestos do Corpo / Gustavo Jordan Castro Brasil. -- 2017.
127 f. : 30 cm.

Dissertação (mestrado)-Universidade Federal de São Carlos, campus São Carlos, São Carlos
Orientador: Luis Carlos Trevelin
Banca examinadora: Luis Carlos Trevelin (UFSCar), Sergio Donizetti Zorzo (UFSCar), Diego Roberto Colombo Dias (UFSJ)
Bibliografia

1. Reconhecimento de Gestos. 2. Realidade Virtual. 3. Reconhecimento de Padrões. I. Orientador. II. Universidade Federal de São Carlos. III. Título.

Ficha catalográfica elaborada pelo Programa de Geração Automática da Secretaria Geral de Informática (SIn).
DADOS FORNECIDOS PELO(A) AUTOR(A)

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia

Programa de Pós-Graduação em Ciência da Computação

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de Dissertação de Mestrado da candidata GUSTAVO JORDAN CASTRO BRASIL, realizada em 18/08/2017.

Prof. Dr. Luis Carlos Trevelin
(UFSCar)

Prof. Dr. Sergio Donizetti Zorzo
(UFSCar)

****

Prof. Dr. Diego Roberto Colombo Dias
(UFSJ)

Certifico que a sessão de defesa foi realizada com a participação à distância do membro Diego Roberto Colombo Dias, depois das arguições e deliberações realizadas, o participante à distância está de acordo com o conteúdo do parecer da comissão examinadora redigido no relatório de defesa do aluno GUSTAVO JORDAN CASTRO BRASIL.

Prof. Dr. Luis Carlos Trevelin
Presidente da Comissão Examinadora
(UFSCar)

Aos meus pais, irmãos e esposa, Pâmela

AGRADECIMENTOS

Agradeço ao professor Dr. Luis Carlos Trevelin por ter me aceito no ingresso ao mestrado e por ter me acompanhado e orientado ao longo dessa jornada com confiança, paciência, dedicação e amizade.

Agradeço aos meus colegas de laboratório do LaVIIC (Laboratório de Visualização Imersiva, Interativa e Colaborativa) e de trabalho acadêmico, que me impulsionaram e colaboraram para o meu sucesso acadêmico: Marcelo de Paiva Guimarães, Bruno Barberi Gnecco, Diego Roberto Colombo Dias, Alexandre Fonseca Brandão e Alfredo Guilherme da Silva Souza.

Agradeço infinitamente a meu pai, por ser meu pai e pelos conselhos e cobranças ao longo da minha vida e minha mãe pela sua energia, dedicação e paciência comigo.

Agradeço a minha esposa Pâmela Diniz de Campos a paciência e confiança comigo durante o Mestrado e pelo seu empenho em me ajudar a superar esta fase de minha vida.

Agradeço a todos os funcionários do Programa de Pós-graduação em Ciência da Computação (PPG-CC) do Departamento de Computação (DC) e da Universidade Federal de São Carlos que proporcionam este lugar incrível de conhecimento e aprendizado.

Por fim agradeço ao meu anjo da guarda pelo apoio, e a Deus por esta oportunidade impar nesta vida.

Você não pode voltar atrás e fazer um novo começo, mas você pode começar agora e fazer um novo fim.

Francisco Cândido Xavier

RESUMO

Interfaces multimodais estão cada vez mais populares, e demandam mais a interação natural como recurso para enriquecer a experiência do usuário. Sistemas computacionais que suportam a multimodalidade provêm um modo mais natural e flexível para execução de tarefas em computadores, uma vez que permitem aos usuários com diferentes níveis de habilidades e de aprendizado, escolha o modo de interação mais adequado a suas necessidades. Dentre as formas de interação natural, estão os gestos, a interação natural através de gestos vem se popularizando cada vez mais, visto que, foge do estilo convencional de interação baseado em teclado e mouse, e ainda pelo crescimento e advento de dispositivos de captura de movimento, com sensores visuais de profundidade de baixo custo. Neste contexto, esta dissertação apresenta um estudo sobre todas as etapas necessárias para a construção de um modelo e ferramenta para reconhecimento de gestos estáticos e dinâmicos, sendo estas: Segmentação; Modelagem; Descrição; e Classificação. Soluções e resultados propostos, são apresentados para cada uma destas etapas e, por fim uma ferramenta que implementa o modelo é avaliada no reconhecimento de gestos, utilizando um conjunto finito de gestos. Todas as soluções apresentadas nesta dissertação foram encapsuladas na ferramenta GG-Gesture, que tem por objetivo simplificar as pesquisas na área de reconhecimento de gestos, permitindo a comunicação com sistemas de interfaces multimodais e interfaces naturais.

Palavras-chave: Reconhecimento de Gestos, Realidade Virtual, Reconhecimento de Padrões, Interface Natural

ABSTRACT

Multimodal interfaces are becoming more popular, and increasingly require natural interaction as a resource to enrich the user experience. Computational systems that support multimodality provide a more natural and flexible way to perform tasks on computers, since they allow users with different levels of skills and knowledge to choose the mode of interaction best suited to their needs. Among natural forms of interaction are the gestures, the natural interaction through gestures is becoming more popular, since it's an alternative to the conventional style of interaction based on keyboard and mouse, and also by the growth and advent of motion capture devices, with low-cost visual depth sensors. In this context, this dissertation presents a study about all the necessary steps for the construction of a model and tool for the recognition of static and dynamic gestures, these being: Segmentation; Modeling; Description; and Classification. Proposed solutions and results are presented for each of these steps and, finally, a tool that implements the model is evaluated in the recognition of gestures, using a finite set of gestures. All the solutions presented in this dissertation were encapsulated in the GGGesture tool, which aims to simplify research in the area of gesture recognition, allowing communication with multimodal interfaces systems and natural interfaces.

Keywords: Gesture Recognition, Virtual Reality, Pattern Recognition, Natural Interface

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ALGORITMOS

LISTA DE ABREVIATURAS E SIGLAS

API – *Application Programming Interface*

CLI – *Command Line Interface*

CMOS – *Complementary Metal-Oxide-Semiconductor*

CSV – *Comma-separated values*

DTW – *Dynamic Time Warping*

GUI – *Graphical User Interface*

HMM – *Hidden Markov Model*

IHCM – *Interaction Multimodal Human-Computer*

IHC – *Interação Humano-Computador*

LGPL – *GNU Lesser General Public Lincese*

MS SDK – *Microsoft SDK*

NUI – *Natural User Interface*

NURBS – *Non Uniform Rational Basis Spline*

RA – *Realidade Aumentada*

RV – *Realidade Virtual*

SDK – *Open Natural Interaction*

SDK – *Software Development Kit*

SVM – *Support Vector Machine*

ToF – *Time-of-Flight*

UML – *Unified Modeling Language*

USB – *Universal Serial Bus*

WIMP – *Windows, Icons, Menus and Pointers*

kNN – *k-Nearest Neighbor*

SUMÁRIO

Capítulo 1

INTRODUÇÃO

Este capítulo apresenta uma visão geral da área de pesquisa, assim como uma breve descrição dos objetivos e motivação do projeto.

1.1 Apresentação

Os atuais avanços e investimentos no desenvolvimento de tecnologias de componentes para *hardware* tem levado a um crescente aumento na capacidade dos computadores, tornando-os menores, mais velozes e *commodities*. Isso vem possibilitando uma ampla difusão desses equipamentos pelos mais diversos ambientes, atingindo um número cada vez maior de usuários, dispositivos como *smartphones*, *notebooks*, sistemas embarcados, e entre outros em geral já se tornaram parte do cotidiano das pessoas.

Entretanto, há um problema inerente ao uso de computadores de que muitas das vezes os usuários são obrigados a aprender a utilizá-los, mudar seus hábitos e métodos de trabalho (DIX, 2004). Embora a evolução dos *software* tenha acompanhado de perto o desenvolvimento do *hardware* dos computadores, há uma enorme demanda por sistemas de alta usabilidade que possuam interfaces de interação natural, dinâmica e acessível de forma que não sejam intrusivas, ou seja, que não interfiram no modo em que as pessoas realizam suas tarefas.

Nesse sentido, os contínuos estímulos empregados em pesquisas na área de Interação Humano-Computador (IHC) resultaram no desenvolvimento de interfaces computacionais que suportam formas mais naturais de comunicação, por meio de voz, visão, toque e gestos (DIX, 2004). Assim, as Interfaces Multimodais surgiram para tomar vantagem dessas novas modalidades de interação, possibilitando uma maior comunicação com o sistema multissensorial humano através do processo simultâneo de diferentes tipos de entrada (OVIATT; COHEN, 2000).

A oportunidade de se interagir com o computador por meio de gestos do corpo, pode enriquecer substancialmente as interfaces de IHC, pois é uma linguagem de comunicação já utilizada pelos humanos e representa uma perspectiva de interação que pode influenciar enormemente a produtividade dos usuários e permite uma maior expressividade na comunicação homem-máquina aumentando a variabilidade de tarefas que podem ser executadas simultaneamente, para assim, unir as demais formas de comunicação e dispositivos hoje existentes, na concepção de Interfaces Multimodais.

Reconhecimento de Gestos é uma área de pesquisa que vem sendo amplamente abordada na academia e indústria, por ser uma tarefa de alta complexidade, o seu desenvolvimento envolve diversos tópicos e áreas do conhecimento como IHC, Visão Computacional, Reconhecimento e Análise de Padrões, entre outros.

A grande maioria dos trabalhos desenvolvidos visa o reconhecimento de interações gestuais do corpo em cenários estritos a uma aplicação, além disso estão limitados a um conjunto de gestos e poses pré-definidos segmentados a partir da utilização de soluções vestíveis ou marcadores para a captura do movimento.

Porém, o modelo de reconhecimento de gestos do corpo para aplicação e implementação em sistemas que adota a Interface Multimodal faz a exigência de aspectos mais extensos, que utilizem outras variáveis que devem ser levadas em consideração, como por exemplo, personalização e segmentação de gestos, identificação e descrição dos membros do corpo e dos gestos, posicionamento relativo entre eles, captura e rastreamento tridimensional dos movimentos e um classificador para o reconhecimento.

Neste contexto, um sistema de reconhecimento de gestos do corpo pode ser subdividido em quatro etapas, como indicado na Figura 1.1:

Fonte: Elaborada pelo autor

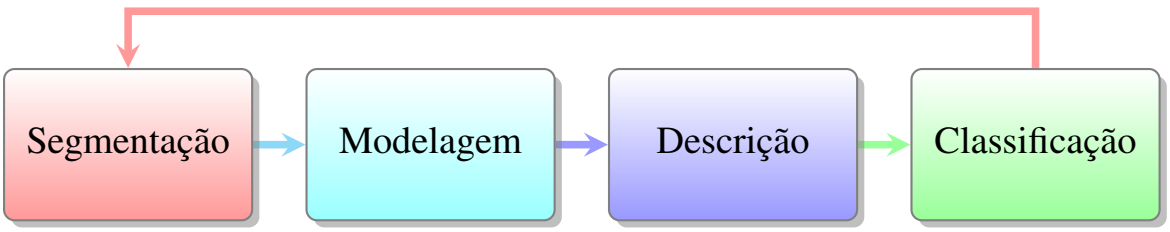


Figura 1.1: Etapas de um sistema de reconhecimento de gestos do corpo

Segmentação: é o processo para particionamento do espaço de dados em regiões que sobressaem e trazem significado ao domínio de aplicação, pré-requisito para o processo de reconhecimento de objetos, a eficácia do reconhecimento é determinante por este processo onde

deve-se atentar as características fundamentais dos dados a serem analisados.

Modelagem: é o processo de organizar e estruturar as etapas para captura e representação das informações obtidas.

Descrição: é o processo que descrever a unidade de um conjunto de informações, parâmetros e características extraídas a partir da segmentação e modelo. A descrição torna-se uma forma para o registro e análise, bem como para o reconhecimento.

Classificação: é o processo de comparação de varias instancias de descrições, no qual busca classificar a descrição mais correspondente, para se alcançar o reconhecimento.

1.2 Motivação e Relevância

Uma das principais características da pesquisa sobre o gesto é a sua natureza multidisciplinar. A pesquisa filosófica sobre gesto permite uma investigação profunda sobre os mecanismos de comunicação entre humanos, como exemplo, nas áreas de psicologia, ciências sociais, artes e humanas. Tais áreas do conhecimento podem ser exploradas em pesquisas de computação como o design de Interação Humano Computador.

Esta natureza interdisciplinar pode beneficiar muito as pesquisas e abrir novas perspectivas em todos os campos. Se, por um lado, a investigação computacional pode crescer a partir de modelos e teorias cedidas da psicologia, ciências sociais, artes e humanas, por outro lado, essas disciplinas podem começar a usar com maior confiança as tecnologias computacionais, fornecendo assim ferramentas para a sua própria pesquisa, ou seja, examinar, a uma profundidade que nunca foi alcançada. Reconhecer e interpretar os gestos do corpo humano é uma tarefa importante para IHC, a comunicação através dos gestos é uma das formas mais naturais de interação dos seres humanos.

O reconhecimento dos gestos em seu caráter natural aplicado a uma interface computacional para gerar ações e interações em sistemas torna-se objetivo principal deste trabalho. A comunicação e interação através dos gestos deve ser obtida sem utilização de quaisquer mecanismos adicionais, que são usualmente utilizados para auxiliar a tarefa de segmentação, modelagem e reconhecimento, para que se torne efetivamente uma interação natural.

Suprimir a necessidade da utilização de marcadores, roupas de sensoriamento, dispositivos magnéticos de captura tridimensional, ou qualquer outro dispositivo mecânico que auxilia a fase de segmentação, modelagem, estimação do posicionamento das partes do corpo humano é essencial para alcançar a naturalidade. Portanto, o foco deste trabalho é o estudo de outros

mecanismos e dispositivos que sejam capazes de obter informações e características partindo de soluções não intrusivas como imagens visuais através de câmeras.

Aplicação do reconhecimento de gestos é vista em ambientes restritos e com uso de mecanismos intrusivos que auxiliam nos processos da classificação do gesto, apesar de resultados sólidos, possuem desvantagens de serem desconfortáveis e não vão de encontro a interação natural (AGGARWAL; RYOO, 2011), neste sentido podemos considerar que quanto mais natural for a forma de interação por gestos do corpo, mais complexa será a tarefa de segmentação e interpretação do reconhecimento de gestos.

A busca por esta solução de reconhecimento de gestos de forma natural é evidenciada por produtos presentes no mercado como LEAP Motion[1], Asus Xtion PRO Live[2], Structure Sensor[3] entre outros que visam o reconhecimento e interpretação de movimentos do corpo e das mãos.

Um dos primeiros dispositivos comerciais a ganhar destaque nesta área foi o Microsoft Kinect [4] , desenvolvido pela Microsoft, é um exemplo de arranjo de software e hardware que está presente no mercado com o objetivo de mapear a movimentação de partes do corpo humano. Sua principal funcionalidade está na simplificação da segmentação das imagens, utilizando a técnica de estimação de profundidade apresentada na Subseção (2.4.4).

A interação natural por meio de gestos é recente e ainda carente de padrões e modelos. Sua aplicação tem sido restrita a cenários específicos, e estão limitados a um conjunto de gestos ou poses pré-definidos a um *software* ou hardware, deste modo, os usuários têm que gastar tempo e esforço para ensaiar os gestos programados para coincidir com os do sistema, não permitindo personalização de gestos. No sentido da interação natural, permitir a personalização de gestos e poses, torna-se um requisito essencial, a falta de modelos definidos e abertos gera aplicações isoladas, que não analisam tópicos importantes de interoperabilidade, no entanto poucos estudos apresentam efetivamente a modelagem de poses e gestos do corpo para aplicação em sistemas.

A possibilidade de uso do reconhecimento de gestos como interface natural, abre espaço para muitas oportunidades no desenvolvimento de novas soluções comerciais e acadêmicas em diversas áreas. Este trabalho tem como o objetivo de abrir tais possibilidades, disponibilizando o reconhecimento de gestos do corpo, através de uma abordagem multidisciplinar para o problema, com base em alguns dos princípios conhecidos de como os seres humanos reconhecem os gestos, juntamente com os métodos da ciência da computação.

[1]Informações: *https://www.leapmotion.com/*
[2]Informações: *https://www.asus.com/us/Multimedia/Xtion_PRO_LIVE/*
[3]Informações: *http://structure.io/*
[4]Informações: *https://www.microsoft.com/en-us/kinectforwindows/*

1.3 Objetivo

Na tarefa do reconhecimento de gestos os seres humanos possuem a capacidade de processar vários fatores, desde o volume muscular, velocidade do movimento e/ou expressões faciais, em múltiplos sentidos, e simultaneamente analisar a ação, enquanto um sistema computacional possui uma limitação de dados disponíveis para um ou dois canais através de sensores (ZHAO; BADLER, 2001) . O reconhecimento de gestos é uma tarefa complexa que envolve muitos aspectos, tais como modelagem do movimento, análise do movimento e padrão de reconhecimento, além disto os estudos psicolinguísticos.

Este trabalho, tem por objetivo o estudo de modelos computacionais de gestos como interface de interação humano computador, e propor uma ferramenta capaz de segmentar objetos que estão se movendo diante de sensores visuais, e dentre estes objetos identificar a representação humana, e modelar as partes do corpo humano necessárias, para posteriormente, reconhecer um gesto executado, comparando em um conjunto de gestos salvos em um dicionário de gestos, tendo como contemplação a execução do reconhecimento dos gestos. Ademais, a proposta deste trabalho é apresentar um estudo comparativo das soluções existentes, em todas as etapas que compõe a complexa tarefa do reconhecimento de gestos, sendo elas: Segmentação; Modelagem; Descrição; e Classificação.

Por fim, tal objetivo almeja resultar numa ferramenta de reconhecimento de gestos disponível para uso em diferentes aplicações, de modo a fornecer uma interface intuitiva e natural, que minimize o esforço cognitivo dos usuários, e que permita interagir e realizar outras tarefas comuns através de gestos. É importante ressaltar que faz parte da proposta utilizar-se de soluções e tecnologias abertas e padronizadas para alcançar o mesmo.

1.4 Organização do Trabalho

Este trabalho está dividido na seguinte disposição: O Capítulo 1, traz a Introdução, Apresentação, Motivação e Relevância, Objetivo e Organização do Trabalho;

O Capítulo 2, é realizado um estudo geral acerca das disciplinas relevantes para este trabalho bem como as soluções propostas que estão subdivididas e organizadas em 5 Subcapítulos: Os Subcapítulos 1; 2; 3; e 4; abordam a revisão bibliográfica sobre IHC, Interface Natural de Usuário, Multimodalidade e Taxonomia dos Gestos; No Subcapítulo 5, é apresentado o contexto e trabalhos relatados de Modelagem do Gesto e Reconhecimento de Gestos;

No Capítulo 3, descreve-se o Modelo Proposto para realizar o reconhecimento de gestos estáticos e dinâmicos, que suporta o Registro e Classificação do Gestos Espaciais Tridimensionais, bem como é demonstrado a utilização do Modelo Proposto com o sensor visual Microsoft Kinect;

Posteriormente no Capítulo 4, apresenta os Experimentos realizados através da analise, desenvolvimento e aplicação do Modelo Proposto na Ferramenta GGGesture. Ademais é relatado Casos de Usos, resultados gerais e avaliações de execuções;

E finalmente no Capítulo 5, é apresentado as Conclusões, Produções Científicas e Tecnológicas da Pesquisa e Trabalhos Futuros.

Capítulo 2

FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados fundamentos e conceitos importantes acerca de várias disciplinas relacionadas a este trabalho, juntamente com referências a trabalhos mais relevantes existentes na literatura. Devido a ampla abrangência dessas disciplinas, os mesmos serão abordados sob o intuito de fornecer toda uma fundamentação sólida para servir de base para algumas decisões abordadas para alcançar seu objetivo final.

2.1 Interação Humano Computador

A área de IHC é a intersecção entre a Ciência da Computação e a Ciência do Comportamento. Envolve o estudo, o planejamento, e o ato de projetar a interação entre as pessoas (referidas neste trabalho como usuários) e os computadores. Dar atenção à IHC é um fator importante no contexto da computação, interfaces mal projetadas podem gerar como consequências, riscos e problemas inesperados.

Segundo Hewett et al. (1992), a área de IHC tem como objetivo principal fornecer aos pesquisadores e desenvolvedores de sistemas explicações e previsões para fenômenos de interação do usuário com o sistema e resultados práticos para o *design* da Interface de Usuário. Com teorias a respeito dos fenômenos envolvidos, seria possível prever antecipadamente se o sistema a ser desenvolvido satisfaz as necessidades de usabilidade, aplicabilidade, acessibilidade e comunicabilidade dos usuários, além da funcionalidade prática de um sistema.

2.1.1 Estilos de Interação

Estilo de Interação é um termo genérico que inclui todas as formas como os usuários se comunicam ou interagem com sistemas computacionais (SHNEIDERMAN; PLAISANT, 2010; PRE-

ECE et al., 1994).

O homem usualmente empenha no processo de comunicação diferentes canais: como voz, gestos, expressões e movimentos. Estudos na área de IHC buscam criar sistemas de interação natural que assimilam tais canais e permitam envolver os indivíduos e o ambiente em um diálogo natural.

A restrição de comunicação do homem com a máquina aos longos dos anos determinada pelas tecnologias existentes, determinou inicialmente que o homem adapta-se a linguagem da máquina, entretanto a evolução e o advento de novas tecnologias tem possibilitado com que as máquinas se adaptem à linguagem humana (NORMAN, 2013). Segundo Valli (2008), o grande desafio dos designers de interação é a concepção de novos paradigmas de interação que explore a capacidade de compreensão das máquinas da mesma maneira em que os humanos descobrem o mundo real.

Algumas tecnologias e estilos de interação aplicados em sistemas têm dispensado a necessidade de aprendizado de novos processos, comandos, linguagens ou dispositivos eletrônicos. Segundo Prates e Barbosa (2003), o processo de interação usuário-sistema atua principalmente do ponto de vista usuário: as ações que ele realiza usando a interface de um sistema, e suas interpretações das respostas transmitidas pelo sistema através da interface, demonstrada na Figura 2.1.

Fonte: Adaptada de (PRATES; BARBOSA, 2003)

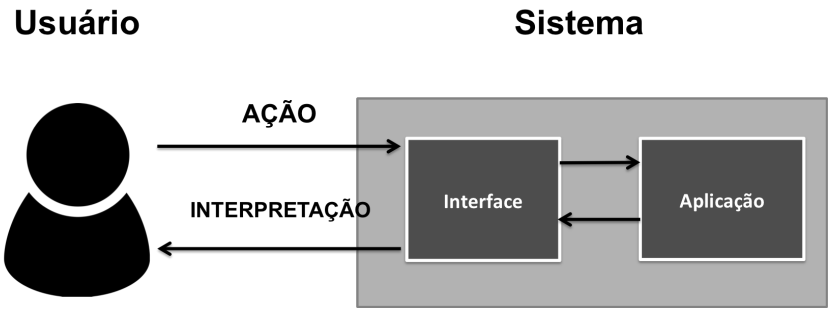


Figura 2.1: Processo Interação Humano-Computador

Pode-se destacar os seguintes estilos de interação como os mais citados: linguagem natural, linguagens de comando, manipulação direta e WIMP (*Window-Icon-Menu-Pointer*) que estão descritos resumidamente a seguir.

- Linguagem Natural

  A comunicação do usuário com o sistema é por meio da fala e/ou expressões escritas, linguagem natural que utilizada pelos seres humanos. A interação em linguagem natural

é bastante atrativa para os usuários com pouco ou nenhum conhecimento e experiência em computação, por ser uma interface de comunicação e interação natural . Para permitir que um usuário interaja com aplicações em linguagem natural, pode-se disponibilizar uma entrada de áudio através de um microfone e suas falas podem expressar comandos ou navegações no sistema, ou até mesmo uma interface textual onde ele possa inserir as frases expressando comandos. Em aplicações que utilizam esse tipo de estilo de interação, tenta-se aproximar a aplicação ao usuário, aumentando o nível de abstrações e privilegiando a forma de comunicação. Exemplos de aplicações deste tipo de interação mais populares: *Google Now* (INC, 2014b), *Siri* (INC, 2014a) e *Cortana* (INC, 2014c).

- Manipulação Direta

  A interação de manipulação direta é aquela que permite ao usuário agir diretamente sobre os objetos da aplicação (representações gráficas e/ou dados) sem a necessidade de comandos de uma linguagem específica ou seleção de menus. Neste tipo de interação, os comandos são geralmente ações baseadas numa analogia entre o cursor e uma mão. Há sistemas operacionais que proporcionam este estilo na interface gráfica no qual os usuários podem interagir com o gerenciador de arquivos do sistema através de manipulação de ícones que representam arquivos, diretórios, discos e outros componentes computacionais. O usuário interage com ícones, utilizando o mouse ou dispositivo equivalente, através de ações do tipo clicar, arrastar (*drag-and-drop*), *swipe* etc. Este estilo vem utilizado por sistemas de Realidade Virtual (RV) , Realidade Aumentada (RA) e *Mobiles*, onde o usuário pode interagir com objetos através de gestos produzidos pelas mãos, dedos e/ou corpo. Podem-se encontrar exemplo em aplicações que utilizam o *design* Interface Natural de Usuário (NUI | *Natural User Interface*) para implementação.

- Linguagem de Comando

  Proporciona ao usuário a possibilidade de enviar instruções diretamente ao sistema através de comandos específicos (PREECE et al., 1994). As linguagens de comandos podem ser consideradas poderosas por oferecerem acesso direto às funcionalidades do sistema. Contudo, este poder implica maior dificuldade de aprendizado para usuários iniciantes, pelo fato da necessidade de aprender os comandos, e que na maioria das vezes são comandos compostos, e precisam ser relembrados. Em uma aplicação com linguagem de comandos, tenta-se aproximar o usuário em direção à aplicação, ao contrário do que acontece com aplicações de linguagem natural. Aplicações que utilizam o *design* de Interface de Comando (CLI | *Command Line Interface*) para implementação de interação são exemplos de uso deste estilo.

- WIMP

  O estilo de interação Janelas, Ícones, Menus e Apontadores permite a interação através de componentes de interação virtuais denominados *widgets* (SHNEIDERMAN; PLAISANT, 2010). Este estilo é implementado com o auxílio das interfaces gráficas, que proporcionam o desenho de janelas e o controle de entrada através do teclado e do mouse em cada uma destas janelas. WIMP não deve ser considerado como um estilo único, mas a junção de vários estilos, a medida, que é muito comum identificar outros estilos como os de menus, manipulação direta, preenchimento de formulário e linguagem de comandos embutidos livremente em interfaces de estilo WIMP. Comumente Interfaces Gráficas de Usuário (GUI | Graphical User Interface) utilizam esse estilo para interação.

2.1.2 Interface

A interface dos computadores é parte de um todo de um sistema com o qual um usuário mantém contato ao utilizar, tanto ativamente quanto passivamente (PREECE et al., 1994). Segundo Moran (1981), a interface de usuário pode ser a parte de um sistema computacional com qual uma pessoa entra em contato fisicamente, perspectivamente e/ou conceitualmente.

Dentro disto, podemos dizer que a dimensão física inclui elementos que o usuário pode manipular, como teclado, mouse, etc. A perspectiva inclui uma interface de interação onde o usuário percebe elementos e analisa. E no conceitual inclui os processos de interpretação e raciocínio do usuário.

2.2 Interface Natural de Usuário

A Interface Natural de Usuário ou *Natural User Interface* e desenvolvedores de interfaces de computador no qual buscam a interação com o computador de um modo natural, ou tornando-se natural com sucessivas interações.

Segundo Wigdor e Wixon (2011), NUI não é uma interface natural, mas sim uma interface que permite seu usuário agir e sentir como se fosse natural, pois descartam o uso de dispositivos artificiais tornando o corpo o instrumento de interação.

Um dos grandes desafios da área de IHC é o de reduzir a curva de aprendizagem do usuário na operabilidade das aplicações computacionais. Uma interface NUI exige apenas que o usuário seja capaz de interagir com o ambiente por meio de interações previamente já conhecidas, como, por exemplo, gestos e voz. Esse tipo de interface também exige aprendizagem, porém

é facilitada, pois não exige que o usuário seja apresentado a um novo dispositivo de controle artificial.

A interação natural baseada em gestos exige que as aplicações estejam preparadas para receber e manipular os eventos entrada do usuário através de novos dispositivos como câmeras e microfones. A maioria das aplicações tradicionais utilizam dispositivos de interface artificial, no qual o conhecimento é adquirido por meio de interações com botões que representam ações, tais como teclado ou controle de vídeo-*game*, mouse etc. cuja a curva de aprendizagem pode ser alta.

Segundo Wigdor e Wixon (2011), as interfaces gráficas tradicionais criam obstáculos aos seus usuários tendo apenas a assistência a visão tradicional da informação e por possuírem dispositivos usuais de entrada como os citados anteriormente, em oposição a NUI recorre de recursos como a interação gestual. Os *games* do dispositivo Microsoft Kinect são exemplos que se utiliza a NUI para gerar uma experiência mais imersiva e interativa que um *game* tradicional que utiliza um controle remoto, justamente por incluir o jogador a um ambiente de interação natural, partindo do preceito que acionar botões em um controle remoto não é uma ação natural humana.

As interfaces naturais devem estabelecer características que as diferencie comparadas as interfaces tradicionais, tornando-as mais intuitivas e naturais como proposto por Wigdor e Wixon (2011), Liu (2010), Saffer (2008) e, Norman e Nielsen (2010). Dentre todas as características propostas dos autores citados descrevemos o estudo (LIU, 2010) que orienta o desenvolvimento de interfaces no intuito de torná-las mais próximas da naturalidade e do usuário:

Design centrado no usuário: provê mudança da interface de usuário de forma externa e interna para atender às necessidades de diferentes usuários. Em sistemas de interfaces tradicionais as pessoas são consideradas como operadoras e assim devem se adaptar à máquina, onde as conhecemos com o termo de usuário, tais interfaces permitem apenas o diálogo, mas nenhum controle ativo. Em sistemas de NUI as pessoas são participantes ativos, assim a máquina irá responder as contínuas ações humanas para cada pessoa que utiliza o sistema;

Multi-canal: visa fazer o uso pleno de um ou mais canais sensoriais e motores humanos para capturar as características complementares sobre a verdadeira intenção de interação do usuário com a máquina, assim aumentar a naturalidade da interação humano-computador. Algumas das modalidades sensoriais são: visão, audição, tato, olfato e equilíbrio; e no canal motor humano tem as mãos, boca, olhos, cabeça, pés e corpo e entre outros. O

uso de multi-canal vai de encontro ao caminho natural de interação, pode-se atingir uma comunicação eficiente humano-máquina, bem como a máquina ou humano escolher o melhor canal de resposta;

Inexato: as ações e pensamentos dos usuários não são precisos, de modo que a interface deve compreender a requisição das pessoas, diferente do usuário de interfaces tradicionais que usa o mouse e teclado com uma finalidade de interação precisa;

Alta largura de banda: a característica de saída das interfaces tradicionais dos computadores é a exibição rápida e contínua de imagens, cores e informações, mas é observado que a quantidade de entrada ainda é baixa, pois a interação é gerada por teclado e mouse, comparado a uma interface de interação natural, a entrada do usuário é gerada pela compreensão de voz, imagem e gestos. Uma NUI deve suportar uma alta largura de banda de entrada e saída para compreender sem erros a interação humano-computador gerada;

Interação baseada por voz: considerada o meio mais conveniente, eficiente e natural de compartilhamento de informações. Em interfaces NUI possui duas abordagens: reconhecimento vocal e tecnologia de compreensão;

Interação baseada por imagens: os humanos utilizam como principal sentido a visão. Para interfaces NUI, as imagens podem ser abordadas em 3 diferentes níveis: processamento de imagens, reconhecimento de imagens e percepção de imagens;

Interação baseada no comportamento: responsável por reconhecer a linguagem corporal do usuário, isto é, movimentos que expressem algum significado; e

Dispositivos comuns para NUI são:

Câmeras de cores aditivas em que o Vermelho (*Red*), o Verde (*Green*) e o Azul (*Blue*) (RGB) e RGB-D onde D vem de profundidade (*Depth*) para compreensão do ambiente e movimento humano.

Microfones para captura da voz e dos sons do ambiente.

2.3 Multimodalidade

A comunicação dos humanos dá-se por meio da percepção dos sentidos humanos, os sentidos humanos podem ser utilizados de maneira isolada ou combinada. Os sentidos constituem em uma rede sensorial que permite obter todo o tipo de informação necessária para interação, ou seja, a troca de informação, seja ela por gestos ou fala, este tipo de comunicação é naturalmente

multimodal. Tal termo "multimodal" tem sido cada vez mais utilizado em vários contextos e através de várias disciplinas como demonstra Bernsen (1997).

Um sistema de *Interaction Multimodal Human-Computer* (IHCM) é aquele que responde às entradas de múltiplas modalidades ou canais de comunicação por exemplo: fala, gestos, escrita e outros. Segundo Jaimes e Sebe (2007), que demonstram uma abordagem centrada no ser humano, as modalidades de muitos dispositivos computacionais de entrada podem ser consideradas como correspondências diretas aos sentidos humanos como: câmeras (visão), superfícies sensíveis ao toque (tato), microfones (audição) e sensores de odor (olfato).

Como exemplo pode-se definir um sistema que responde a expressões faciais e gestos com as mãos, utilizando somente câmeras como dispositivo de entrada. Este sistema não é considerado como multimodal, mesmo se os estímulos de entrada forem capturados através de várias câmeras. Mas um sistema com mouse e teclado como entrada já possui uma multimodalidade. Um sistema de IHCM permite ao usuário escolher a modalidade com a qual ele queira interagir, e que ainda seja possível para este usuário realizar a mesma tarefa de acordo com qualquer uma das modalidades de interação sugeridas e/ou suportadas pelo sistema.

Neste sentido, este trabalho tem como foco apenas na combinação visual (câmera) para entrada de gestos do corpo com outros tipos de entrada como teclado, mouse e outros para a interação humano-computador.

2.4 Taxonomia dos Gestos

A investigação de maior importância para este trabalho é sobre os gestos, antes de modelar um gesto para o computador com a finalidade de realizar a tarefa de reconhecimento, é necessário saber o máximo possível sobre as suas características. Como eles são definidos, as suas principais características, as notações usadas, que características são importantes para extrair e como extraí-las.

A análise dos gestos humanos com foco na compreensão conceitual do gesto veem sendo foco de extensas pesquisas em diversas áreas do conhecimento. Dentre as mais específicas no domínio podemos citar: linguística, antropologia, biologia, ciência cognitiva, psicologia, neurologia, teatro, artes visuais, dança, fisioterapia, etc.

Em relação a cada uma das áreas se encontram diferentes definições e convenções sobre gestos. No que diz respeito às referências deste trabalho, investigou o trabalho dos principais autores nos seus respectivos domínios de investigação. Estes incluem: Kendon (KENDON,

1994); McNeill & Levy (MCNEILL; LEVY, 1980); Rector & Trinta (RECTOR; TRINTA, 1999); Corradini (CORRADINI, 2001); entre outros trabalhos de grande relevância para este estudo sobre gestos.

Embora as abordagens destes estudos não visam justificar uma teoria de análise de gestos através do desenvolvimento de um modelo computacional, eles fornecem uma análise profunda qualitativa e teórica sobre gesto humano para uma melhor interpretação da captura do gesto.

Esta seção resume várias definições de gesto e algumas abordagens para a classificação de um gesto, e define o gesto espacial tridimensional e suas características importantes.

2.4.1 Definição de Gestos

Há muitas definições de gesto, McNeill (1992) foca a definição de gesto como uma relação entre um discurso falado e movimentos dos braços e mãos. Ademais, podemos encontrar uma definição mais ampla no dicionário Aurélio:

> Movimento do corpo, principalmente das mãos, dos braços e da cabeça; Mímica, aceno, sinal: com um simples gesto, expressou o pensamento; Aspecto, aparência; Semblante. (FERREIRA; FERREIRA; ANJOS, 2009)

Neste sentido, estas definições podem ser úteis para um propósito geral, mas é necessária uma definição mais clara sobre o gesto e, de como realmente aplicar esse termo para implementação computacional na IHC.

Rector e Trinta (1999) em seu estudo representam gesto como uma ação corporal visível, em que um certo significado é transmitido por meio de uma expressão voluntária. Dessa forma, gesto é um recurso esclarecedor de mensagens, tanto em processos de comunicação verbal em que reforça a linguagem falada, quanto em processos de comunicação não verbal.

Hickson e Stacks (1985) definem gestos como um processo em que as pessoas manipulam intencionalmente ou não suas ações e expectativas exprimindo experiências, atitudes e sentimentos, para si mesmo, o ambiente e os outros.

Na mesma acepção, Bobick (1997) descreve o gesto em três aspectos relevantes: o movimentos, atividade e ação:

- Os movimentos representam a forma mais primitiva de ação que pode ser interpretada semanticamente;
- A atividade é uma sequência de movimentos ou posições estáticas; e

- As ações são as entidades de alto nível no qual as pessoas normalmente usam para descrever o que está acontecendo.

Em contraposto Luciani compara três termos relevantes: gesto, movimento e ação. Segundo Luciani et al. (2010), o movimento e gesto são termos semelhantes, e, na verdade, o movimento se refere uma evolução produzida por um sistema físico, qualquer que seja, corpo humanos, objetos mecânicos etc. Como por exemplo, o movimento de um corpo humano, de uma folha e de uma fonte de som. O movimento é considerado como o resultado da evolução executada, isto é, uma saída de um sistema em evolução. A ação é um resultado de alto nível por exemplo, "vou pegar a bola" , que é causada por um conjunto de gestos de entrada. Cada ação pode ser executada por muitos movimentos de recursos diferentes e podem ser descritos em vários níveis simbólicos.

O aspecto temporal também é destacado nas definições de gestos, os gestos são definidos como movimentos ou sequências. Corradini (2002), Kim, Song e Kim (2007) caracterizam os gestos como sequências de posturas conectadas por movimentos durante um período de tempo. À vista disso, podemos considerar os aspectos temporais dos gestos para definir e modelar um gesto.

2.4.2 Classificação de Gestos

Quek forneceu um estudo (QUEK, 1994) na área de IHC sobre a taxonomia de um gesto com foco em movimentos das mãos. Ele faz a classificação em duas classes: gestos e movimentos involuntários. Se os movimentos das mãos não transmitem qualquer informação significativa, eles são considerados como os movimentos involuntários. Enquanto os gestos são movimentos que tem algum significado e se dividem em subclasses: manipulativos e comunicativos.

Gestos de manipulação são usados para movimentar e/ou girar objetos. Gestos comunicativos são usados para um propósito comunicacional inerente. Esses tipos de gestos podem ser usados tanto para atos quanto para a interpretação do movimento e ou símbolos para um sentido linguístico.

Para interpretar os movimentos do corpo, a pessoa tem que classificá-los de acordo com as propriedades comuns dos movimentos que podem expressar de fato um gesto. Por exemplo, em linguagem de sinais cada gesto representa uma palavra ou frase. Kendon (1980) analisou a relação entre o gesto e a fala. Ele classifica um gesto como um rótulo para ações que tenham características de manifestação de expressão.

Feldman e Rimé (1991) propuseram a taxonomia gestual consistindo como: simbólico,

dietéticos, icônico e pantomímico: Gestos simbólicos, tem um único significado dentro de cada cultura, como um emblema; Gestos dietéticos, são usados para apontar ou direcionar a atenção do ouvinte para um evento específico ou de um determinado objeto de destino; Gestos icônicos, são para transmitir informações sobre o tamanho, forma ou orientação do objeto no discurso, outro exemplo, de um gesto icônico é visualizar um caminho de voo de um avião, ao mover sua mão no ar; e Gestos pantomímico, são normalmente utilizados em mostrar os movimentos.

A taxonomia de gestos, influência em grande parte a forma como pode ser modelado um gesto. Como é observado nos trabalhos dos autores citados anteriormente, dentro deste conjunto de informações consideramos nesse trabalho os parâmetros de espaço e intervalo de tempo em que é executado um gesto, que são substanciais para modelagem de gesto.

Alguns gestos são naturais, que com um pouco de aprendizagem conseguimos reproduzir. Em contrapartida alguns gestos são concebidos artificialmente, por uma pessoa ou um grupo com intuito especial, um exemplo é a comunicação visual de esportes e operações militares. Gestos artificiais podem requerer um nível de percepção elevado para aprendizagem e interpretação.

Através dos estudos citados anteriormente podemos definir que um gesto é oriundo e preconcebido da concepção mental do usuário, tal gesto após a concepção mental pode se utilizar contexto como pessoas, objetos e ambiente para reproduzi-lo.

2.4.3 Descrição de Gestos Espaciais

Nesta subseção, a visão geral da descrição de gestos espaciais. Em especial é dado maior importância a gestos baseados em representações espaciais tridimensionais.

A interação espacial é um elemento substancial em NUI, no qual um movimento pode representar um gesto e consequentemente um evento de entrada para um sistema. Pressupondo que os gestos são baseados no corpo humano e expressados através do movimento das diferentes partes do corpo humano, utilizando-se de sensores os gestos do corpo podem ser descritos e caracterizados através de suas propriedades espaciais como dois tipos de sinais: contínuo e discreto.

Outra característica observada pelo estudo da taxonomia dos gestos, é que uma interação espacial, no processo de comunicação gestual, envolve simultaneamente o ambiente de interação e o próprio gesto, através da manipulação de objetos físicos, pessoas e uma gama complexa de eventos proveniente do ambiente de interação, que definimos neste trabalho como contexto, e por consequência, a análise do gesto pode se tornar mais complexa e difícil.

Neste sentido, uma captura da interação espacial pode ser baseada em sensores e técnicas diferentes como demonstra Pavlovic et al. (1997). No entanto, sensores apropriados devem ser selecionados para atender as quantidades contínuas previstas e o grau de liberdade no movimento de um gesto esperado. Assim, no desenvolvimento de um sistema de reconhecimento de gestos, é importante modelar o movimento (características temporais) e forma (características espaciais) do corpo.

Em particular, uma maior importância é colocada em gestos com representações espaciais, baseado no modelo tridimensional que depende do tipo de dados de entrada do sensor a ser utilizado. Existem dois tipos principais de descrição de gestos espaciais tridimensionais, definidos na literatura, a Figura 2.2 apresenta os dois tipos de modelos de descrição de gestos espacial : os baseados em reconstrução tridimensional que possuem mapas de profundidade, e os baseados em aparência que possuem apenas *pixels* de imagens em duas dimensões, que são discutidos mais à frente.

Fonte: Adaptada de (PAVLOVIC et al., 1997)

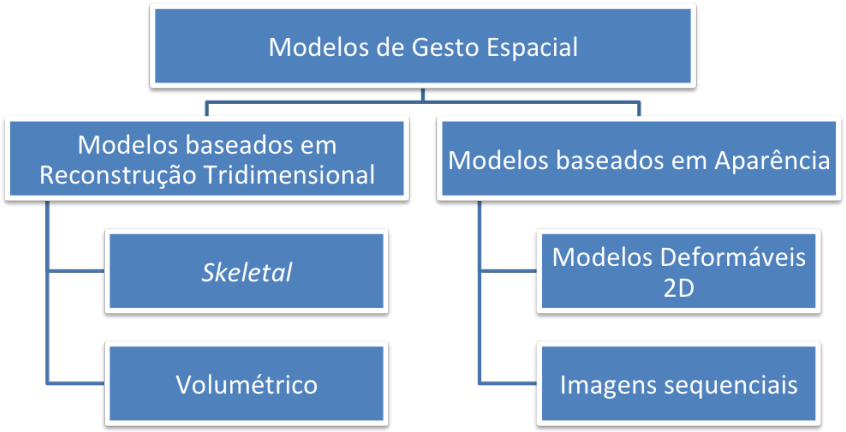


Figura 2.2: Diagrama representando diferentes tipos de modelos para descrição do gesto espacial

2.4.3.1 Modelos baseados em Aparência

Os modelos baseados em aparência através de imagens sequenciais ou de imagens deformáveis de duas dimensões, sua interpretação é diretamente ligada aos *pixels* que compõe uma imagem de duas dimensões, um modelo baseado em aparência é apropriado para gestos de duas dimensões, por exemplo, ao pressionar uma tela e/ou tecla de pressão sensível, pode ser considerado um gesto de uma dimensão (1D) e movendo as mãos sobre a superfície plana pode

ser considerado um gesto de duas dimensões (2D) muito comum em *touch-pads, touch-screens* conhecido como gesto *swipe*. Os modelos baseados em aparência mais utilizados são:

- Modelo baseado em cor : em geral, usando marcadores do corpo para controlar o movimento da parte do corpo;
- Modelo deformáveis binário : são geralmente baseados em contornos ativos deformáveis;
- Modelo baseado em movimento : com base no movimento de *pixels* individuais ou descrição da imagem; e
- Modelo baseado em silhueta : modelos baseados nas propriedades geométricas da silhueta.

Mas tal limitação dimensional que os modelos baseados em aparência possuem, torna mais complexa a tarefa de reconhecimento de um gesto do corpo, foco deste trabalho, o gesto do corpo humano em sua naturalidade é realizado no espaço com propriedades tridimensionais, pois estão intimamente ligados à habilidade motora que é a base para os movimentos humanos. Nossos movimentos cotidianos no mundo real estão em cenários em que por exemplo manipulamos objetos físicos, tais como dirigir automóveis, jogar futebol, cozinhar etc. Assim a próxima subseção discute com mais profundidade os Modelos baseados em Reconstrução Tridimensional.

2.4.3.2 Modelos baseados em Reconstrução Tridimensional

Os modelos baseados em Reconstrução Tridimensional, são originados através de mapas de profundidade de um sensor. Os mapas de profundidade compõem-se de imagens que contém valores de profundidades associados para cada pixel, um mapa de profundidade pode ser interpretado como uma imagem bidimensional que possui informações de profundidade, podendo ainda conter informações de cor, curvatura dos pontos, etc. capturados da cena visualizada pelo sensor.

Os sensores que capturam mapas de profundidade podem ser classificados em ativos e passivos: Os ativos fornecem e gerenciam a sua própria iluminação projetando um padrão simples ou codificado para extração de profundidade da cena; Os passivos apenas absorvem as radiações do ambiente, e a partir delas buscam extrair informações de profundidade de um ponto. (YOUNG, 1994).

Para sensores ativos, os métodos mais conhecidos são: triangulação, *Time-of-Flight* (ToF) e luz estruturada.

A triangulação, é obtida através das propriedades geométricas do triângulo para calcular a localização de um centro de interesse ou ponto, sendo considerado o método mais antigo para medição de intervalo entre pontos remotos (BESL, 1988; POSDAMER; ALTSCHULER, 1982). O mapa de profundidade é obtido a partir da triangulação feita entre uma lente de uma câmera receptora e um emissor de feixe de laser.

Como demonstrado na Figura 2.3, (d) indica a distância horizontal entre o feixe emissor de laser e o eixo óptico da lente da câmera, e (α) é o ângulo entre o feixe do laser e a linha horizontal, portanto o ângulo de canto da lente da câmera (β) pode ser determinado observando a posição do ponto do feixe de laser no eixo óptico da lente da câmera. Ambos são fixos e conhecidos quando o ponteiro laser está instalado. O ponto *P* é o ponto de intersecção do feixe de laser e a superfície do objeto.

Fonte: Elaborada pelo Autor

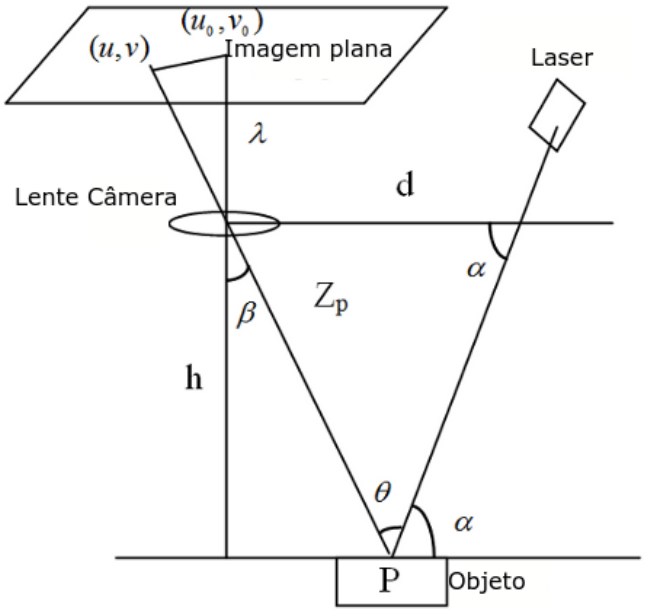


Figura 2.3: Cálculo da Profundidade de um ponto usando Triangulação

A função de profundidade (Zp), se obtém os índices de pixel (u, v), tal função é derivada através da aplicação da trigonometria (TRUCCO; VERRI, 1998). O (Zp) profundidade calculada pela Equação 2.1 pode ser usada para aproximar a profundidade de cada ponto na característica de superfície planar do objeto alvo.

$$Zp = \frac{d \sin a}{cos(\alpha - \beta)} \tag{2.1}$$

O método de luz estruturada por triangulação, demonstrado na Figura 2.4, tem a triangulação baseada na deformação geométrica de uma única faixa de luz emitida pelo laser que é projetada sobre uma superfície tridimensional. O deslocamento das linhas, ou seja, a interseção entre a visão da câmera e luz produzida pelo feixe do sensor permite uma recuperação exata de coordenadas tridimensionais sobre a superfície do objeto.

Fonte: Adaptada de (HECHT, 2011)

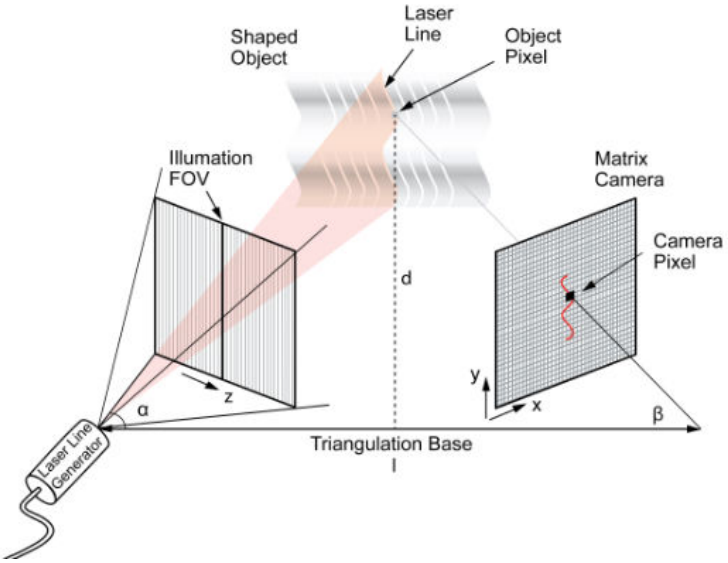


Figura 2.4: Triangulação por Luz Estruturada

No método de ToF, a profundidade da cena é obtida através do tempo de duração entre a emissão da luz do emissor e o retorno dela ao sensor. O sistema de medição da distância tipicamente é composto por dois componentes principais: um emissor de luz e um detector de luz (câmera), como mostrado na Figura 2.5. Medindo-se o tempo de vôo de um pulso luminoso entre a câmera e o objeto se obtém a profundidade para cada pixel da imagem.

Fonte: Adaptada de (GOKTURK; YALCIN; BAMJI, 2004)

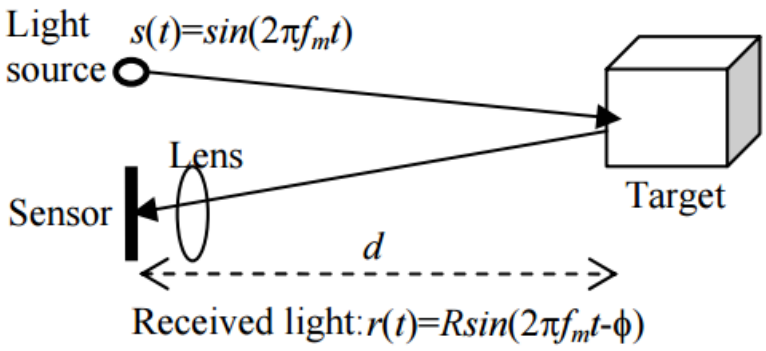


Figura 2.5: Sistema de Medição ToF

A função $s(t) = \sin(2\pi f_m t)$ representa a luz transmitida onde $f_m t$ representa a frequência de modulação. A luz refletida do alvo é registrada em um pixel do sensor através de um deslocamento de fase ϕ :

$$r(t) = R\sin(2\pi f_m t\phi) = R\sin(2\pi f_m(t - \frac{2d}{c})) \tag{2.2}$$

R é a amplitude da luz refletida, d é a distância entre o sensor e o alvo, e c é a velocidade da luz, $(\approx 3 \times 10^8 m/s)$. A distância Equação 2.3 d pode ser calculada a partir da mudança de fase como se segue:

$$d = \frac{c\phi}{4\pi f_m} \tag{2.3}$$

Segundo Weise et al. (2011), sensores que aplicam o método de luz estruturada comparado ao ToF possuem vantagem de prover mapas de profundidade densos com alta precisão.

Assim, os modelos baseados em Reconstrução tridimensional, definem a descrição espacial tridimensional das partes do corpo humano, que podem ser classificados em dois grandes grupos: Modelos volumétricos; e Modelos esqueletais (*Skeletal*).

2.4.3.3 Modelos Volumétricos

As abordagens baseadas em modelos volumétricos para os gestos do corpo possuem a finalidade de descrever a aparência dos membros do corpo, regularmente tal abordagem é utilizada em animação de computador, mas tem sido utilizado em aplicações de visão computacional

(PAVLOVIC et al., 1997).

A superfície de modelos volumétricos pode ser representada com diversas técnicas como a *Non Uniform Rational Basis Spline* (NURBS) (PIEGL; TILLER, 2012) muito popular em programas gráficos. Tal técnica é complexamente custosa para processamento de reconhecimento de gestos em tempo real, alternativas a esta técnica é a utilização de formas geométricas, tais como cilindros, esferas e elipsoide.

2.4.3.4 Modelos baseados em Algoritmos *Skeletal*

Outro modelo de gesto espacial são os baseados em algoritmos *Skeletal* , que capturam as articulações do esqueleto humano através da reconstrução tridimensional e sua descrição espacial é representada tridimensionalmente. Um exemplo é o algoritmo *Skeletal* de Shotton et al. (2013) que foi aplicado para o Microsoft Kinect, gera parâmetros importantes como ângulos e posições tridimensionais de articulações do corpo em tempo real, que são rastreados pelo mapa de profundidade produzido pela reconstrução tridimensional das câmeras do dispositivo, tal modelo faz a estimação de pose através de uma abordagem de classificação por *pixel* como demonstra a Figura 2.6.

Fonte: (SHOTTON et al., 2013)

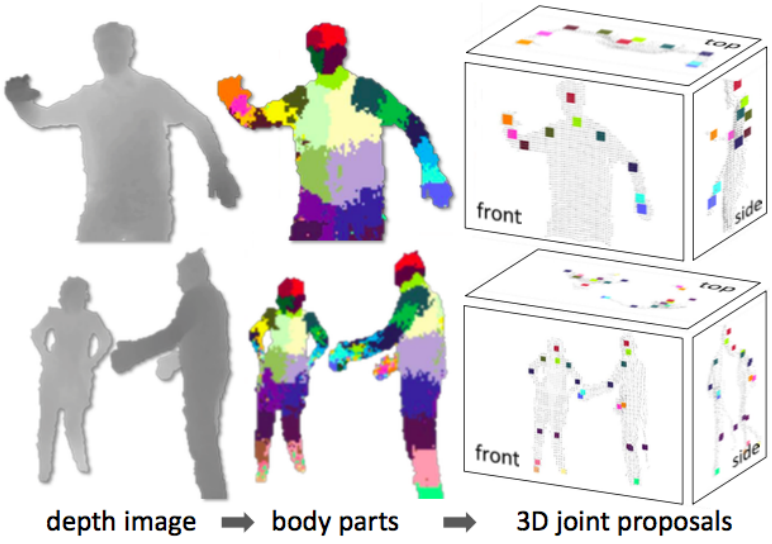


Figura 2.6: Segmentação de articulações do corpo.

A descrição de gestos espaciais através de modelos volumétricos baseados na reconstrução tridimensional de sensores, podem transmitir as informações necessárias para uma análise

elaborada da interação gestual simultaneamente com o seu contexto, pela sua quantidade de dados, porém a aplicação de uma estrutura de dados para análise dessas informações pode requerer muito processamento computacional, a fim de ser implementado para a análise em tempo real. No entanto, a tarefa de reconhecimento de gestos espaciais baseado em algoritmos *Skeletal*, gera uma limitação para gestos que aplicam o contexto pois somente é capturado a interação gerada pelo movimento do corpo, em contraponto, traz grandes benefícios para IHC em aplicação a uma NUI (LIEBLING; MORRIS, 2012; WINKLER et al., 2012; FARHADI-NIAKI; GHASEMAGHAEI; ARYA, 2012; GNECCO et al., 2012; BRASIL et al., 2014) e por ser um conjunto menor de dados, o tempo de processamento pode ser reduzido.

2.4.4 Captura do Gesto

Do ponto de vista da engenharia, cada quantidade física pode ser medida por meio de diferentes tipos de sensores. Para usar a interação espacial como uma entrada, os movimentos realizados devem ser quantificados através de sensores. A determinação de escolha de um sensor é geralmente baseada nas precisões, taxas de captura, finalidade da aplicação, infraestrutura e custo.

Existem quatro tipos principais de sensores de captura de movimento, são eles: magnético, eletromagnético, mecânico, óptico e inercial (FURNISS, 1999).

Neste trabalho apresentamos alguns dos sensores de captura do movimento do corpo humano no espaço tridimensional (3D) que são usualmente dedicados para interface humano-computador e que definimos em dois tipos: (I) Sensores corporais; e (II) Sensores visuais.

Sensores corporais, são geralmente vestidos ao corpo e a captura é baseada nas características magnéticas, mecânicas e inerciais, tais sensores oferecem precisão, e relativamente uma amplitude para o movimento do corpo humano, e as taxas de atualização são geralmente rápidas (TECHNOLOGIES, 2014; FELS; HINTON, 1993; BAUDEL; BEAUDOUIN-LAFON, 1993; MOESLUND; GRANUM, 2001). Algumas roupas de captura de movimento possuem rastreadores ópticos (pontos ou pequenas bolas que servem como marcadores) ou eletromecânicos (CORPORATION, 2014; TECHNOLOGY, 2014), contudo necessitam de câmeras para o rastreamento do movimento. As abordagens baseadas em sensores corporais exigem que sejam vestidos e geralmente carregam uma carga de cabos que conectam o dispositivo a um computador, assim dificultando o movimento e a livre circulação do usuário ao ambiente e não indo ao encontro da NUI, foco deste trabalho. Um dos sensores mais populares para captura do movimento e de baixo custo é o acelerômetro, que mede tanto a aceleração ou ângulos de rotação ao longo de um determinado eixo, e que também é utilizado para captura de gestos. (DONG; WU; CHEN, 2007; NGUYEN et al.,

2011).

Sensores visuais, como câmeras, capturam as formas e propriedades, tais como textura, profundidade e cor do mundo real. Algumas comunicações gestuais também se utilizam da face, como em língua de sinais. Neste cenário sensores visuais são utilizados para capturar movimentos faciais e labiais (NEFIAN et al., 2002). As câmeras também são usadas para capturar movimentos executados em *desktop*. Algumas abordagens utilizam câmeras que são colocadas atrás de telas transparentes ou semitransparentes, elas monitoram a posição e movimento das mãos em cima da tela (BAE et al., 2004).

Para captura de atividade do movimento do corpo, alguns sistemas foram desenvolvidos para análise de posturas corporais, acompanhando os movimentos do corpo inteiro com câmeras (AGGARWAL; PARK, 2004). Há várias abordagens para a construção de um sistema de captura de movimentos que personaliza diferentes tipos de sensores corporais e visuais para o rastreamento do movimento (STIEFMEIER et al., 2006; NEWMAN et al., 2004; PUSTKA; KLINKER, 2008).

Ambas as categorias têm suas vantagens e desvantagens, por exemplo, sensores corporais requerem a cooperação do usuário para seu uso, e pode ser desconfortável para usar por muito tempo, mas podem ser mais precisos. Sensores visuais não requerem a cooperação do usuário, mas eles são mais difíceis de configurar e sofrem de problemas de oclusão. Os sensores corporais também sofrem de oclusão quando são obstáculos de metal ou mecânicos, outro aspecto é de saúde, alguns sensores corporais podem trazer problemas de alergia ou exposição aos dispositivos magnéticos

No entanto, sensores visuais para captura de movimento espaciais tridimensionais são necessários um conjunto de algoritmos e modelos demonstrados na Subseção (2.4.3), que são baseados na extração das propriedades ópticas, que demandam um maior poder computacional e tempo, outro aspecto de desvantagem, é que a interação se limita a visão da câmera. Porém as abordagens de sensores visuais, trazem maior naturalidade, pelo fato de serem sensores visuais, não interferindo na interação espacial, indo ao encontro da NUI, foco deste trabalho.

2.4.5 Sensores Visuais de Interação Natural

Nesta subseção, é apresentado os principais sensores visuais para Interação Natural. Tais sensores são baseados em Modelos de Reconstrução Tridimensional, neste sentido foram levantadas as principais características de cada sensor, bem como suas vantagens e desvantagens para empregar neste trabalho.

2.4.5.1 Microsoft Kinect

Microsoft lançou em 2010, um conjunto de sensores em um único dispositivo revolucionando a indústria de jogos. Inicialmente o Microsoft Kinect XBOX 360^{TM}foi desenvolvido como um dispositivo periférico para uso com o console Xbox 360^{TM}(GNECCO et al., 2012). Seus três principais sensores, ou seja, RGB, áudio e profundidade (projetor de feixe de laser infravermelho e um sensor óptico).

O dispositivo juntamente com bibliotecas de desenvolvimento permite detectar movimentos, identificar rostos e fala. Assim, cria-se uma interface de controle de jogos, onde o controle é o próprio corpo do jogador. Embora seu objetivo inicial era para jogos, o Microsoft Kinect abriu a porta para um grande número de aplicações úteis.

Fonte: Elaborada pelo Autor

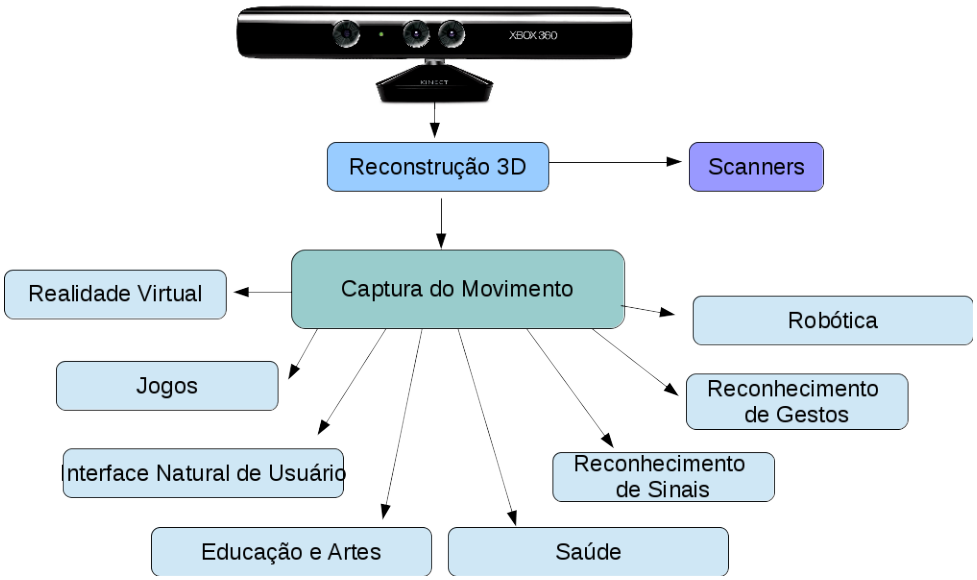


Figura 2.7: Categorias de Aplicações do Kinect

A Figura 2.7 mostra as principais categorias de aplicações do dispositivo, abrangendo desde a saúde, à educação, jogo, robótica, interface natural do usuário, reconhecimento de sinais, bem como a reconstrução 3D, que é um grande passo para a revolução da impressão 3D. Exceto para aplicações de escaneamento, todas as outras aplicações exigem a captura de movimento que o dispositivo provê.

O sensor de profundidade do Microsoft Kinect é composto por um projetor infravermelho e um sensor *Complementary Metal-Oxide-Semiconductor* (CMOS) monocromático. A câmera RGB refere-se as cores que ela detecta, ou seja, vermelho, verde e azul. A captura de som é

feita por uma matriz de quatro microfones que possuem isolamento de ruídos do ambiente e detecção de distância. A Figura 2.8 apresenta o dispositivo e seus componentes.

Fonte: Elaborada pelo Autor

Figura 2.8: Dispositivo Microsoft Kinect

O sensor de profundidade, utiliza-se do método conhecido de triangulação de luz-estruturada. Trata-se de um projetor de laser infravermelho combinado com um sensor CMOS monocromático. O projetor gera uma grade/padrão de pontos de luz infravermelha no campo de visão. Logo após um mapa de profundidade é criado com base nos raios que o sensor recebe a partir de reflexões da luz de objetos na cena.

O dispositivo é capaz de capturar dados de profundidade em tempo real a partir do chip SoC PS1080 (Primesense, 2012). Esta tecnologia foi desenvolvida pela empresa israelense PrimeSense, que é capaz de interpretar a informação de cena 3D a partir de uma luz infravermelha continuamente estruturada-projetada.

O Microsoft Kinect não é o único dispositivo que usa o design de arquitetura de referência PrimeSense, como demonstra a Figura 2.9, a arquitetura interliga o chip a sensores de profundidade, áudio e cor, a memória flash e a uma interface *Universal Serial Bus* (USB) . Outros fabricantes também têm aplicado o chip aos seus dispositivos, como o ASUS Xtion Pro LIVE.

Fonte: (DIAS et al., 2013b)

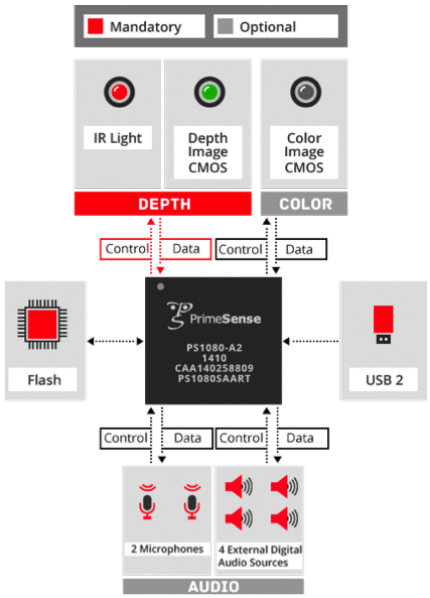


Figura 2.9: Design PrimeSense Chip SoC PS1080

Uma versão revisada chamada Kinect para Windows, foi lançada para desenvolvimento de aplicativos em computadores em fevereiro de 2012. O Kinect para Windows oferece um modo de rastreamento do movimento para perto, chamado de *near mode*, mas todas as outras especificações de hardware continuam a ser o mesmo que o sensor Kinect original.

A segunda versão do Kinect (referido como Kinect v2 neste trabalho) foi lançada em setembro de 2014. Na segunda versão, a captura do mapa de profundidade é baseada no método ToF, e possui resolução mais alta na câmera de cor comparada a primeira versão. As principais características para as duas versões do dispositivo são apresentadas na Tabela 2.1.

2.4.5.2 ASUS Xtion PRO Live

O sensor ASUS Xtion PRO Live [1] é desenvolvido pela fabricante de hardware ASUS, que trouxe como proposta o desenvolvimento de soluções para NUI em computadores. O dispositivo é composto por um sensor de profundidade e uma câmera RGB. Suas principais características são descritas na Tabela 2.2 .

Comparado ao Kinect ele possui as mesmas referências de design da PrimeSense, ou seja, traz seus componentes básicos para o chip SoC PS1080, mas em relação ao Microsoft Kinect o dispositivo pode ser alimentado diretamente pela interface USB, e suas dimensões de tamanho

[1] Informações: *https://www.asus.com/us/3D-Sensor/Xtion_PRO_LIVE/*

Tabela 2.1: Comparação das principais características das duas versões do sensor Kinect

Características	Kinect v1	Kinect v2
Distância de uso	0.8m - 4m 0.4m - 3m (near mode)	0.5m - 4.5m
Mapa de Profundidade	Triangulação com Luz Estruturada	TOF
Campo de visão (FoV)	Horizontal: 57graus. Vertical: 43 graus.	Horizontal: 70graus. Vertical: 43 graus.
Motor de Inclinação Vertical	$\pm$ 27 graus	
Resolução da Imagem de RGB	640x480 30fps 1920x1080 30fps 1280x960 12fps	1920x1080 30fps 1280x960 12fps
Resolução da Imagem de Profundidade	640 x 480 30fps 640 x 480 15fps 1280 x 960 12fps	512x424 30fps
Interface	USB 2	USB 3
Dimensões	7.3 cm x 28.3cm x 7.28 cm	24.9cm x 6.6cm x 6.7cm

Tabela 2.2: Características do ASUS Xtion Live Pro

Características	ASUS Xtion PRO Live
Distância de uso	0.8m – 3.5m
Mapa de Profundidade	Triangulação com Luz Estruturada
Campo de visão (FoV)	Horizontal: 58 graus. Vertical: 45 graus.
Motor de Inclinação	-
Resolução da Imagem de RGB	1280x1024 15fps 640x480 30fps
Resolução da Imagem de Profundidade	640x480 30fps 320x240 60fps
Interface	USB 2.0/3.0
Dimensões	18 cm x 3.5 cm x 5 cm

são reduzidas. Mas não possui o motor de inclinação. O dispositivo é pouco documentado e a sua venda comercial não é amplamente globalizada que acaba gerando uma dificuldade e prejudicando o desenvolvimento de aplicações para o dispositivo.

2.4.5.3 Carmine's

Os sensores Carmine's desenvolvido pela PrimeSense, possuem semelhanças ao dispositivo ASUS Xtion Pro Live. Sua principal diferença está em seu tamanho que é reduzido comparado aos concorrentes.

Os sensores Carmine's são dispositivos de interação natural desenvolvido pela PrimeSense, segue a referência do chip SoC PS1080 com suporte à captura de áudio, cor e profundidade. Sua aparência é bastante semelhante ao ASUS Xtion PRO Live e também utiliza alimentação e interface via USB. As diferenças entre os sensores Carmine 1.08 e Carmine 1.09, é de que

o 1.09 possui um alcance menor na captura de profundidade. As principais características dos dispositivos Carmines são descritas na Tabela 2.3.

Tabela 2.3: Comparação das principais características das versões dos sensores Carmine's e Capri

Características	Carmine 1.08/1.09	Capri 1.25
Distância de uso	0.8m – 3.5m / 0.35m – 1.4m	0.8m-3.5m
Mapa de Profundidade	Triangulação com Luz Estruturada	Triangulação com Luz Estruturada
Campo de visão (FoV)	Horizontal: 57.5 / 56.5 graus. Vertical: 45 graus.	Horizontal: 56.5 graus. Vertical: 45 graus.
Motor de Inclinação	-	-
Resolução da Imagem de RGB	640x480 30fps	-
Resolução da Imagem de Profundidade	640x480 30fps	640x480 30fps
Interface	USB 2.0 e 3.0	USB 2.0 e 3.0
Dimensões	18 cm x 2.5 cm x 3.5 cm	Sem Informações

A PrimeSense produz ainda uma solução embarcada denominada Capri 1.25, que fornece uma plataforma de sensoriamento tridimensional com características reduzidas, para soluções embarcadas como PCs, *All–in-One* PCs, tablets, laptops, smartphones, TVs e Robôs. A Figura 2.10 mostra o sensor Capri 1.25 acima do Carmine 1.09 .

Fonte: (GUIZZO, 2013)

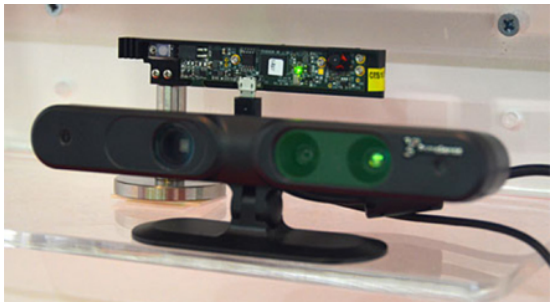


Figura 2.10: Sensor Capri 1.25 e Carmine 1.09

A produção dos sensores Carmine's e Capri foram descontinuados após a aquisição da PrimeSense pela Apple, tornando-se sua utilização em projetos acadêmicos e comerciais pouco viável.

2.4.5.4 Structure Sensor

O sensor Structure Sensor [2] é uma solução para dispositivos móveis para utilização em aplicações como: Mapeamento de espaços internos, medições instantâneas e redecoração virtual em tempo real; Jogos de Realidade Aumentada; e Escaneamento de objetos. A Figura 2.11 apresenta o sensor acoplado a um iPad [3].

[2]Informações: *https://structure.io/*
[3]Informações: *https://www.apple.com/br/ipad/*

Fonte: Elaborada pelo Autor

Figura 2.11: Structure Sensor

Por ser um sensor que é projetado para dispositivos móveis diferentemente dos outros citados, as dimensões e características são reduzidas e compactadas, com otimização no alcance e fornecimento de energia.

2.4.5.5 DepthSense 325 (DS325)

O sensor DS325 é desenvolvido para Interação Natural fabricado pela empresa SoftKinect, sua característica é de ser um dispositivo de interação de curto alcance. Sua precisão de captura de profundidade é alta comparada aos outros sensores, e a sua distância de início de captura é bem maior, no qual contribui para aplicações de rastreamento de mãos e dedos. A empresa Intel criou um programa chamado de Intel® Perceptual Computing que renomeou e renovou o design do sensor Figura 2.12, agora chamado Intel® RealSense™ Camera (F200) [4], o sensor possui as mesmas características do DepthSense 325.

Fonte: Elaborada pelo Autor

Figura 2.12: Intel® RealSense™ Camera (F200)

[4]Informações: *https://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html*

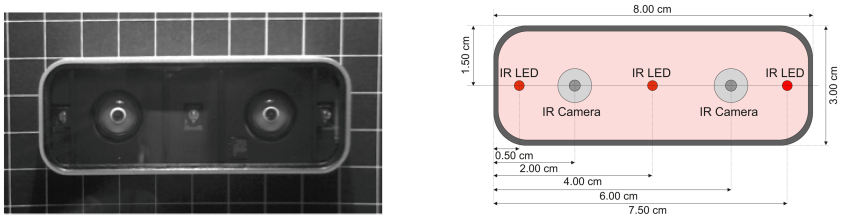
O sensor possui câmera RGB de alta definição, dois microfones e uma câmera que aplica o método ToF, tal método é chamado DepthSense. A sua interface USB também fornece alimentação de energia para o dispositivo, assim dispensa o uso de outras fontes de alimentação.

2.4.5.6 LEAP Motion

O sensor LEAP Motion [5], é desenvolvido e projetado para captura do movimento e detecção de gestos de mãos e dedos pois seu alcance estende-se desde aproximadamente 25 a 600 mm acima do sensor, e seu campo de visão a 150 graus (GUNA et al., 2014). Suas principais aplicações são em NUI utilizando gestos das mãos como entrada.

O método de captura de profundidade também é o de luz estruturada. O sensor possui alta taxa de quadros por segundo, cerca de 200 quadros por segundo, isso por conta de sua arquitetura e configuração dos sensores internos, o Leap Motion é equipado com três emissores de infravermelhos e duas câmeras. A Figura 2.13 apresenta o sensor e visualização e disposição dos sensores.

Fonte: (GUIZZO, 2013)



(**a**) (**b**)

Figura 2.13: (a) Visão do hardware real e (b) Visão da configuração dos sensores.

2.5 Modelagem do Gesto

As seções anteriores descreveram como os movimentos tridimensionais espaciais são adquiridos através de diferentes sensores e métodos. Nesta seção é apresentado como os dados adquiridos serão modelados para suportar a tarefa de reconhecimento de gesto do corpo.

Fornecemos em aspectos gerais a compreensão global dos elementos necessários para o processamento de gestos espaciais tridimensionais baseado nos conceitos apresentados sobre a Taxonomia dos Gestos no Capítulo (2.4).

[5] Informações: *https://www.leapmotion.com/*

2.5.1 Produção do Gesto

A Produção do Gesto, demonstra como um gesto é produzido a partir de um usuário e como é observado por um sistema. Gestos se originam de um conceito mental humano, e que também se situam com o contexto (ambiente de execução) e são expressados através de movimentos que são observados como sinais. Segundo Kwon (2008), durante a produção do gesto, podem existir duas fontes de erros que podem ser estimadas neste processo: o de empenho e medição.

O erro de empenho é gerado a partir de movimentos do usuário ao executar um gesto. Um exemplo é quando usuários executam um determinado gesto, o resultado real, muitas vezes difere de sua expectativa devido à imprecisão do controle de seu corpo e até mesmo incapacidade de gerar exatamente as posturas e gestos desejados. No erro de medição, é gerado a partir do sistema de sensor por causa de ruídos, distorções e variâncias causadas pela abordagem de captura do sensor.

2.5.2 Percepção do Gesto

Na produção do gesto discutido anteriormente, demonstra que um gesto produzido gera um conjunto de sinais a ser observado. A percepção de um gesto é o processo inverso, ou seja, a conversão de observações gestuais para um conceito mental. Segundo Kwon (2008), existem dois principais processos de percepção do gesto: análise e reconhecimento.

2.5.2.1 Análise

O objetivo da análise, é estimar os parâmetros de um gesto e um contexto, usando medidas da observação do gesto exercido. No entanto, considerando o esforço computacional o mapeamento direto de observações de um gesto e contexto, ao mesmo tempo como parâmetro, seria extremamente complexo, devido ao grande volume de características a serem analisadas.

Desta forma, podemos concluir que o conveniente, seja que as características de um gesto e características de contexto sejam extraídas a partir de observações, mas os parâmetros de gesto e parâmetros de contexto, sejam estimados a partir das características extraídas individualmente. A análise pode compor-se de duas funções:

- *Detecção de características relevantes:* uma vez definida as características, é necessário selecionar os sensores apropriados para a captura do gesto e contexto de interesse, de modo que a observação pode se aproveitar do mesmo sensor para estimar gesto e contexto, com intuito de minimizar o número de sensores. Neste trabalho, identificamos duas

típicas características que apresentamos na Taxonomia dos Gestos (2.4), são elas: posição e rotação. Tais características são variantes para diferentes posições e orientações na execução de um gesto, e tal variância deve ser considerada na análise. Embora possa haver várias características relevantes que são originadas do contexto para a percepção de um gesto mais preciso, este trabalho ira aplicar somente as características do próprio gesto devido à grande complexidade na análise paralela; e

- *Estimativa dos parâmetros com base em um modelo:* nesta função pode se também definir quais os parâmetros, por exemplo, se somente há necessidade de comparar a execução do movimento que se leva em consideração de características como posição e rotação para o reconhecimento do gesto, ou não, o conhecimento detalhado de características de um gesto é necessário quando o modelo é usado para avaliar as qualidades especificas de um gesto.

2.5.2.2 Reconhecimento

A tarefa de reconhecimento dos gestos, é de buscar um dos modelos de gesto que mais se aproxima do gesto de entrada. O reconhecimento de gesto na sua forma primária como comparação de movimentos, pode-se utilizar de uma abordagem heurística que observa tendências simples ou picos em um ou mais dos valores do sensor. Entretanto, como foi observado na Taxonomia dos Gestos os gestos espaciais tridimensionais são ações espaço-temporais que são executadas com variação de duração do tempo. Assim, a execução do gesto se observa a natureza temporal como uma característica, podendo ser descrita como uma série temporal, e utilizar-se de técnicas de reconhecimento e comparação de séries temporais (JUNIOR et al., 2012).

2.5.3 Modelagem

As seções anteriores, descreveram a captura e observação do gesto, também é discutido a produção e percepção do gesto em aspectos gerais, a partir do conhecimento de variadas áreas. O conjunto desse conhecimento nos leva a elaboração de um modelo de gesto que se compõe de subsistemas que suportam o desenvolvimento de uma interface de reconhecimento de gesto, foco deste trabalho.

Na produção do gesto sensores visuais, atuam no processo de observação, pela captura dos dados brutos, a partir disto geram informações que são processadas. Primeiro, o sistema processa em duas etapas: segmentação e representação.

A primeira etapa, a de segmentação, é a extração das características a partir dos dados brutos do sensor para que se torne características e parâmetros para a representação.

A segunda etapa, é a representação, que é unidade de um gesto, ou seja, um conjunto de parâmetros e características extraídas a partir da segmentação. A unidade torna-se uma forma para o registro e avaliação, bem como para o reconhecimento.

O modelo pode ter a percepção do gesto, através dos componentes de subsistemas de: registro, reconhecimento e avaliação. Usando o registro, os usuários podem criar seus próprios gestos e adicioná-los ao sistema. O reconhecimento identifica o tipo de gesto de entrada desconhecido. A avaliação mede a qualidade do gesto de entrada, e os resultados da avaliação são apresentados através da interface do gesto.

2.5.4 Segmentação e Rastreamento

A segmentação, é um processo que consiste em buscar um objeto de interesse em determinado sinal. No contexto de reconhecimento de gestos, o objetivo deste trabalho é buscar a presença do corpo humano em movimento como objeto de interesse em sensores e câmeras que capturem os movimentos tridimensionais, podemos distinguir duas abordagens para essa busca: detecção baseada sobre um sinal, extrai-se de um único quadro ou utilizar-se do espaço temporal, analisando assim uma sequência de quadros.

Segundo Rautaray e Agrawal (2015), as técnicas de reconhecimento de gestos utilizando-se de sensores visuais de interação natural, é composta por três fases fundamentais: detecção, rastreamento e reconhecimento.

Fonte: Adaptada de (RAUTARAY; AGRAWAL, 2015)

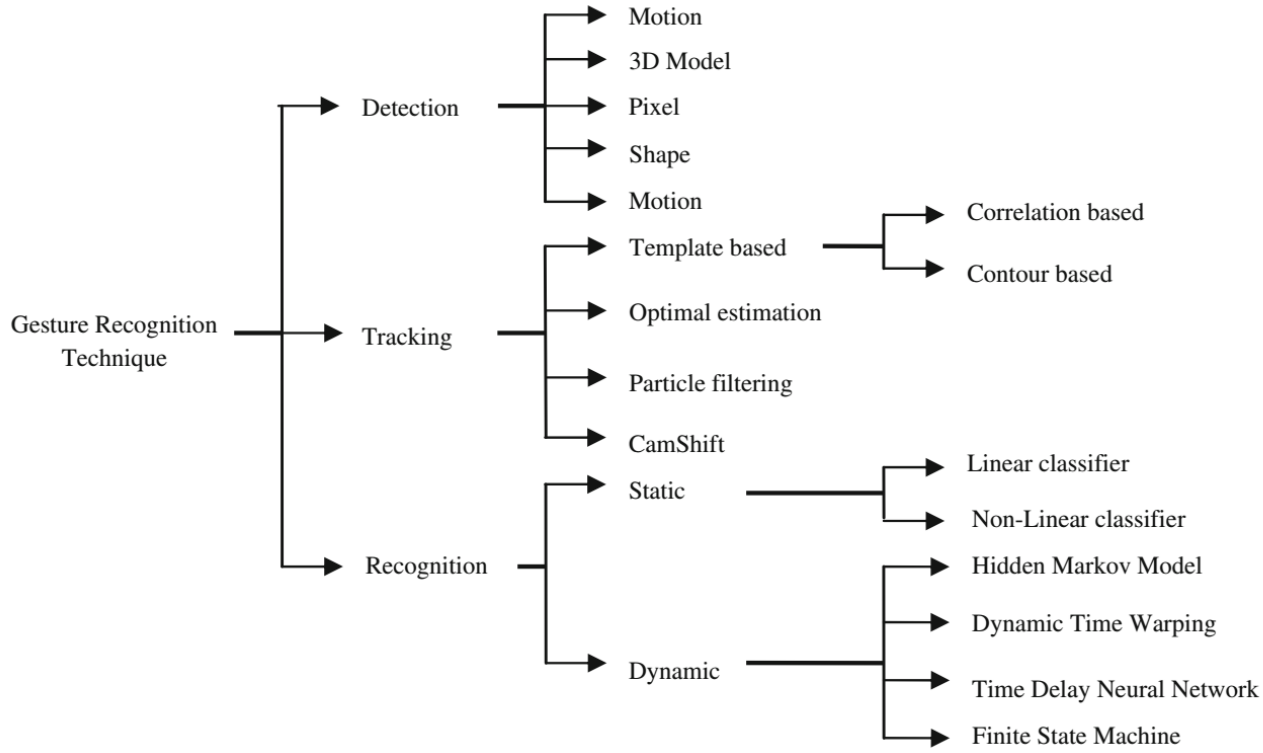


Figura 2.14: Técnicas de Reconhecimento de Gestos de Sensores Visuais de Interação Natural

O primeiro passo consiste na detecção, as técnicas desenvolvidas para localizar e segmentar é fundamental para isolar os dados relevantes para as próximas fases. Um grande número de métodos tem sido proposto na literatura que utilizam vários tipos de características visuais e, em muitos casos, a sua combinação.

A detecção consiste na busca pelo reconhecimento do corpo, tal detecção determina a segmentação, esta tarefa traz dados e informações para na sequência ser rastreado e reconhecido o gesto.

- *Cor :* a cor fornece informações substanciais para o reconhecimento do alvo, existem diferentes espaços de cores como: RGB, HSV, YCrCb, YUV e HSL. A escolha de qual utilizar é relativa à sua robustez em relação as mudanças na iluminação e orientação superficial do alvo. (RAUTARAY; AGRAWAL, 2015) apresentam trabalhos que revisam diferentes modelos da pele e avaliam seu desempenho para cada espaço. Para aumentar a invariância contra a variabilidade da iluminação, alguns métodos tentam buscar a cromaticidade da pele em vez do seu valor de cor aparente. Eles tipicamente eliminam o componente de luminância, para remover o efeito de sombras, mudanças de iluminação, bem como modulações de orientação da superfície da pele em relação às fontes de luz. No entanto,

a maioria destes métodos ainda são sensíveis a mudanças rápidas ou condições de iluminação mista. Em geral, a segmentação de cores pode ser confundida por objetos de fundo que têm uma distribuição de cores semelhante à da pele humana, uma forma de lidar com esse problema é baseando em técnicas de subtração de fundo;

- *Forma :* é uma característica que tem sido utilizada para detecção em imagens de várias maneiras. Muitas informações podem ser obtidas apenas extraindo os contornos de objetos na imagem, mas utilizando câmeras bidimensionais a forma pode ser impedida a detecção por oclusões ou pontos de vista degenerados. Em geral, a extração de contorno para se obter a forma baseia na detecção de bordas que resulta em um grande número de arestas que podem pertencer ao objeto de interesse, mas também podem pertencer a objetos irrelevantes, e por meio disto é necessárias técnicas de pós-processamento para aumentar a fidedignidade; e

- *Pixel :* proporcional um conjunto de vetores de movimento que definem a translação dos *pixels* numa região, o fluxo óptico associa um vetor que aponta para a posição do mesmo pixel na sequência. Esta associação é realizada usando uma restrição no brilho, assumindo a constância dos *pixels* correspondentes em quadros consecutivos. Esse recurso é comumente usado para segmentação baseada em movimento e aplicações de rastreamento.

- *Modelos tridimensionais :* uma das vantagens destes métodos é que eles podem conseguir a detecção independente do campo bidimensional e detectam a profundidade.

O rastreamento, que em sua versão básica, pode ser pensado como o problema de identificar um alvo, ou múltiplos alvos, situados em um plano de imagem, e seguindo seu movimento. Um algoritmo de rastreamento, consiste em três fases-chave: A primeira fase, a detecção de objetos em movimento de interesse; Segunda fase, o rastreamento de tais objetos ao longo do tempo, ou mais especificamente quadro a quadro; e A análise do alvo para reconhecer seu comportamento para o rastreamento.

- *Detectores de ponto :* através de pontos de interesse em cada quadro, como as bordas ou cantos dos objetos, desse modo os pontos de interesse devem ser invariantes em relação à pose da câmera e às mudanças de condição de luz;

- *Aprendizagem supervisionada :* neste caso, há um que sistema aprende a detectar um alvo através de um treinamento em conjunto que são compostos por diferentes visões do mesmo objeto;

- *Subtração de fundo :* a detecção é realizada através da construção de modelo de fundo e, em seguida, para cada quadro se busca as diferenças e ou mudanças relevantes, em seguida, as regiões modificadas são agrupadas, se possível, em conjunto ao alvo; e
- *Segmentação:* o quadro é segmentado em regiões com o objetivo de simplificar a forma como a imagem é representada. Desta forma, os *pixels* estão agrupados em regiões, o alvo pode ser localizado pesquisando particular características, como intensidades de cores, texturas ou bordas.

2.5.5 Métodos de reconhecimento de gestos

O reconhecimento de gestos por dispositivos computacionais é frequentemente realizado através de métodos que comparam um sinal de entrada a um conjunto de sinais de um modelo armazenado previamente (AGGARWAL; RYOO, 2011; MOESLUND; GRANUM, 2001; DUDA; HART; STORK, 2012). Estes métodos se originam do reconhecimento de padrões, uma habilidade humana e também presente em outros animais. Padrões são os meios pelos quais o mundo é interpretado e, a partir dessa interpretação, nos elaboramos atitudes e decisões.

Segundo Tou e Gonzalez (1974), padrões são as propriedades que possibilitam o agrupamento de objetos semelhantes dentro de uma determinada classe ou categoria, mediante a interpretação de dados de entrada, que permitam a extração das características relevantes desses objetos. De acordo com Castro e Prado (2002) define que, o reconhecimento de padrões é um procedimento, em que se busca, a identificação de certas estruturas nos dados de entrada, em comparação a estruturas conhecidas, e sua posterior classificação dentro de categorias, de modo que, o grau de associação seja maior entre estruturas de mesma categoria, e menor entre as categorias de estruturas diferentes. A busca por similaridades entre as classes é bastante subjetiva, pois depende de inúmeros fatores pois podem pertencer a um padrão ou não, os fatores que exercem influência são suas estruturas que tem origem do domínio de aplicação, e as características do método escolhido para buscar as similaridades.

Segundo Aggarwal e Ryoo (2011), Moeslund e Granum (2001), os métodos de reconhecimento de padrões, que mais se destacam na tarefa de reconhecimento de gestos são eles: *machine learning* e *template matching*. Entretanto, deve-se considerar vários aspectos para a escolha do método e da abordagem, tais como incertezas nas medições, semântica, tempo de processamento, custo computacional, quantidade de observações e ruídos.

Os algoritmos de *machine learning*, possuem como base a característica de aprendizado, neste sentido possuem um grande número de parâmetros, que devem ser determinados no trei-

namento, uma base de treinamento deve ser estabelecida que incluí gestos a ser reconhecido, e gestos a não ser reconhecidos, para alguns modelos e algoritmos os parâmetros devem ser tratados manualmente para uma boa precisão de classificação.

Um modelo de reconhecimento de gestos pode ser referido como um classificador, no entanto, algumas abordagens de *machine learning* podem ser formuladas utilizando o problema de regressão. Ao contrário de um classificador comum, que exibe um valor discreto em relação a qual classe o gesto de teste pertence com a maior probabilidade, a saída de um regressor geralmente é um valor contínuo dentro de um intervalo predefinido. O uso de um regressor tem a vantagem de fornecer não apenas a classificação usando um limite heurístico, mas também fornece informações sobre o estado do gesto durante sua execução no contexto do gesto e execução, em desvantagem, em geral nas abordagens de *machine learning* quanto maior o conjunto de recursos usados para o reconhecimento, é também necessário um conjunto de dados de treinamento maior e parâmetros a serem analisados, os algoritmos mais utilizados no reconhecimento de gestos são Support Vector Machine (SVM) , Hidden Markov Model (HMM) e redes neurais (AGGARWAL; RYOO, 2011; MOESLUND; GRANUM, 2001).

O método de *template matching* é uma técnica simples de implementação, pois as classes (gestos) que se pretendem a reconhecer são padrões. Um novo gesto (P), para reconhecer é necessário correlaciona-lo com os diversos padrões de exemplo que estão presentes num conjunto de pré-armazenado $Cg = (P_1, P_2, ..., P_n)$, e determinar um grau de correspondência entre eles.

A implementação mais simples para cada padrão é estabelecendo uma classe (C), assim, não existindo classes vazias $(C \neq \emptyset)$. Para se obter maior correspondência cada classe, pode ser representada como um subconjunto de Cg , sendo $C' \subseteq Cg$, ou seja, cada classe pode conter N padrões de exemplo para comparação. Para a classificação Px, compara-se com os elementos de Cg, sendo conferido a classe do elemento que apresenta maior correspondência.

O custo de comparar P' com todo o conjunto Cg tende a aumentar proporcionalmente com a cardinalidade de Cg, pois a complexidade computacional é $0(n)$, e isto pode se exigir mais processamento. Há abordagens no qual tentam diminuir a cardinalidade de Cg, por sua vez diminuir o tempo de processamento empregado na classificação. De uma forma geral, todas elas tentam normalizar o conjunto de exemplos, ou aplicar operações que compensem variações irrelevantes na posição, tamanho ou orientação do gesto a uma só classe, a geração de um padrão médio $\overline{P}$.

Assim, o método de *template matching*, apresenta a vantagem de ser fácil de treinar porque os protótipos das classes ou gestos que se pretendem reconhecer são simplesmente padrões de exemplo. No entanto, há necessidade de utilizar um grande número de protótipos para tornar

o sistema mais semântico às várias formas de representação dos gestos, em relação as suas características, desta forma este método possa ser considerado computacionalmente inapropriado. Um aspecto relevante a considerar, é que entre as classes elas podem ter dimensões diferentes, um exemplo de dimensão temporal, neste sentido também deve ser aplicado técnicas para apoio ao método.

Na Percepção do Gesto demonstrado anteriormente, através de observações da Análise que é composta da detecção de características relevantes. A Taxonomia dos Gestos, nos demonstra que ainda assim podemos identificar características especificas e passiveis, a serem analisados de forma independe ou em conjunto, de modo que, em alguns gestos que podem ser executados através da dimensão temporal.

Portanto, podemos classificar os gestos em dois tipos estáticos e dinâmicos: Os estáticos, como exemplo a observação de uma postura; Os dinâmicos, como exemplo a observação de uma caminhada ou mesmo um aceno. Assim esta definição dos gestos estáticos e dinâmicos trazem um contexto semântico substancial do domínio e natureza estatística, para a escolha de quais métodos comparativos a se utilizar na tarefa de reconhecimento gestos.

2.5.5.1 Gestos Estáticos

Um gesto estático, pode ser entendido como uma observação que se encontra estática, e que não possui uma continuidade na dimensão temporal, utilizando estes sinais estáticos podem ser aplicados diversas abordagens para encontrar similaridades entre eles.

Para medir a similaridades entre as classes, podemos considerar os dados de várias formas para análise, podemos utilizar os sinais originais diretamente, ou aplicar transformações e métricas.

A aplicação de métricas para medição de similaridade, tem como objetivo de ressaltar características especificas, porém não pode servir como base para utilização genérica, pois vai de encontro ao domínio em especifico, e cabe uma análise prévia. Em paralelo, pode se utilizar de técnicas de transformação, em geral, segundo Junior et al. (2012), possui dois objetivos principais, o de reduzir a quantidade informações possíveis e isolar características especificas de seus componentes, dentro os principais exemplos são *Principal Components Analisys* (PCA) e Transformada de *Fourier*.

O uso dos sinais originais dos gestos estáticos sem nenhuma alteração, ou não, nos permitem calcular a distância entre as classes através do espaço funcional. Utilizando a análise funcional das métricas do espaço L_p derivadas da Norma L_p.

Segundo Junior et al. (2012), as medidas da Norma L_p estão entre as medidas de distâncias mais conhecidas e exploradas na literatura, em especial a distância Euclidiana. A distância Euclidiana é uma medida intuitiva de considerar a distância, sua simplicidade de implementação e interpretação favorece a seleção desta medida para calcular a similaridade em um grande número de aplicações, o seu uso pode ser aplicado em dados bidimensionais, tridimensionais e em maior número de dimensões.

A notação do cálculo da distância baseado na Norma L_p se obtém no espaço R_n, onde cada vetor é considerado um ponto no espaço $N-$dimensional. Assim a similaridade entre esses vetores é dada pela diferença desses pontos. Essa norma é definida pela Equação 2.4.

$$L_p(R,S) = \left(\sum_{i=1}^{N} |S_i - r_i|^p \right)^{\frac{1}{p}} \tag{2.4}$$

Onde R e S são vetores $N-$dimensional e p define a medida de distância a ser utilizada. De acordo com o valor de p, obtemos medidas de distâncias diferentes, no qual possuem comportamento diversos e específicos. Através da Figura 2.15 é aplicado a uma distribuição em duas dimensões, podemos ver o efeito da variação de p, no mesmo espaço geométrico em 0.5, 1, 2 e ∞.

Fonte: Elaborada pelo Autor

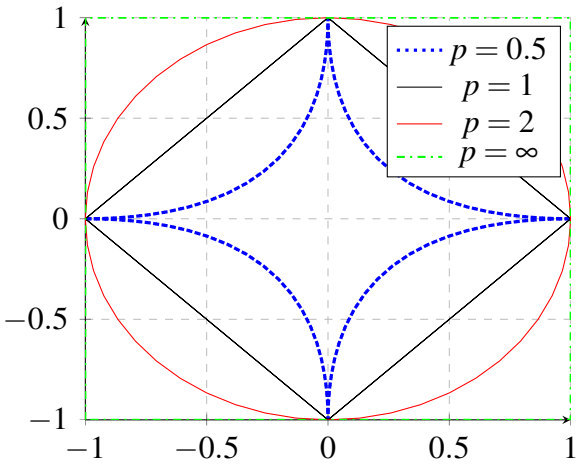


Figura 2.15: Variação de p na Norma L_p para $p = 0.5$, $p = 1$, $p = 2$ e $p = \infty$

Na Figura 2.16 já é aplicado a uma distribuição em três dimensões e podemos também observar o efeito da variação de p em 1, 1.5, 2, 3, 4, 8, 16, 32 e ∞.

Fonte: (AKLEMAN; CHEN, 1999)

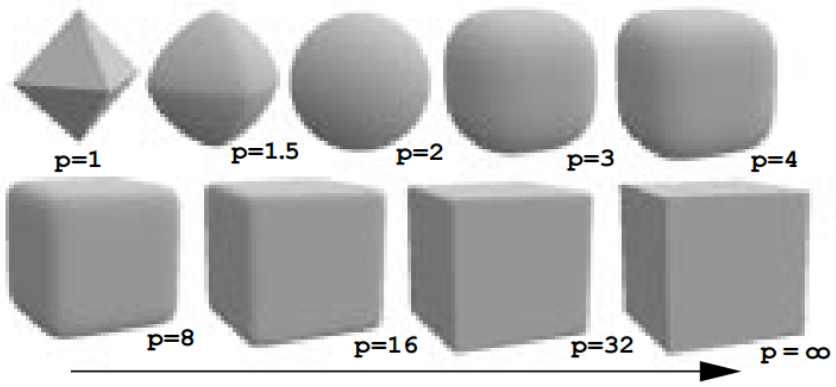


Figura 2.16: Variação de p na Norma L_p para $p = 1$, $p = 1.5$, $p = 2$, $p = 3$, $p = 4$, $p = 8$, $p = 16$, $p = 32$ e $p = \infty$

Através da variação demonstrada nas Figuras 2.15 e 2.16, podemos observar que a distância fracionaria, atribui mais pesos a pequenas variações entre os dados e pode ser visualizada no valor de p = 0.5 e 1.5, e tendem a se retrair em comparação as inteiras. A distância de p = 1, define um espaço geométrico onde todos os pontos possuem o mesmo valor da soma das diferenças absolutas de cada ponto. A distância de p = 2, define um espaço geométrico em forma de circunferência onde todos os pontos estão equidistantes em relação ao centro. A medida que se aumenta o valor de p ao ∞, pode se alcançar um espaço quadrático e um custo computacional maior.

A Norma L_p, possui medidas empregadas em diversos trabalhos em variados domínios, segundo Junior et al. (2012), as mais conhecidas são a distância Euclidiana p = 2 e a distância Manhattan p = 1. A forma de comparação de p = 1, é tipicamente mais tolerante a *outliers*, e tende a encolher coeficientes esparsos dinamicamente, enquanto que p = 2, encolhe todo o coeficiente pelas mesmas proporções, mas não elimina nenhum.

A distância de p = 1, fornece uma solução quando o sinal de dados é esparso, entretanto p = 2, é adequada para soluções não esparsas limitados a uma largura. Em resumo a distância de p = 1, tem a função de minimizar a soma das diferenças absolutas, enquanto p = 2 minimiza a soma do quadrado das diferenças entre o valor alvo e os valores estimados.

2.5.5.2 Gestos Dinâmicos

Um gesto dinâmico, pode ser entendido como uma observação que se encontra em movimento e possui uma continuidade na dimensão temporal, no qual podem ser aplicadas diversas

abordagens para encontrar similaridade entre estes sinais contínuos.

Se observa que um gesto dinâmico, pode ser entendido como uma Série Temporal, segundo Morettin e Toloi (2006), uma *ST*, pode ser entendida, como qualquer conjunto de observações que se encontram ordenadas no tempo e que são originadas a partir da saída de um sistema dinâmico, o sistema dinâmico no contexto de gestos, podemos atribuir à Percepção do Gesto. Nota-se um sistema dinâmico pela Equação 2.5.

$$\begin{aligned} v_1 + 1 &= f(v_t, u_t) \\ z_t &= g(v_t) \end{aligned} \tag{2.5}$$

Em que u e v, representam os estados das entradas do sistema e $u \in \mathbb{R}$ e $v \in \mathbb{R}$; f e g, são funções não-lineares desconhecidas; z_t é uma saída escalar conhecida, portanto, pode-se definir uma série temporal Z, de tamanho m como um conjunto ordenado de valores, ou seja, $Z = (z_1, z_2, ..., z_m)$ onde $z_t \in \mathbb{R}$ representa uma observação z em um instante t.

Segundo Giorgino et al. (2009), o *Dynamic Time Warping* (DTW) é uma técnica muito popular para medir a similaridade entre duas sequências, que podem variar em tempo ou velocidade utilizando *template matching*. Seu principal destaque foi a partir dos anos 70, para o reconhecimento de fala e caracterização fonéticas das palavras (SAKOE; CHIBA, 1978; RABINER; JUANG, 1993).

O DTW, pode ser aplicado a qualquer sinal que possa ser convertido em uma representação linear. Sua aplicação mais conhecida tem sido o reconhecimento de fala, mas segundo Senin (2008), também é utilizado em outras aplicações que incluem : escrita manuscrita e correspondência de assinaturas online; reconhecimento de gestos; *data-mining*; séries temporais; visão e animação computacional; vigilância; engenharia química; música; e processamento de sinal.

O algoritmo se destaca onde as variações de tempo indicam um problema, onde é necessário o alinhamento temporal para uma comparação. Em geral, o DTW é um método que calcula uma combinação ideal entre duas sequências dadas como séries temporais. As sequências são deformadas não linearmente na dimensão temporal para determinar uma medida de sua similaridade independente de certas variações não-lineares na dimensão temporal, o DTW se torna uma solução para um dos principais problemas das distâncias da Norma L_p, adicionando robustez à comparação na dimensão temporal (JUNIOR et al., 2012).

Posteriormente que o DTW, mede a similaridade entre a entrada e os modelos de comparação, a entrada pode ser admitida como um membro da mesma classe que o modelo para o qual é mais semelhante e/ou mais próximo, ou rejeitada como pertencente a nenhum das possíveis

classes, se a medição for superior ao limiar de similaridade.

A correspondência entre os modelos não possui uma abordagem formal e iterativa ao treinamento que classificadores estatísticos que redes neurais possuem, mas por ser uma técnica não paramétrica pode empregar os quadros originais diretamente para o reconhecimento de gestos, funcionando mesmo quando apenas um conjunto de dados de treinamento está disponível.

Em uso no reconhecimento de gestos, por exemplo, as semelhanças nos padrões de uma caminhada podem ser detectadas, mesmo em um vídeo em que uma pessoa estivesse caminhando lentamente e, se em outro ela estivesse caminhando mais rapidamente, ou mesmo se houvesse acelerações e desacelerações durante o curso de uma observação de um sensor, com diferentes usuários.

O DTW, não pode se adaptar a natureza probabilística do sinal, assim ocorre a necessidade de adaptação e normalização dos dados, que podem exercer um papel crítico no desempenho do sistema na aplicação do reconhecimento de gestos dinâmicos, uma vez que a maioria dos gestos não é reproduzida de forma invariante, conforme apresentado na Produção do Gesto.

Neste capítulo, observamos toda fundamentação teórica sobre o tema de reconhecimento de gestos do corpo humano, neste sentido concluímos que a partir de uma observação do movimento através de dispositivos e sensores visuais conseguimos fazer a captura e rastreamento das informações para descrever a unidade de um gesto e por fim utilizar de técnicas e métodos para a medição de similaridade entre outros gestos.

Capítulo 3

MODELO PROPOSTO

Este capítulo, descreve o modelo proposto para realizar o reconhecimento de gestos estáticos e dinâmicos. Uma visão geral do modelo é apresentada na Figura 3.1, há três partes que compõe este modelo a Aquisição do Gesto, Modelagem do Gesto e Interface Multimodal.

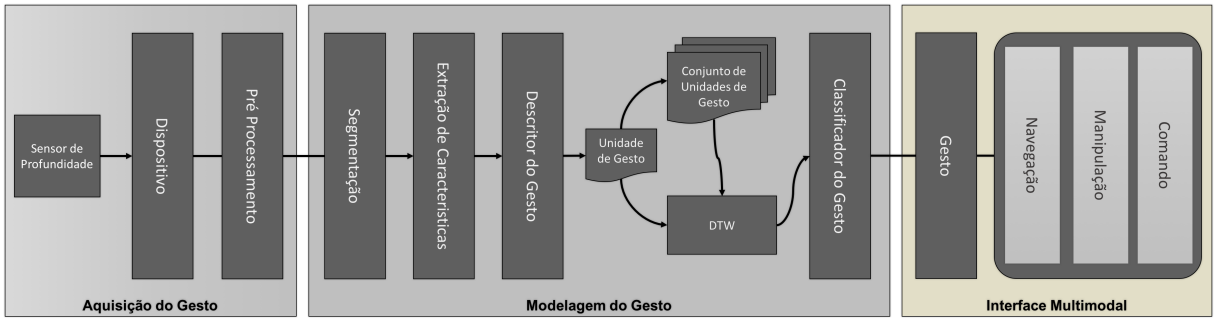


Figura 3.1: Modelo Proposto

Aquisição do Gesto: Durante a Aquisição do Gesto, os gestos podem ser adquiridos através de um Sensor de Profundidade, assim o Dispositivo faz um Pré Processamento e disponibiliza os dados brutos para a parte seguinte de Modelagem do Gesto.

Modelagem do Gesto: Em seguida, na Modelagem do Gesto os dados brutos adquiridos são processados através da Segmentação, que busca definir e buscar o objeto de interesse na cena visualizada pelo sensor, o corpo humano. Na próxima etapa, a Extração de Características foca na seleção e extração de características relevantes do corpo humano, que será utilizada para auxiliar na tarefa do reconhecimento do gesto, e as normaliza, a normalização se torna necessária devido as fontes de erros estimadas definidas na Produção do Gesto.

Após as informações coletadas na Extração de Características, elas são estruturadas e representadas pelo Descritor do Gesto, que gera uma Unidade de Gesto. Uma Unidade

de Gesto pode ser atribuída a um conjunto, que possui outras unidades que já foram descritas anteriormente, e assim formam o Conjunto de Unidades de Gesto.

Por sua vez, o algoritmo DTW é utilizado para comparar a similaridade entre Unidades de Gestos. O DTW, auxilia o Classificador do Gesto, que possui a função de buscar o gesto mais correspondente, comparando o gesto a ser conhecido ao Conjunto de Unidades de Gesto.

Interfaces Multimodais: Por fim, o gesto reconhecido pode ser utilizado como interface de entrada, em sistemas de interfaces multimodais para navegação, manipulação ou comando.

Com a finalidade de demonstrar o uso do Modelo Proposto, é apresentado o emprego do modelo com o sensor visual Microsoft Kinect, utilizando a segmentação do corpo humano com a extração de características de origem das juntas para suportar o registro, classificação e reconhecimento dos gestos espaciais tridimensionais.

3.0.1 Aquisição dos Gestos

A aquisição dos dados é o início para o modelo. Nesta fase que intitulamos de aquisição dos gestos, é importante a definição do sensor corretamente, pois a escolha atuará diretamente no desempenho, estrutura, restrições dos movimentos, posição do sensor em relação ao objeto e frequência de amostragem. Para obtenção da informação citamos diversos sensores para este objetivo.

Neste modelo abordamos o dispositivo Microsoft Kinect para a aquisição da informação a partir das imagens de profundidade disponibilizadas por este sensor. Para o acesso ao sensor e às imagens de profundidade foram testados dois frameworks, Microsoft SDK (MS SDK) e Open Natural Interaction (OpenNI) . Os dois frameworks revelaram um bom desempenho, porém o MS SDK, é de código fechado e desenvolvido apenas para suportar o dispositivo Microsoft Kinect, gerando uma desvantagem para o framework e objetivo deste trabalho. Enquanto o OpenNI de código aberto, foi desenvolvido independente do sensor. Por fim, a escolha foi a utilização do dispositivo Microsoft Kinect juntamente com o framework OpenNI, esta escolha também se deve ao fato de maior sucesso em aplicações comerciais e não comerciais e maior abrangência da comunidade de desenvolvedores.

3.0.1.1 Drivers e Frameworks Open Source

No seu lançamento, a Microsoft não divulgou seus *drivers* para que o Microsoft Kinect pudesse ser utilizado em computadores pessoais. Mas a comunidade de software livre dedicou a explorar os dados de entrada e saída que o dispositivo possui, e deixa-lo disponível para uso em computadores pessoais.

Desde então, alguns *drivers* de código aberto, *Software Development Kit* (SDK) e *Application Programming Interface* (API) surgiram. A Tabela 3.1, apresenta e compara os principais *drivers* de código aberto.

Tabela 3.1: *Drivers* e *Frameworks Open Source*

Nome	Linguagens	Plataformas	Recursos	Hardware
OpenNI	C, C++, Java e Processing	Windows, GNU/Linux, OS/X, Android	Identificação do usuário Reconhecimento de gestos das mãos. Rastreamento de juntas. Cor e profundidade imagens. Registro de dados de cor e profundidade em arquivo	Qualquer compatível com padrão OpenNI. Referência PrimeSense
Robot Operating System (ROS)	Python, C++	GNU/Linux	Cor e profundidade imagens. Motor e controle de LED.	Referência PrimeSense
CL NUI SDK and Driver	C, C++, WPF/C#	Windows	Cor e profundidade imagens. Dados do acelerômetro. Motor e controle LED.	Microsoft Kinect
OpenKinect / libfreenect	C, Python, actionscript, C#, C++, Java etc.	Windows, GNU/Linux, OS/X	Cor e profundidade imagens. Dados do acelerômetro. Motor e controle LED. Fakenect Kinect Simulator. Gravação de todos os dados em arquivo	Microsoft Kinect

Buscando-se o melhor *driver* e *framework* multi-plataforma e *open source* para execução em diversos sistemas operacionais e dispositivos, juntamente com a característica para segmentação do corpo e rastreamento das juntas do corpo. A OpenNI se torna a opção mais adequada, contendo funcionalidades úteis que são convenientes para este trabalho e modelo proposto.

3.0.1.2 OpenNI/NITE

A Open Natural Interaction (OpenNI) (DIAS et al., 2013b) fornece uma API para o desenvolvimento de aplicações que necessitam de interação natural. A API foi desenvolvida pela organização OpenNI que foi fundada pela PrimeSense. A OpenNI é uma organização sem fins lucrativos, formada pela própria indústria de fabricantes para certificar e promover a compatibilidade e interoperabilidade de equipamentos de interação natural. A OpenNI está disponível para as plataformas Windows, OS/X e GNU/Linux é escrito originalmente em C++ e disponibilizado sob a GNU Lesser General Public Lincese (LGPL) , sendo o código fonte livremente distribuído e disponível ao público geral.

Características da API é de ser multi-linguagem e multi-plataforma, existir componentização modulares (chamados *Node Generators*), responsáveis por diferentes tarefas de detecção e rastreamento. Estes componentes podem ser utilizados em separado ou em conjunto.

A API abrange a comunicação com dispositivos de baixo nível (por exemplo, sensores de visão e áudio), no caso deste trabalho o dispositivo Microsoft Kinect, bem como soluções de alto nível (por exemplo, o acompanhamento visual utilizando visão computacional). Toda sua estrutura é baseada em eventos, realizando *call-backs* sempre que um evento ocorre. Permite que dados da execução de uma aplicação sejam gravados para serem reproduzidos posteriormente, funcionando como uma simulação.

Por padrão, a OpenNI permite que os desenvolvedores de aplicativos de interação natural extrair elementos de um ambiente real através do *input* de dados feito por sensor de profundidade como dispositivo o Microsoft Kinect (por exemplo, a representação de uma pessoa, a representação de mãos, uma matriz de *pixels* em um mapa de profundidade etc.)

A arquitetura da OpenNI pode ser descrita em três camadas, como demonstra a Figura 3.2. A camada inferior contém o hardware ou os dispositivos que coletam os dados do ambiente real (ou seja, elementos visuais e/ou de áudio). A segunda camada contém os componentes de *middleware* que interpretam e analisam os dados do sensor. Finalmente, a camada superior contém o software que implementa aplicações de interação naturais.

Fonte: (GNECCO et al., 2012)

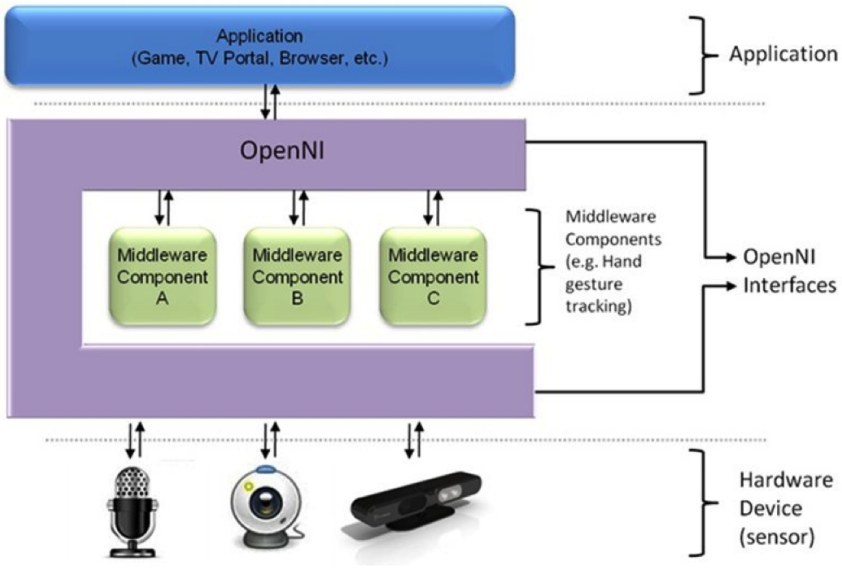


Figura 3.2: Arquitetura OpenNI

Como componente de *middleware* a NITE, uma implementação encapsulada na API da OpenNI, para dispositivos de arquitetura de *design* PrimeSense, e no qual foi desenvolvido pela

própria empresa. Apesar de ser de código fechado, é gratuito e pode ser usado comercialmente. Ele é responsável de prover a biblioteca que detecta pessoas a partir de uma sessão em que é aguardado a silhueta e, após identificado e finalmente rastreado. Possui três modos básicos de operação: um que permite rastrear e detectar gestos das mãos, análise de cena (o que é chão e parede) e outro permite rastrear o corpo todo (esqueleto), dando informações sobre as juntas do corpo.

Como apresentado a OpenNI, funciona por componentes modulares chamado de *Production Nodes* (nós de produção), que são um conjunto de componentes, que estabelecem o processo de criação de dados necessários para aplicações baseadas em NUI (GNECCO et al., 2012). Cada nó encapsula a funcionalidade que se relaciona com a geração do tipo de dados específicos.

Esses nós de produção, são os elementos fundamentais da interface OpenNI. No entanto, a API dos nós de produção apenas define a linguagem. A lógica de geração de dados deve ser implementada pelos módulos que se conectam com a OpenNI.

Cada nó de produção da OpenNI, tem um tipo e pertence a uma das seguintes categorias:

- Produção de nós de dispositivos (sensor)

  Dispositivo:

  Um nó que representa um dispositivo físico (por exemplo, um sensor de profundidade, ou uma câmera RGB). O papel principal deste nó é permitir a configuração do dispositivo.

  Gerador de Profundidade:

  Um nó que gera uma profundidade e mapa. Este nó deve ser implementado por qualquer sensor 3D que deseja ser certificado como compatível OpenNI. A classe *DepthGenerator*, gera um mapa de profundidade como uma matriz de *pixels*, em que cada *pixel* é um valor de profundidade representando uma distância a partir do sensor em milímetros.

  Gerador de Imagem:

  Um nó que gera imagens coloridas de mapas. Este nó deve ser implementado por qualquer sensor de cor que deseja ser certificado como compatível OpenNI

  IR Gerador:

  Um nó que gera imagem IR de mapas. Este nó deve ser implementado por qualquer sensor infravermelho que deseja ser certificado como compatível OpenNI.

Gerador de Áudio:

Um nó que gera um fluxo de áudio. Este nó deve ser implementado por qualquer dispositivo de áudio que deseja ser certificado como compatível OpenNI.

- Produção de nós de *middleware*

Componente de detecção das mãos:

Componente do *middleware HandsGenerator*: Processa os dados dos sensores e gera a localização de um ponto central de cada mão (tipicamente estrutura de dados que descreve posições X,Y,Z) com *ID's* persistentes. O *HandsGenerator*, é um tipo específico de classe geradora, derivada da classe *Generator*. Componente de detecção e alerta de gestos: componente do *middleware GestureGenerator*: Identifica gestos predefinidos executados e gera um evento para o aplicativo. Um exemplo de gesto típico de ser aplicado é o gesto clique: como o usuário humano seria clicar em um botão, ou seja, fazer um movimento de empurrar e, em seguida, puxando para trás. Os principais gestos são detalhados abaixo.

Componente de análise de cena:

Componente do middleware *SceneAnalyzer*: Analisa a imagem da cena, a fim de produzir informações, tais como a separação entre o primeiro plano da cena (pessoas) e o fundo. As coordenadas do plano de chão e a identificação individual de objetos compostos na cena também são feitas. O analisador de cenas gera um mapa de profundidade em que cada pixel contém um rótulo que indica que se representa uma figura ou é parte do fundo.

Componente de análise do corpo:

Componente do *middleware UserGenerator*: Processa os dados sensoriais, identificando individualmente cada usuário e, assim, permitindo que ações sejam feitas em usuários específicos. A implementação do *middleware* NITE gera informação do corpo e permite obter a posição de 15 juntas (*joints*) do corpo. Cada conjunto é identificado com o seu nome e coordenadas X, Y, Z da posição são dadas em milímetros entre a origem e o dispositivo. O eixo X é no plano horizontal, o eixo Y está no plano vertical e no eixo Z a profundidade da direção do dispositivo.

- Para utilização de registros, os seguintes tipos de nós de produção são suportados:

***Recorder*:**

Implementa as gravações de dados

***Player*:**

Lê os dados de uma gravação e executa.

***Codec*:**

Usado para comprimir e descomprimir dados em gravações

3.0.2 Segmentação

A segmentação, neste modelo consiste em definir o objeto de interesse, o corpo humano. Uma das principais dificuldades encontradas para realizar esta tarefa é a segmentação do corpo em movimento com fundos complexos em diversas características. Com o uso do Microsoft Kinect e a OpenNI/NITE conseguimos segmentar o corpo humano através das abordagens aplicadas por ambos citadas anteriormente.

A segmentação do corpo de um usuário é obtida através do centro de massa do usuário, utilizando o *UserGenerator* a OpenNI/NITE (Seção 3.0.1.2) obtemos a máscara de *pixels* (p) que corresponde a cada usuário a frente do dispositivo e assim calculamos o Centro de Massa *CM* 3.1 do usuário:

$$\begin{aligned} CM_z &= \sum_p depth(p) \\ CM_x &= \sum_p RealWorldCoordinates_x(p) \\ CM_y &= \sum_p RealWorldCoordinates_y(p) \end{aligned} \tag{3.1}$$

O *depth* retorna à profundidade do pixel p, e função *RealWorldCoordinates* da OpenNI/NITE retorna a conversão da posição discreta do pixel na imagem para a posição deste ponto no mundo real, assim (CM_x, CM_y, CM_z) como um ponto no espaço tridimensional correspondente ao centro da massa do usuário de origem do Microsoft Kinect, como demonstra a Figura 3.3 . Após obter o corpo do usuário, podemos diferenciar ele de outros, e por fim extrair suas características especificas.

A Figura 3.4 mostra os *pixels* do usuário em azul ao resto da cena, por meio disto obtemos os *pixels* da câmera RGB do dispositivo Microsoft Kinect e fazemos uma extração de *pixels* de acordo com as coordenadas obtidas pelo *CM*.

Fonte: Elaborada pelo Autor

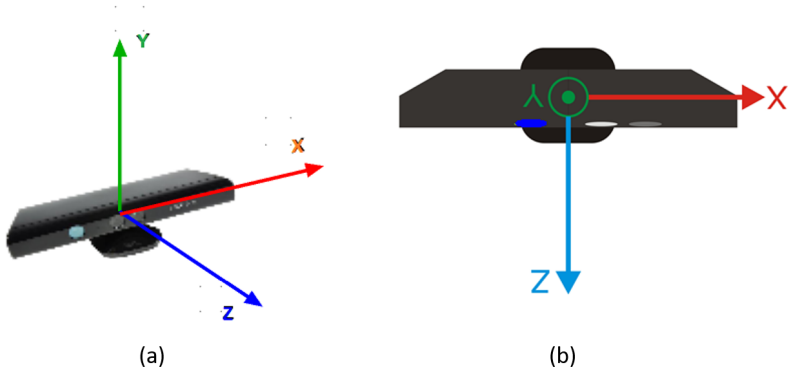


Figura 3.3: (a) Visão Lateral Espaço Tridimensional (b) Visão Superior Espaço Tridimensional

Fonte: Elaborada pelo Autor

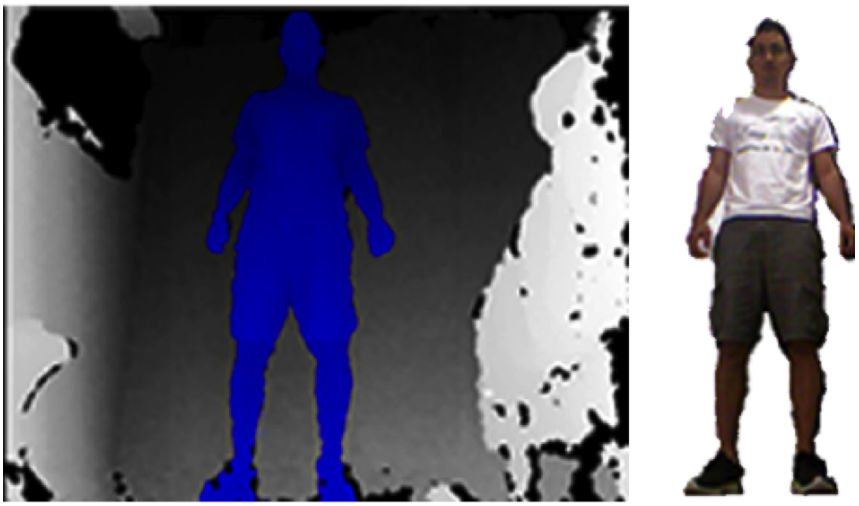


Figura 3.4: Centro de Massa do Usuário

3.0.3 Extração de Características

Tipicamente a extração de características possuem três abordagens: *appearance-based*, *view-based* e *feature-based*. As técnicas nas abordagens de *appearance-based* e *view-based* se baseiam em características da imagem como todo num espaço dimensional, enquanto a *feature-based* se baseia em características particulares presentes na imagem, tais como áreas, contornos e retângulos. Neste modelo utilizamos a *feature-based* que se entendeu mais coerente com a necessidade pois extraímos as características das juntas do corpo.

A extração e seleção de características relevantes para o gesto, se torna um passo fundamental para o processo de reconhecimento. A tarefa de extração de características concentra-se

em encontrar um grupo de características relevantes, no qual representam o gesto. Este passo é importante pois é fonte de informação para etapa de classificação e reconhecimento. Anteriormente foi demonstrado que através do sensor Microsoft Kinect podemos identificar o usuário e fazer o rastreamento das juntas do corpo no espaço tridimensional.

Através das juntas podemos construir o esqueleto, assim se tornando uma representação do corpo humano, onde as posições de juntas são representadas por pontos no espaço tridimensional. A representação do esqueleto de uma pessoa, pode ser usada para rastrear as atividades ou posturas (FUJIYOSHI; LIPTON, 1998). Shotton et al. (2011) destacam a importância em fragmentar o esqueleto em partes para obter maior precisão nos testes com reconhecimento de imagens em humanos.

Fonte: Elaborada pelo Autor

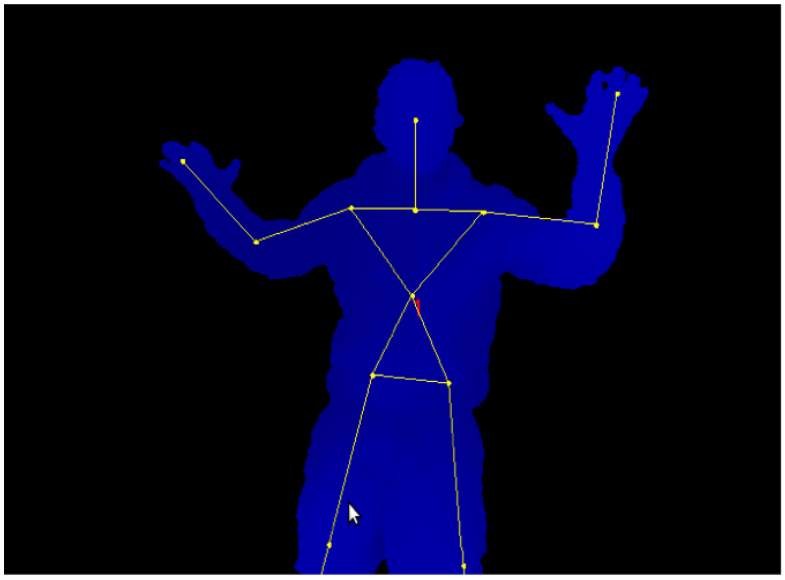


Figura 3.5: Captura das Juntas do Corpo

Nesse sentido a OpenNI/NITE (Seção 3.0.1.2) pode rastrear 15 juntas (*joints*) do corpo através do *UserGenerator*, o fornecimento das coordenadas das juntas do esqueleto em tempo real com precisões milimétricas e sua modularidade nos ajudam nesta tarefa, como demonstra a Figura 3.5. A Tabela 3.2, apresenta as juntas e descrição por nome representando as posições nas coordenadas (x, y, z) no espaço tridimensional da câmera, como demonstra a Figura 3.3.

Na tarefa de reconhecer gestos dinâmicos descritos anteriormente, neste modelo definimos que as características temporais são fundamentais, assim podemos definir em que os gestos são executados em um vetor de temporal (VT), onde é composto pelas informações das coordenadas das juntas (x, y, z), t momento de tempo e j junta do esqueleto dado pela equação.

Tabela 3.2: Juntas do Corpo OpenNI/NITE

Juntas	Descrição
XN_SKEL_HEAD	Cabeça
XN_SKEL_NECK	Pescoço
XN_SKEL_TORSO	Tronco
XN_SKEL_LEFT XN_SKEL_RIGHT	Onde LEFT e RIGHT pode se obter: Ombro, Cotovelo, Mão, Quadril, Joelho e Pé.

$$VT_t^j = (x_t^j, y_t^j, z_t^j) \tag{3.2}$$

Inicialmente este modelo trata de três tipos de situações que podem trazer variâncias nas juntas do corpo, devido, as fontes de erros estimadas na Produção do Gesto.

A primeira situação, é em relação a plano do corpo em relação ao sensor, como demonstra a Figura 3.6 , usuários em planos diferentes em relação ao sensor. A segunda situação, é em relação a posição do corpo em relação ao Microsoft Kinect, como demonstra a Figura 3.7, onde usuários podem estar mais próximo ao sensor e distante, ou mais esquerda, ou à direita do sensor no mesmo plano. A terceira situação, se dá pela altura do corpo do usuário que se diferencia, a Figura 3.8 demonstra, no mesmo plano a altura dos usuários se diferem.

Portanto, é necessário a normalização das juntas para aumentar a precisão do reconhecimento, as abordagens neste modelo proposto são baseadas em estudos que se buscam o reconhecimento de gestos e calibração das coordenadas utilizando se o dispositivo Microsoft Kinect (WEI; QIAO; LEE, 2014; TAHA et al., 2015).

Fonte: Elaborada pelo Autor

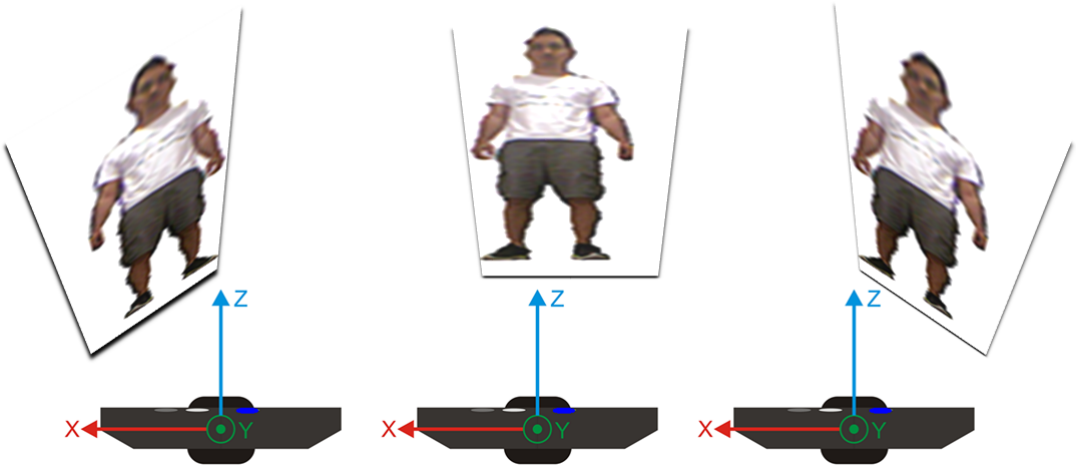


Figura 3.6: Situação de variância de plano do corpo

Fonte: Elaborada pelo Autor

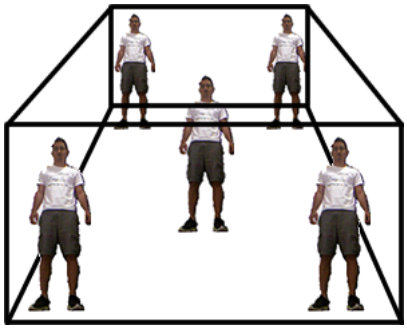


Figura 3.7: Situação de variância de posição

Fonte: Elaborada pelo Autor

Figura 3.8: Situação de variância de tamanho

No intuito de normalizar a primeira situação, a de variância de posição do plano do corpo do usuário em relação Microsoft Kinect. Se obtém o ângulo do plano do corpo em relação ao sensor, assumindo que todas as juntas estão em linha reta no mesmo plano quando o usuário projeta seus ombros a uma direção, podemos dizer que o plano do corpo sé da através das coordenadas das juntas (x, y, z) do ombro direito OD e do ombro esquerdo OE do usuário , onde o plano do corpo se torna perpendicular à direção do eixo z de origem do C centro do sensor como demonstra a Figura 3.9.

Fonte: Elaborada pelo Autor

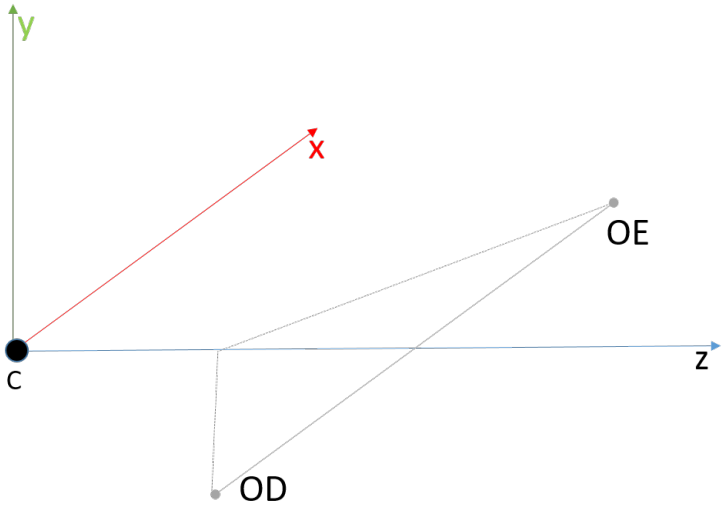


Figura 3.9: Plano do corpo em relação ao sensor

Por fim, podemos calcular o ângulo do plano do corpo em relação ao sensor, e utilizar como característica para as seguintes etapas do modelo pela equação:

$$\alpha = tan^{-1}\left(\frac{z_d - z_e}{x_d - x_e}\right) \tag{3.3}$$

Na segunda situação de variância, a posição, a distância no eixo Z, que possui o significado de profundidade em relação ao sensor pode causar variação nas coordenadas espaciais nos eixos X e Y, variando as distâncias entre as juntas. Assim, observa-se que as coordenadas das juntas, devem ser traduzidas para um outro sistema de coordenadas, onde sua origem é um ponto no corpo humano e não no sensor. Desta forma, o fator de distância entre o corpo e sensor é neutralizado, isto permite que as coordenadas sejam invariantes para a translação e rotação do corpo em relação ao sistema de referência do sensor. Assume-se o centro do corpo o XN_SKEL_TORSO, no qual iremos expressar como T, como nova origem de coordenadas. Suponha que as coordenadas das juntas de T, sejam (x,y,z). Assim, para cada junta do esqueleto i, com coordenadas (x,y,z) as coordenadas são convertidas em (x',y',z') e serão calculadas com a equação:

$$x'_i, y'_i, z'_i = (x_i - x, y_i - y, z_i - z) \tag{3.4}$$

Assim para evitar a variância de posição as coordenadas (x,y,z) do esqueleto, são convertidas em coordenadas esféricas, o sistema de coordenadas esféricas é um sistema espacial

tridimensional com três componentes: a distância do ponto da origem (distância radial r), o ângulo polar ϕ e o ângulo de azimute θ, como demonstrado na Figura 3.10. Ao normalizar uma junta nas coordenadas espaciais, todos os eixos x, y e z são alterados.

Fonte: Elaborada pelo Autor

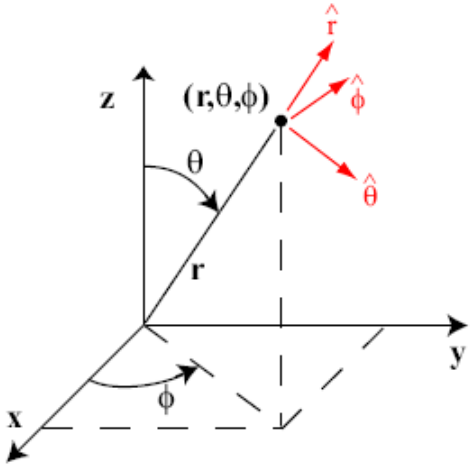


Figura 3.10: Sistema de Coordenadas Esféricas

A distância radial r, será expressada por D de um vetor (x,y,z) entre o T e a junta correspondente i onde $D_i = (T_{xyz} - i_{xyz})$. A inclinação, ou ângulo polar ϕ é o ângulo entre o T e a junta correspondente i onde $\phi_i = (T_{xyz} - i_{xyz})$, enquanto o azimute ou longitude θ, é o ângulo entre o T e a junta correspondente i onde $\theta_i = (T_{xyz} - i_{xyz})$, as seguintes equações descrevem as conversões.

$$\sum_{i=1}^{n} D_{(i)} = \sqrt{(i_x - T_x)^2 + (i_y - T_y)^2 + (T_z - i_z)^2} \tag{3.5}$$

$$\sum_{i=1}^{n} \phi_{(i)} = atan2\left(\sqrt{(i_x - T_x)^2 + (i_y - T_y)^2}, (T_z - i_z)\right) \tag{3.6}$$

$$\sum_{i=1}^{n} \theta_{(i)} = atan2\left((i_y - T_y), (i_x - T_x)\right) \tag{3.7}$$

Em sequência, a terceira situação notada a variância de tamanho do usuário, pode se tornar uma característica que traz uma diferença expressiva nas coordenadas das juntas. Segundo Bogin e Varela-Silva (2010), seres humanos embora tenham estaturas diferentes, tem proporção corpórea similar em relação a XN_SKEL_HEAD a cabeça C e tronco T, após a conversão da distância radial entre as juntas C e T temos D_{CT}, este valor nos informa a altura do usuário,

através dela podemos converter as demais distâncias D_i em relação a ela, em distâncias que consideram esta variância temos Da_i através da seguinte equação:

$$\sum_{i=1}^{n} Da_{(i)} = \frac{D_i}{D_{CT}} \tag{3.8}$$

A medida da amplitude do movimento, é um importante parâmetro utilizado na avaliação e análise do movimento do corpo (WATKINS; PORTNEY, 2009), o movimento articular ocorre em torno de um eixo que está sempre perpendicular a um plano, e é medido em graus como demonstra a Figura 3.11 (SATO; HANSSON; COURY, 2010).

Fonte: (SATO; HANSSON; COURY, 2010)

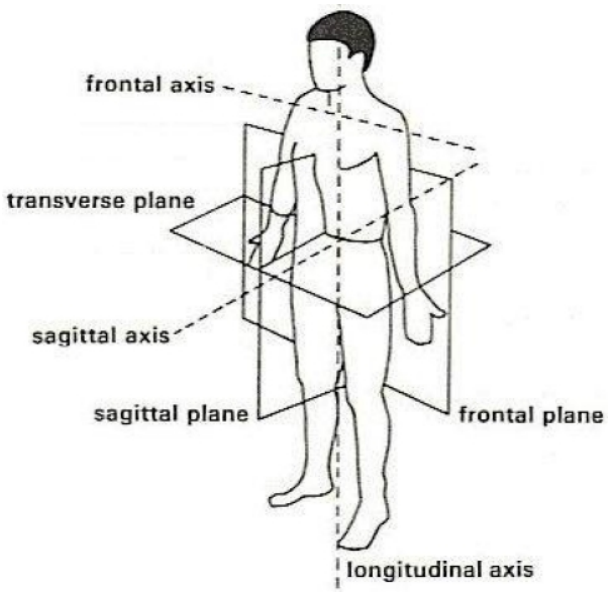


Figura 3.11: Planos e eixos do corpo

Definimos que para se obter a amplitude do movimento, é necessário obter os graus do movimento através das juntas do esqueleto, utilizando as coordenadas no espaço tridimensional (x,y,z), inicialmente abordamos o plano frontal com os movimentos do eixo sagital e o plano sagital com os movimentos do eixo transverso, pois o sensor sempre estará posicionado na horizontal. Necessariamente precisaremos de três juntas j_1 , j_2 e j_3, onde a junta objetiva, no qual desejamos obter os graus JO, será o centro e ligação das outras duas juntas que formam dois vetores A e B onde encontraremos o ângulo θ, através do θ convertemos para graus e teremos o valor desejado, a Figura 3.12 demonstra a junta j_1 como JO que buscamos o θ.

Digamos que as distâncias entre as juntas sejam $A = (j_1 - j_2)$, $B = (j_2 - j_3)$ e $C = (j_3 - j_1)$, através de $\vec{A} \cdot \vec{B} = \|\vec{A}\| \, \|\vec{B}\| \cos\theta$, onde $\| * \|$ mede a magnitude do vetor, $(\cdot)$ é o produto escalar,

Fonte: Elaborada pelo Autor

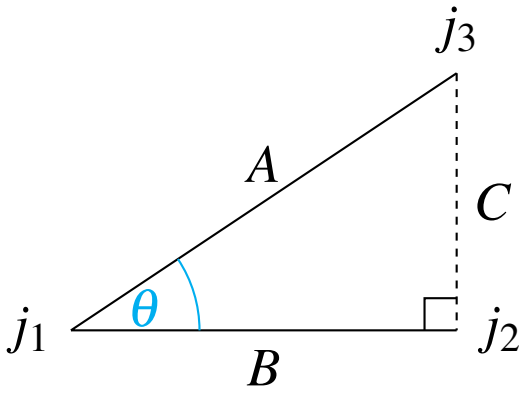


Figura 3.12: θ entre as juntas j_1 , j_2 e j_3

θ é dado pela seguinte equação:

$$\theta = \arccos\left(\frac{\vec{A}\cdot\vec{B}}{\|\vec{A}\|\,\|\vec{B}\|}\right) \tag{3.9}$$

Para termos a amplitude do movimento (AM) é necessário a conversão de θ pois o valor retornado é em radianos, para conversão de radianos em graus se dá pela seguinte equação:

$$AM = \theta \times \frac{2}{\pi} \tag{3.10}$$

3.0.4 Descritor do Gesto

Por meio da extração de características detectadas de origem da aquisição de gestos, é necessário representar as mesmas em um conjunto de características, no qual ira compor uma unidade de um gesto. Esta unidade G se torna uma classe de comparação para classificadores, em que cumpre a tarefa de comparação e reconhecimento do gesto, onde T é o tempo de execução do gesto, J a junta do corpo, JO junta objetiva e α ângulo do plano do corpo.

$$G = \left\{ \sum_{t=1}^{T-1} \left(\left(\sum_{j=1}^{J} \left(D_j^t, Da_j^t, \theta_j^t, \phi_j^t \right) \right), \left(\sum_{jo=1}^{JO} \left(AM_{jo}^t \right) \right), \left(\alpha^t \right) \right) \right\} \tag{3.11}$$

3.0.5 Classificador do Gesto

O classificador do gesto, tem a tarefa de dizer qual gesto é o mais correspondente comparado a um conjunto previamente de gestos gravados. Desta forma a unidade de gesto a ser comparado deve seguir o Descritor Gesto, assim como os gestos do conjunto. O classificador, através dos conjuntos de gestos previamente descritos, e com o gesto a ser comparado, deve dizer o coeficiente de similaridade para após classificar a unidade que é mais correspondente.

- *DTW:*

  Dynamic Time Warping é um algoritmo para medir a semelhança entre duas sequências temporais que podem variar em velocidade como já citado anteriormente. Em geral, DTW é um algoritmo que calcula uma combinação ideal entre duas sequências determinadas com certas restrições com o custo computacional $O(N^2)$, apresentamos o pseudo-código do DTW no Algoritmo 1.

Algoritmo 1 Dynamic Time Warping

```
1:  procedure DTWDistance(s : array[1..n], t : array[1..m])
2:      for for i := 1 to n do
3:          DTW[i,0] := infinity
4:      for for i := 1 to m do
5:          DTW[0,i] := infinity
6:      DTW[0,0] := 0
7:      for for i := 1 to n do
8:          for for j := 1 to m do
9:              cost := d(s[i],t[j])
10:             DTW[i, j] := cost + minimum(DTW[i−1, j],        ▷ insertion
11:             DTW[i, j−1],                                    ▷ deletion
12:             DTW[i−1, j−1])                                  ▷ match
13:     return DTW[n,m]
```

A fim de demonstrar Algoritmo 1, assumimos que temos duas sequências A e B, onde B é a sequência de teste a ser comparada com a amostra A.

$$\begin{aligned} A &= \{1,2,3,4,4,3,1\} \\ B &= \{1,2,3,4,3,1\} \end{aligned} \tag{3.12}$$

Através da Figura 3.13 é demonstrado, os valores plotados através do tempo (t).

Neste sentido definimos que ambas sequências são semelhantes na medida em que são únicas pois representam o mesmo valor semântico. No entanto, A é mais longo do que a sequência B, e a descida de valores é precedente no B. Com isto , temos que as duas

Fonte: Elaborada pelo Autor

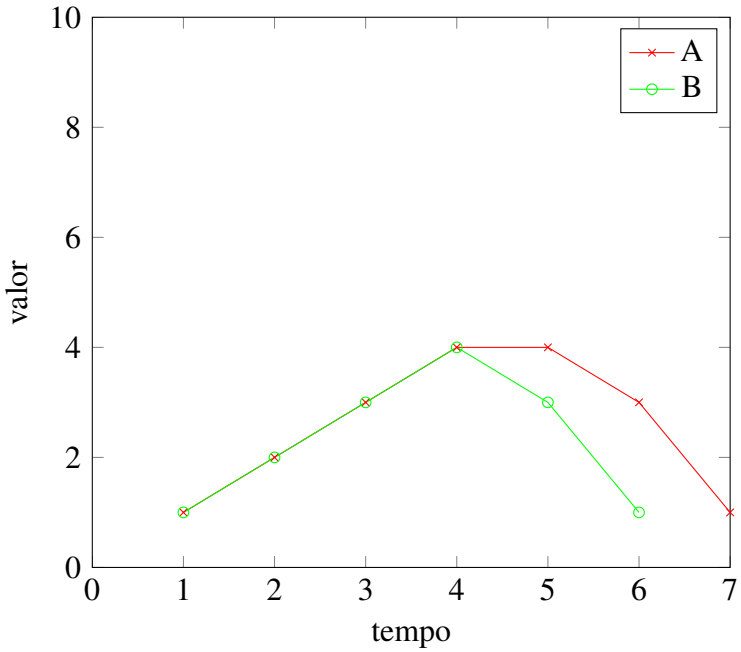


Figura 3.13: Valores de A e B no eixo x o tempo (t) e no eixo y o valor

Tabela 3.3: Matriz bidimensional de A e B

	0	1	2	3	4	4	3	1
0								
1								
2								
3								
4								
3								
1								

sequências não estão sincronizadas no tempo. Para descobrirmos a combinação ideal entre essas duas sequências. Em primeiro lugar, definimos a distância entre as duas sequências, $d = (x, y)$ onde x e y representam os dois pontos entre as sequências:

$$d = |x - y| \tag{3.13}$$

Através das duas sequências iremos demonstrar em uma matriz bidimensional disposto na Tabela 3.3, o uso do DTW. Em sequência devemos calcular as distâncias entre cada ponto de amostra com todos os pontos de teste, e encontrar a melhor combinação entre eles.

Seguindo o algoritmo, primeiramente teremos a primeira linha como infinita, o mesmo

Tabela 3.4: Distância entre A e B

	0	1	2	3	4	4	3	1
0	0	∞	∞	∞	∞	∞	∞	∞
1	∞	0	1	2	3	3	2	0
2	∞	1	0	1	2	2	1	1
3	∞	2	1	0	1	1	0	2
4	∞	3	2	1	0	0	1	3
3	∞	2	1	0	1	1	0	2
1	∞	0	1	2	3	3	2	0

Tabela 3.5: Distância ótima entre A e B

	0	1	2	3	4	4	3	1
0	0	∞	∞	∞	∞	∞	∞	∞
1	∞	0	1	3	6	9	11	11
2	∞	1	0	1	3	5	6	7
3	∞	3	1	0	1	2	2	4
4	∞	6	3	1	0	0	1	4
3	∞	8	4	1	1	1	0	2
1	∞	8	5	3	4	4	2	0

vale para a primeira coluna e a distância entre 0 e 0 é 0. Em sequência apresentamos a distância entre cada ponto em questão na Tabela 3.4.

Agora, para cada passo, consideramos a distância entre cada ponto em questão adicionando a distância mínima que encontramos. Isso nos dará a distância ótima de duas sequências na Tabela 3.5.

Agora, se voltarmos até o ponto de partida $(0,0)$, obtemos uma linha longa que pode se mover horizontalmente, verticalmente e diagonalmente na Tabela 3.6 apresentamos o movimento através da cor vermelha. Sendo assim cada movimento tem seu próprio significado no DTW:

- Um movimento horizontal representa a exclusão. Isso significa que nossa sequência de teste acelerou durante esse intervalo;
- Um movimento vertical representa a inserção. Isso significa que a sequência de teste desacelerou durante esse intervalo;
- Um movimento diagonal representa a correspondência. Durante este período, teste e amostra foram os mesmos.

Por fim temos que a distância máxima entre A e B é: 0, que obtemos no ponto $(7,8)$. Acrescentando a um classificador *Nearest Neighbor Classifier*, segundo Han, Pei e Kamber (2011), o algoritmo *k-Nearest Neighbor* (kNN) de aprendizagem supervisionada

Tabela 3.6: Movimento de *backtracking* das sequências *A* e *B*

	0	1	2	3	4	4	3	1
0	0	∞	∞	∞	∞	∞	∞	∞
1	∞	0	1	3	6	9	11	11
2	∞	1	0	1	3	5	6	7
3	∞	3	1	0	1	2	2	4
4	∞	6	3	1	0	0	1	4
3	∞	8	4	1	1	1	0	2
1	∞	8	5	3	4	4	2	0

através de uma medida de similaridade, os k exemplos mais próximos de um exemplo ainda não-rotulado e, baseado nos rótulos desse k exemplos próximos, rotular o novo exemplo, assim o vizinho mais próximo é unidade de gesto que desejamos encontrar. Através da distância do DTW comparamos com cada unidade de nosso conjunto de gestos, o gesto no qual tem o vizinho mais próximo é o gesto reconhecido.

- *FastDTW:*

  O algoritmo FastDTW foi introduzido por (SALVADOR; CHAN, 2007) e inicialmente foi projetado para diminuir o custo computacional do DTW para $O(n)$. O FastDTW basicamente consiste em dividir a complexidade do DTW padrão por meio de amostragem recursiva das séries temporais. O caminho de trajetória encontrado em cada iteração do algoritmo reduz a complexidade computacional, ao reduzir espacialmente a área manipulada pela programação dinâmica, a complexidade FastDTW é $O(n)$, e é conhecido por encontrar um caminho de trajetória de distância mínima preciso entre duas séries de tempo que é quase ideal a Figura 3.14 apresenta o algoritmo do FastDTW.

Fonte: Adaptada de (SALVADOR; CHAN, 2007)

Function FastDTW()

Input: *X* – a TimeSeries of length |*X*|
Y – a TimeSeries of length |*Y*|
radius – distance to search outside of the projected warp path from the previous resolution when refining the warp path

Output: 1) A min. distance warp path between *X* and *Y*
2) The warped path distance between *X* and *Y*

```
 1| // The min size of the coarsest resolution.
 2| Integer minTSsize = radius+2
 3|
 4| IF (|X|≤minTSsize OR |Y|≤minTSsize)
 5| {
 6|   // Base Case: for a very small time series run
 7|   //   the full DTW algorithm.
 8|   RETURN DTW(X, Y)
 9| }
10| ELSE
11| {
12|   // Recursive Case: Project the warp path from
13|   //   a coarser resolution onto the current
14|   //   current resolution. Run DTW only along
15|   //   the projected path (and also 'radius' cells
16|   //   from the projected path).
17|   TimeSeries shrunkX = X.reduceByHalf()
18|   TimeSeries shrunkY = Y.reduceByHalf()
19|
20|   WarpPath lowResPath =
21|       FastDTW(shrunkX,shrunkY, radius)
22|
23|   SearchWindow window =
24|       ExpandedResWindow(lowResPath, X, Y,
25|                          radius)
26|
27|   RETURN DTW(X, Y, window)
28| }
```

Figura 3.14: Algoritmo FastDTW

Capítulo 4

EXPERIMENTOS

Este capítulo aborda os experimentos e estudos de casos realizados a fim de validar este trabalho. Pretende-se, desta forma, validar as etapas do modelo proposto de reconhecimento de gestos por meio da ferramenta GGGesture. Apresentar os resultados na tarefa de reconhecimento de gesto do corpo, e aplicações em Casos de Uso onde se utiliza etapas do modelo apresentado.

4.1 Ferramenta GGGesture

Neste trabalho, é proposto a implementação do sistema computacional baseado no modelo de reconhecimento de gestos descrito na seção anterior em uma ferramenta (*software*) denominada GGGesture, esta seção tem como objetivo apresentar de forma detalhada a análise de requisitos, as tecnologias utilizadas para implementação, apresentação de alguns diagramas comportamentais da *Unified Modeling Language* (UML) bem como a modelagem da interação do usuário e os fluxos do sistema.

4.1.1 Análise de Requisitos

De acordo com os objetivos definidos e apresentados neste trabalho, é necessário realizar uma análise para se chegar de quais requisitos são importantes para a realização deste trabalho.

4.1.2 Requisitos Funcionais

Aqui são apresentados os requisitos que são considerados imprescindíveis para o concebimento da ferramenta proposta pelos objetivos deste trabalho

Deve possuir uma interface simples e intuitiva : A interface gráfica deve ser o mais simples, intuitiva e comunicativa o possível, com o objetivo de diminuir o esforço cognitivo do usuário quando o mesmo interagir com ela.

Fornecer *feedbacks* e dicas : O sistema, através de sua interface gráfica, deve fornecer mensagens de retorno (*feedbacks*) para o usuário sempre que possível. Essas mensagens podem ser mensagens de erro, de sucesso, de confirmação ou até mesmo dicas sobre o que o usuário pode realizar em um determinado momento. E importante salientar que a utilização de tais mensagens deve ser feita apenas quando necessário, ou seja, quando os recursos da interface por si só não são suficientes para indicar a mesma mensagem ao usuário.

Suportar ações básicas: Essas ações (ou funcionalidades) básicas devem ser ações referentes às funcionalidades de um reconhecedor de gestos

Exemplos comum de ações:

- Gravar Gesto
- Reconhecer Gestos

Deve suportar visualizações de informações Essas visualizações devem ser referentes ao reconhecimento de gestos;

Exemplos de visualizações comuns ao contexto de reconhecimento:

- Mostrar qual a porcentagem de classificação

Exemplos de visualizações comuns ao contexto de amplitude de movimento:

- Mostrar a quantidade em graus das juntas

4.1.3 Tecnologias de Suporte

No desenvolvimento da ferramenta utilizou-se a linguagem de programação Java, por ser uma linguagem multiplataforma (Linux, Mac OS X e Windows). Também se utilizou do *wrapper* da biblioteca OpenNI/NITE para Java.

4.1.4 Diagrama de Caso de Uso do Sistema

Para facilitar o entendimento dos requisitos do sistema foi desenvolvido o diagrama de Caso de Uso apresentado na Figura 4.1. Observa-se que existem dois atores, o Usuário e Sistema de

Arquivos. O Usuário, age com o subsistema de aplicação, sensor de profundidade referência PrimeSense o Microsoft Kinect, e reconhecimento de gestos. O Sistema de Arquivos, age com o subsistema de Reconhecimento de Gestos e Recuperação de Arquivos.

O subsistema de Aplicação, consiste em ter a interface para iniciar a aplicação que incluem a visualização e captura dos quadros do sensor, também é possível o ajuste das configurações do sensor no qual é acionado pelo ator Usuário.

O subsistema sensor Kinect, consiste em operar os componentes no qual conecta o sensor e muda o estado do sensor para captura das informações necessárias para o subsistema de Aplicação e também é acionado pelo ator Usuário. O subsistema de Reconhecimento de Gestos e Amplitude do Movimento, consiste em ter os componentes no qual permite as funcionalidades de análise ou treinamento através do modelo de reconhecimento de gestos implementado. O ator Usuário pode gravar seus movimentos, analisar os gestos e obter a análise da amplitude do movimento.

Durante a execução dos gestos através análise de gestos e gravação dos movimentos é extraído as características estáticas e dinâmicas das juntas do corpo, normalização das juntas e descrição do gesto. Na análise dos gestos é feito o reconhecimento dos gestos, através das informações do classificador de gestos de origem do subsistema de Recuperação de Arquivos. O subsistema de Recuperação de Arquivos, consiste em componentes que recuperam os gestos descritos salvos no sistema de arquivos e também classificam os gestos salvos anteriormente.

O subsistema de Aplicação, consiste em ter a interface para iniciar a aplicação que incluem a visualização e captura dos quadros do sensor, também é possível o ajuste das configurações do sensor no qual é acionado pelo ator Usuário. O subsistema sensor Kinect, consiste em operar os componentes no qual conecta o sensor e muda o estado do sensor para captura das informações necessárias para o subsistema de Aplicação e também é acionado pelo ator Usuário.

O subsistema de Reconhecimento de Gestos e Amplitude do Movimento, consiste em ter os componentes no qual permite as funcionalidades de análise ou treinamento através do modelo de reconhecimento de gestos implementado. O ator Usuário pode gravar seus movimentos, analisar os gestos e obter a análise da amplitude do movimento.

Durante a execução dos gestos através análise de gestos e gravação dos movimentos é extraído as características estáticas e dinâmicas das juntas do corpo, normalização das juntas e descrição do gesto. Na análise dos gestos é feito o reconhecimento dos gestos através das informações do classificador de gestos de origem do subsistema de Recuperação de Arquivos. O subsistema de Recuperação de Arquivos, consiste em componentes que recuperam os gestos

Fonte: Elaborada pelo autor

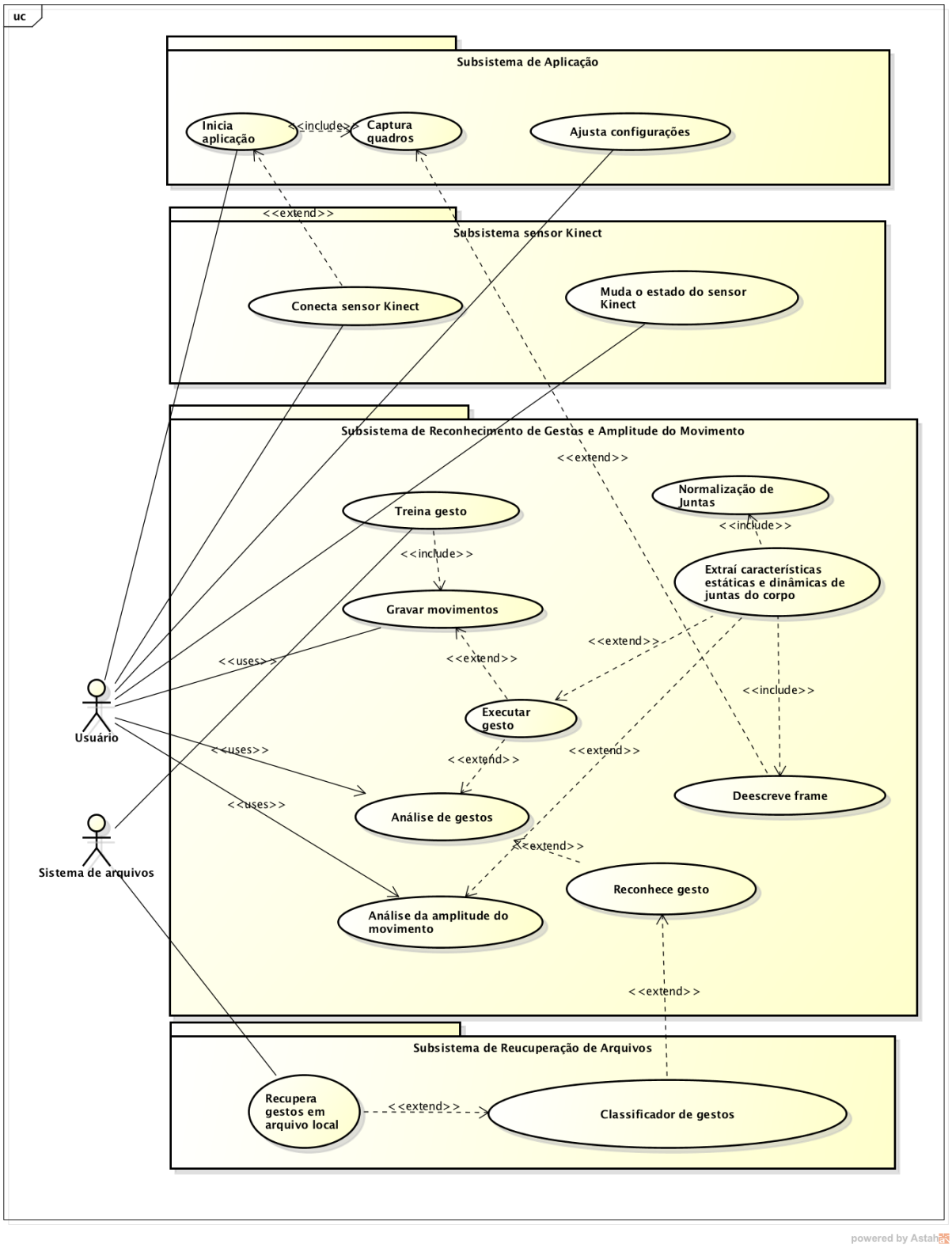


Figura 4.1: Diagrama de Caso de Uso do Sistema

descritos salvos no sistema de arquivos e também classificam os gestos salvos anteriormente.

4.1.5 Diagrama de Atividades do Sistema

O diagrama de atividades faz parte do UML. O principal objetivo do diagrama de atividades é modelar o comportamento do sistema (através da definição de caminhos lógicos que um processo pode seguir) a partir de representações gráficas. O diagrama está apresentado na Figura 4.2. O sistema possui um fluxo interno de atividades que desejamos ir de encontro ao objetivo do reconhecimento de gestos, através do diagrama de atividades conseguimos mapear o fluxo.

O estado inicial é requerer imagens de profundidade do dispositivo Microsoft Kinect, no qual as imagens são a base para nosso modelo reconhecimento de gestos, após é exibido a imagem e o usuário segmentado.

Em sequência temos a condição de aguardo do usuário para utilização da aplicação, durante o uso ele possui dois estados disponíveis o de treinar ou analisar.

No estado de Treina gesto o usuário faz o treinamento do gesto para um reconhecimento futuro, logo na sequência temos o estado onde-se obtém as juntas do corpo e amplitude do movimento, e por diante as normaliza. Tendo as informações necessárias para compor o modelo é feita a descrição do mesmo e por fim o modelo é gravado no dicionário de gestos.

Seguindo através do estado de Analise Gesto após a descrição do modelo se obtém todos os modelos de gestos previamente gravados e executa o classificador, caso não nenhum gesto é classificado volta se ao estado de análise de gesto, caso o gesto seja classificado e reconhecido como gesto válido e finaliza o fluxo de atividades.

Fonte: Elaborada pelo autor

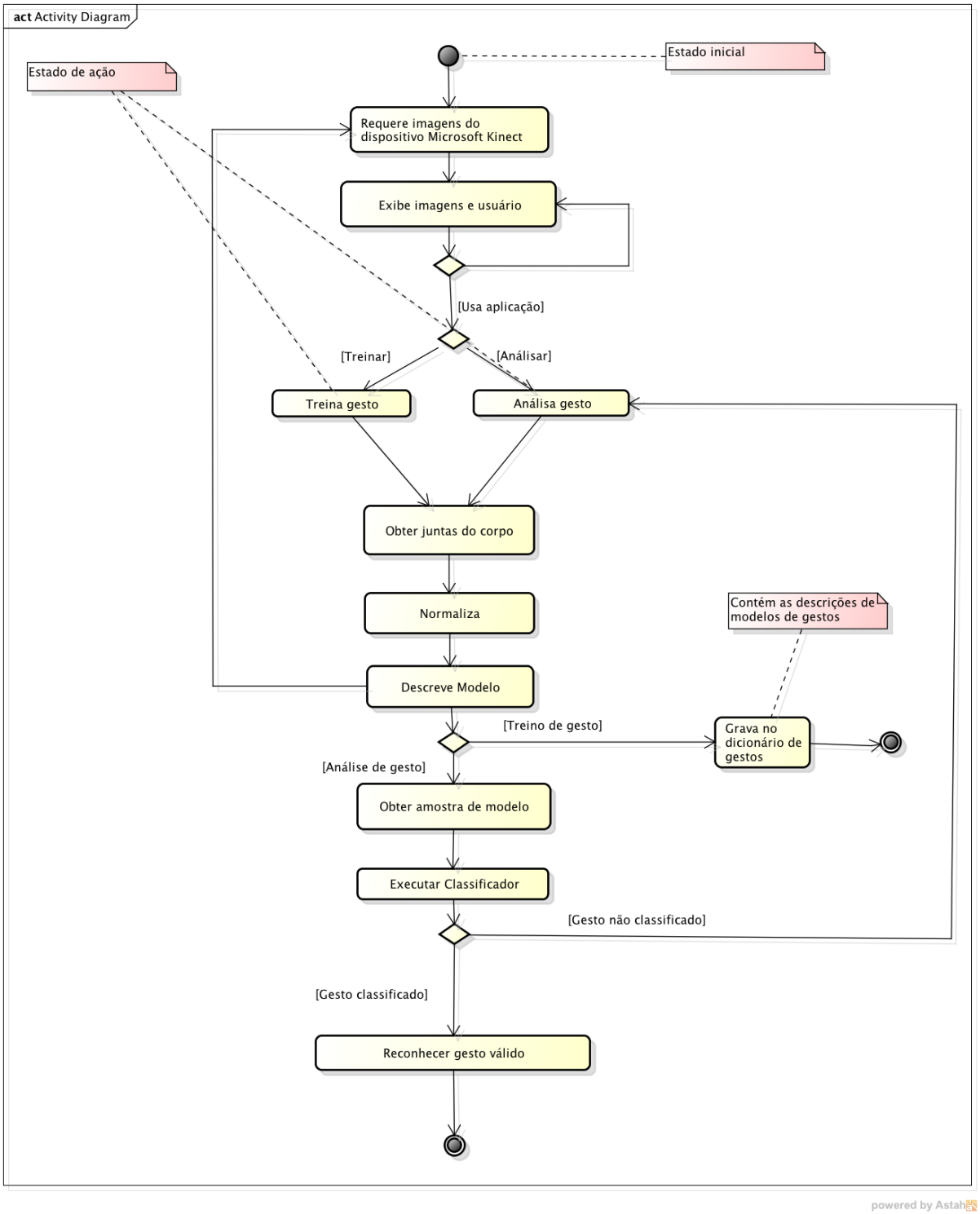


Figura 4.2: Diagrama de atividades

4.1.6 Elementos do GGGesture

Os elementos implementados possuem características próprias e estão separados, no entanto interligados para o objetivo comum. No intuito de melhor distribuição de características e tarefas do sistema os mesmos foram separados, os principais elementos que integram o GGGesture são:

- FileSystem
  - Descriptor
  - Persistence
- GUI
- User
  - UserTracker
  - Normalize
  - Cartesian
- FastDTW
- DTW
- Formato GGGesture

O formato do GGGesture é representação do gesto descrito para persistência no sistema de arquivos do sistema operacional. Ele é composto por um arquivo de extensão *Comma-separated values* (CSV) , no qual os dados das juntas normalizadas, ou não, são concatenados e separadas por uma virgula em uma linha, cada linha representa um único quadro no espaço no tempo.

O FileSystem, possui componentes que armazenam as descrições de gestos que serão utilizados para comparação com os dados de entrada através do Descriptor e Persistence. Os dados de entrada devem seguir o formato do GGGesture, especificado anteriormente. É suportado o armazenamento de quantos segundos de dados o usuário escolher, cuja taxa de transmissão pode ser de 30 quadros por segundo ou 1 quadro por segundo. Embora os dados a serem processados sejam normalmente gravados através do Persistence, não há uma dependência direta, sendo possível receber os dados de qualquer outro local. Assim facilita, por exemplo, a criação de casos de testes, que podem ser realizados com dados capturados do sensor e armazenados diretamente.

A GUI, é composto por uma interface gráfica, apresentada na Figura 4.3 onde o usuário tem as seguintes funcionalidades:

Fonte: Elaborada pelo autor

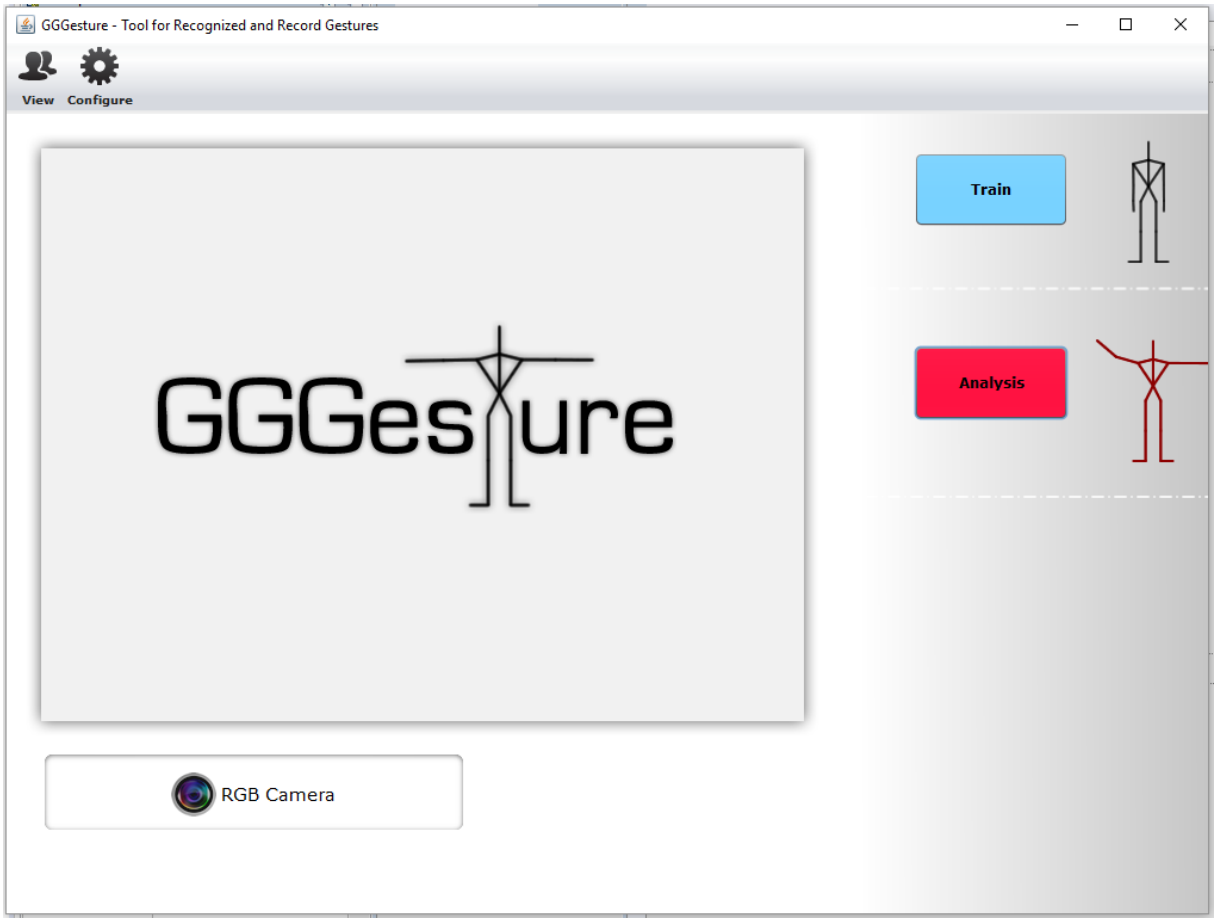


Figura 4.3: GUI - Interface Gráfica da Ferramenta GGGesture

- Configure - neste painel o usuário pode alterar os dados gerais da OpenNI apresentado nas Figuras 4.4 e 4.5.
- View – exibe o painel de visualização, treinamento e análise apresentado na Figura 4.3.
- Visualização da Câmera – exibe a câmera do sensor
- Visualização do Esqueleto e Amplitude do Movimento - exibe as juntas em formato de esqueleto no qual facilita o usuário na gravação
- Train - permite ao usuário gravar as propriedades de determinado gesto nome e intervalo de tempo.
- Analysis - permite ao usuário analisar e comparar todos os gestos e verificar qual o gesto do dicionário que possui mais similaridade.

Fonte: Elaborada pelo autor

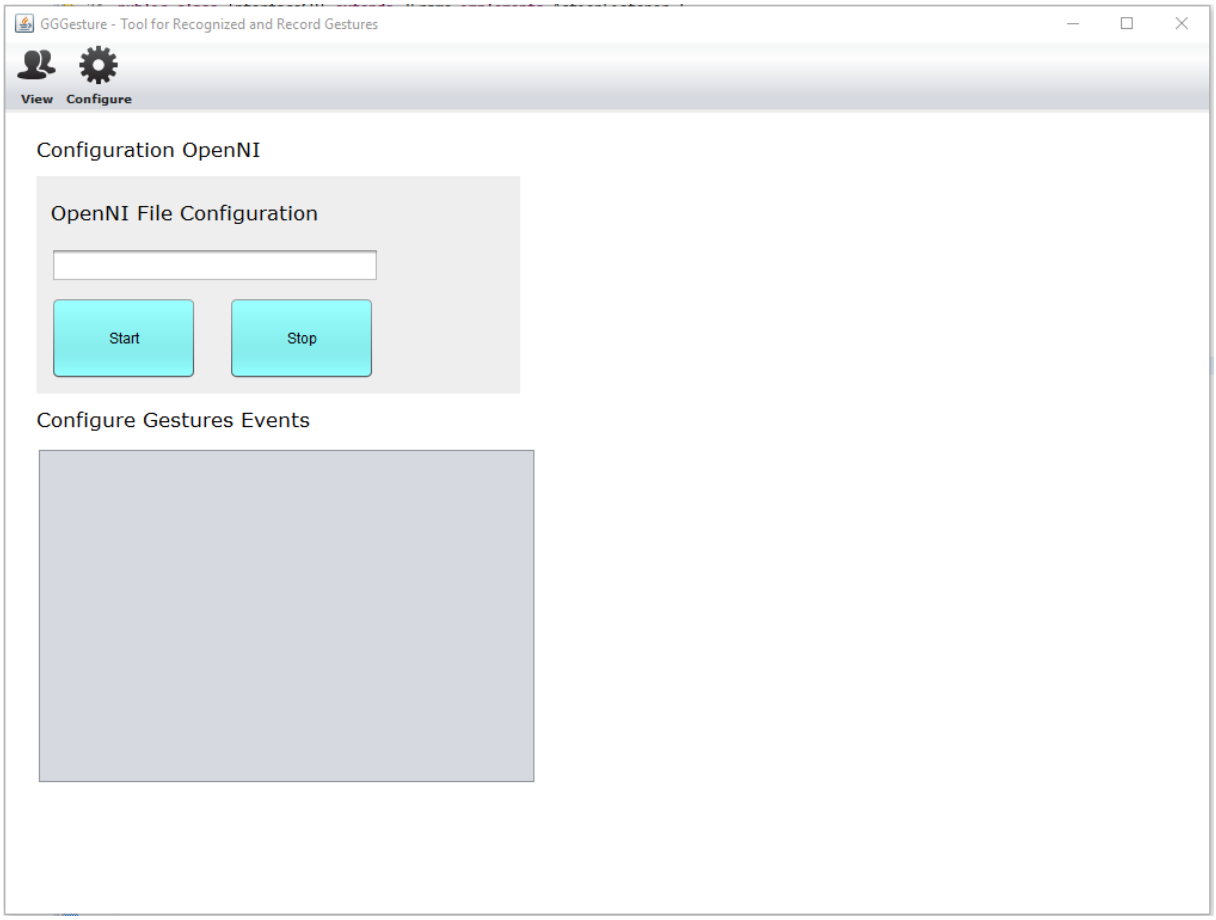


Figura 4.4: Configure - Funcionalidade da GUI

Fonte: Elaborada pelo autor

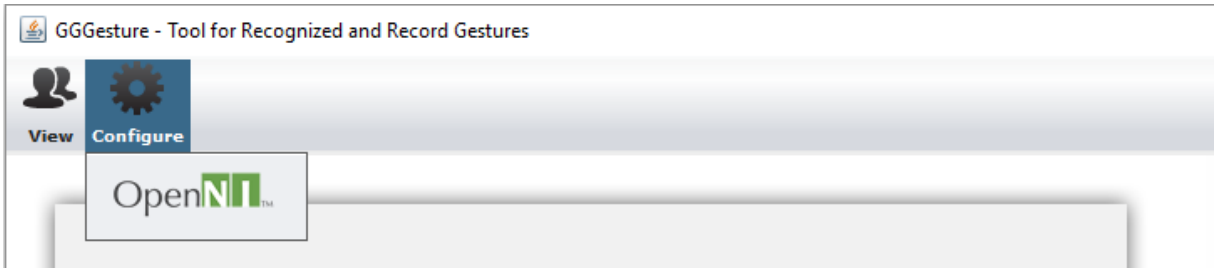


Figura 4.5: Configure Menu - Funcionalidade da GUI

O User, possuí componentes no qual é capturado, normalizado e armazenado as juntas do corpo em memória. O UserTracker tem a funcionalidade de capturar em tempo real as juntas do corpo, enquanto o Normalize após a captura pelo UserTracker normaliza as juntas conforme o modelo de reconhecimento de gestos expostos anteriormente, o Cartesian armazena as juntas no plano cartesiano plano de origem da biblioteca OpenNI. O FastDTW e DTW são

os componentes em que se implementa os respectivos algoritmos.

4.2 Experimentos GGGesture

Nesta subseção a precisão do sistema implementado na ferramenta GGGesture é analisado, a configuração e abordagem do modelo de reconhecimento de gestos são parâmetros para a avaliação. Utilizamos um conjunto finito de 14 gestos e movimentos para os experimentos, os gestos e movimentos enumerados para identificação são:

1. Levantar os dois braços (Abdução de 0 a 180º.)
2. Levantar o braço direito (Abdução de 0 a 180º.)
3. Levantar o braço esquerdo (Abdução de 0 a 180º.)
4. Levantar os dois braços (Abdução de 0 a 90º.)
5. Levantar o braço direito (Abdução de 0 a 90º.)
6. Levantar o braço esquerdo (Abdução de 0 a 90º.)
7. Flexão do cotovelo braço esquerdo e direito (Flexão de cotovelo de 0 a 90º)
8. Flexão do cotovelo braço esquerdo (Flexão de cotovelo de 0 a 90º)
9. Flexão do cotovelo braço direito (Flexão de cotovelo de 0 a 90º)
10. Bater palmas
11. Tchau com a mão direita
12. Tchau com a mão esquerda
13. Agachar
14. Levantar

Os treinamentos dos gestos são realizados através da ferramenta GGGesture, no mesmo ambiente e por 6 usuários diferentes, como demonstra as Figuras 4.6, 4.7, 4.8, 4.9, 4.10 e 4.11

Fonte: Elaborada pelo Autor

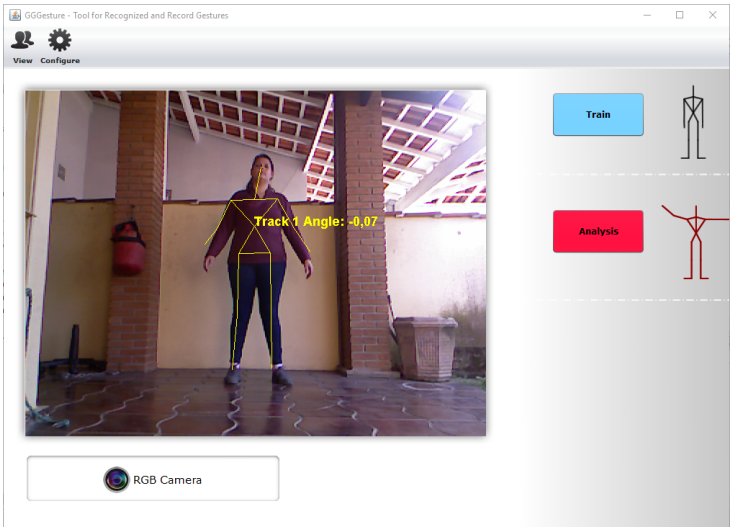


Figura 4.6: Treinamento Usuário

Fonte: Elaborada pelo Autor

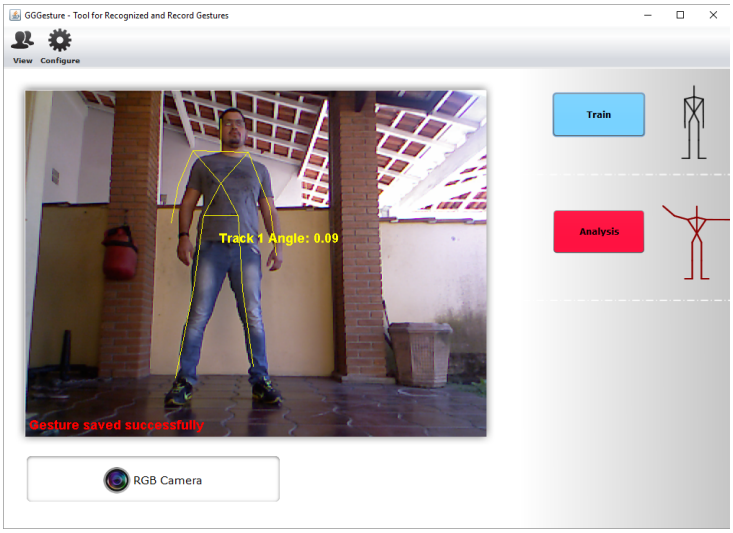


Figura 4.7: Treinamento Usuário

Fonte: Elaborada pelo Autor

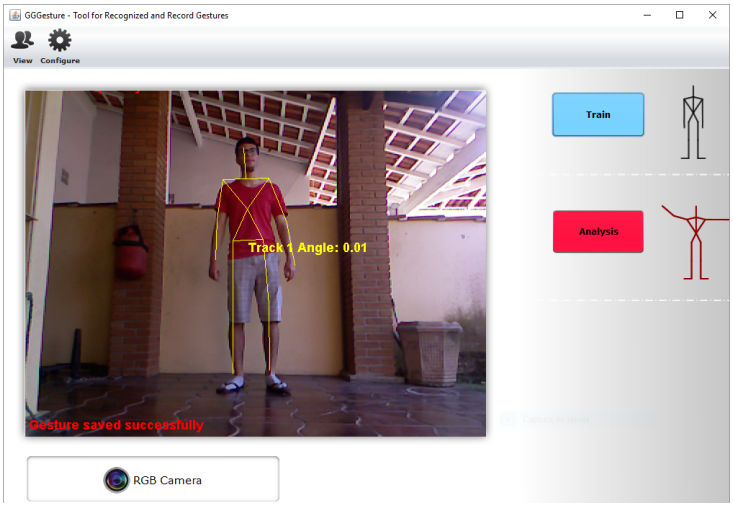


Figura 4.8: Treinamento Usuário

Fonte: Elaborada pelo Autor

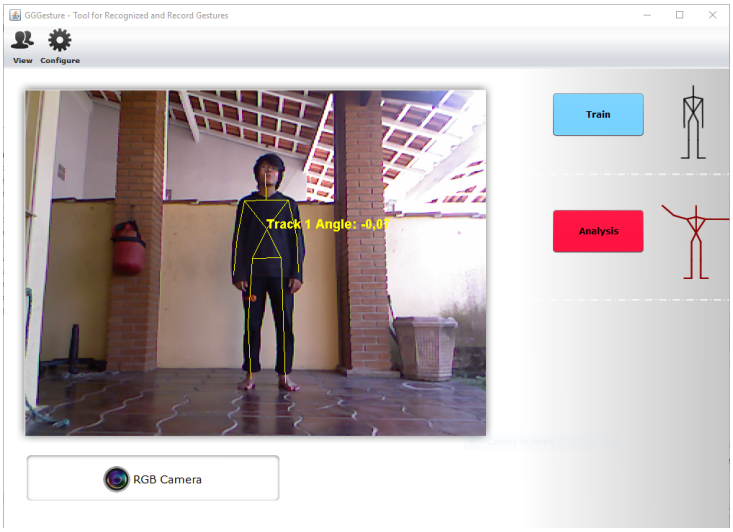


Figura 4.9: Treinamento Usuário

Fonte: Elaborada pelo Autor

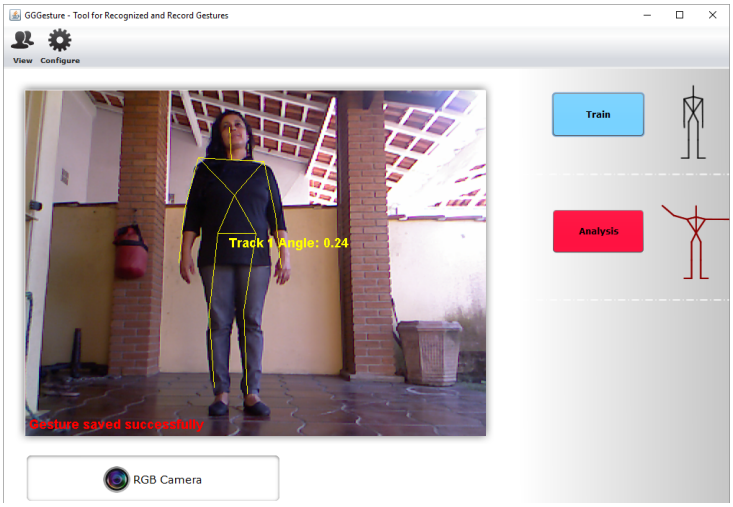


Figura 4.10: Treinamento Usuário

Fonte: Elaborada pelo Autor

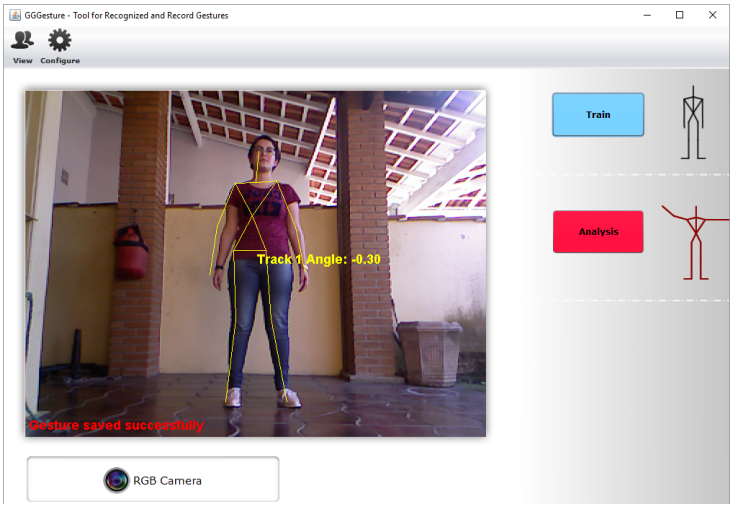


Figura 4.11: Treinamento Usuário

Primeiramente, é gravado as amostras dos 14 gestos que serão base para comparação, todas essas amostras são gravadas pelo mesmo usuário. Na sequência o conjunto de gestos são gravados por 5 pessoas diferentes, em que será comparado com amostra de gestos do primeiro usuário.

As abordagens para avaliação e validação do modelo de reconhecimento de gestos se dá através de classificadores utilizando descritores com juntas normalizadas e não normalizadas comparadas através do algoritmo FastDTW com as distâncias Euclidiana e Manhattan, e ainda

comparando 30 quadros por segundos e 1 quadro por segundo, implementados no componente de Persistence.

Durante a execução do treinamento, foi detectado algumas falhas na detecção e rastreamento das juntas de origem da OpenNI, é observado a ocorrência em virtude da iluminação e o fundo com regiões de grande e baixa profundidade que podem ter influenciado no sensor e no algoritmo de *Skeletal*, mas o treinamento foi prosseguido para discussão e análise. A Tabela 4.1, demonstra os resultados somados de maneira geral :

Tabela 4.1: Resultados de reconhecimento através de todas as abordagens para cada usuário e gesto

Usuário / Gesto	1	2	3	4	5	Total Reconhecido
1	6	12	0	2	0	20
2	6	0	3	6	6	21
3	0	6	6	0	0	12
4	2	0	0	6	0	8
5	8	11	0	2	0	21
6	6	9	3	0	15	33
7	0	0	0	0	0	0
8	2	2	0	4	2	10
9	0	0	0	0	6	6
10	0	0	0	0	0	0
11	0	0	0	0	0	0
12	0	0	6	0	0	6
13	0	0	8	2	0	10
14	0	0	0	0	0	0

Nota-se que alguns gestos com zero de reconhecimento em todas as abordagens (7,10,11 e 14), no caso do gesto 10 (Bater palmas) que não se obteve nenhum reconhecimento, isto ocorre devido a oclusão das juntas durante o movimento, assim o rastreamento e obtenção das juntas é perdido em alguns instantes do tempo.

Enquanto o 11 e 12 (Tchau com mãos diferentes), o baixo reconhecimento ocorre-se devido ao fato de que alguns usuários empenharem o movimento com a mão parada, e outros com a mão em movimento junto com o braço.

O gesto 14, o movimento de levantar a partir da posição de agachamento, o baixo reconhecimento deve-se ao fato das coordenadas se perderem muito neste movimento, alguns usuários quando se agachavam-se, as juntas dos mesmos se perdiam por conta dos joelhos que se aproximavam, e perdiam a forma de silhueta do corpo, silhueta no qual o algoritmo *Skeletal* estima as juntas.

No gesto 7, evidenciamos que em alguns treinamentos o reconhecimento das juntas se perdia, devido a parte superior da câmera existia uma profundidade maior devido ao local que foi gravado, com isto a captura de profundidade pode afetar o algoritmo *Skeletal*.

Os demais gestos, em todas as abordagens só se obteve o reconhecimento por um usuário dentre os 5 usuários, os gestos 9 e 12. Mas o gesto 9 também se observou a perda das juntas, durante o treinamento.

Na Tabela 4.2, apresentamos o reconhecimento utilizando o algoritmo FastDTW, com a distância Euclidiana, com todas as juntas do corpo não normalizadas, ou seja, utilizando as coordenadas (x,y,z) de origem do sensor com apenas 1 quadro por segundo, observa-se que foram reconhecidos no total 3 gestos.

Tabela 4.2: Distância Euclidiana de todas as juntas não normalizadas 1 quadro por segundo.

Gesto	Qtd. Usuários
1	1
2	0
3	0
4	0
5	2
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0

Na Tabela 4.3, apresentamos o reconhecimento utilizando o algoritmo FastDTW, com a distância Euclidiana, com todas as juntas superiores do corpo não normalizadas, ou seja, utilizando as coordenadas (x,y,z) de origem do sensor com apenas 1 quadro por segundo, observa-se que foram reconhecidos no total 7 gestos. Outro ponto observado é uma diferença relacionado a primeira comparação, pois em alguns gestos definidos as juntas que tange o movimento são somente as dos membros superiores, e descartando as demais que prejudicam a comparação.

Tabela 4.3: Distância Euclidiana de todas as juntas superiores não normalizadas 1 quadro por segundo.

Gesto	Qtd. Usuários
1	1
2	0
3	0
4	0
5	0
6	0
7	0
8	4
9	1
10	0
11	0
12	0
13	1
14	0

Na Tabela 4.4, apresentamos o reconhecimento utilizando o algoritmo FastDTW, com a distância Euclidiana, com todas as juntas do corpo não normalizadas, ou seja, utilizando as coordenadas (x,y,z) de origem do sensor com 30 quadros por segundo, observa-se que foram reconhecidos no total 11 gestos, um acréscimo na quantidade de reconhecimento através do aumento para 30 quadros por segundo, pois há mais informações em 30 quadros, assim ampliando a consistência das características do movimento.

Tabela 4.4: Distância Euclidiana de todas as juntas não normalizadas 30 quadros por segundo.

Gesto	Qtd. Usuários
1	1
2	3
3	0
4	0
5	1
6	3
7	0
8	0
9	1
10	0
11	0
12	0
13	2
14	0

Na Tabela 4.5, apresentamos o reconhecimento utilizando o algoritmo FastDTW, com a dis-

tância Euclidiana, com todas as juntas superiores do corpo não normalizadas, ou seja utilizando as coordenadas (x, y, z) de origem do sensor com 30 quadros por segundo, observa-se que foram reconhecidos no total 9 gestos, um declínio na quantidade devido ao gesto 13 e 9, o gesto 13 utiliza-se todas as juntas do corpo, no entanto foram gravados só as juntas dos membros superior, enquanto o gesto 9 durante o treinamento houveram perdas nas características discutidas anteriormente.

Tabela 4.5: Distância Euclidiana de todas as juntas superiores não normalizadas 30 quadros por segundo.

Gesto	Qtd. Usuários
1	1
2	3
3	0
4	0
5	1
6	3
7	0
8	0
9	0
10	0
11	0
12	0
13	1
14	0

Na Tabela 4.6, apresentamos o reconhecimento utilizando o algoritmo FastDTW, com a distância Euclidiana, com todas as juntas do corpo normalizadas, com 1 quadro por segundo, observa-se que foram reconhecidos no total 4 gestos.

Tabela 4.6: Distância Euclidiana de todas as juntas normalizadas 1 quadro por segundo.

Gesto	Qtd. Usuários
1	0
2	0
3	2
4	0
5	1
6	1
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0

Na Tabela 4.7, apresentamos o reconhecimento utilizando o algoritmo FastDTW, com a distância Euclidiana, com todas as juntas superiores do corpo normalizadas, com 1 quadro por segundo, observa-se que foram reconhecidos no total 4 gestos, a mesma quantidade anterior.

Tabela 4.7: Distância Euclidiana de todas as juntas superiores normalizadas 1 quadro por segundo.

Gesto	Qtd. Usuários
1	0
2	0
3	2
4	0
5	1
6	1
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0

Na Tabela 4.8, apresentamos o reconhecimento utilizando o algoritmo FastDTW, com a distância Euclidiana, com todas as juntas do corpo normalizadas, com 30 quadros por segundos, observa-se que foram reconhecidos no total 6 gestos, um acréscimo na quantidade de reconhecimento através do aumento para 30 quadros por segundo, pois há mais informações em 30 quadros, assim ampliando a consistência das características do movimento.

Tabela 4.8: Distância Euclidiana de todas as juntas normalizadas 30 quadros por segundo.

Gesto	Qtd. Usuários
1	1
2	0
3	0
4	1
5	1
6	2
7	0
8	0
9	0
10	0
11	0
12	1
13	0
14	0

Na Tabela 4.9, apresentamos o reconhecimento utilizando o algoritmo FastDTW, com a distância Euclidiana, com todas as juntas superiores do corpo normalizadas, com 30 quadros por segundo, observa-se que foram reconhecidos no total 6 gestos.

Tabela 4.9: Distância Euclidiana de todas as juntas superiores normalizadas 30 quadros por segundo.

Gesto	Qtd. Usuários
1	1
2	0
3	0
4	1
5	1
6	2
7	0
8	0
9	0
10	0
11	0
12	1
13	0
14	0

Nesta nova abordagem utilizamos a distância de Manhattan junto ao algoritmo FastDTW para comparar seus efeitos, na Tabela 4.10 demonstra-se a comparação dos gestos de todas as juntas não normalizadas, com 1 quadro por segundo, no qual foram reconhecidos 5 gestos. Inicialmente observa-se uma diferença de 2 gestos a mais reconhecidos em comparação com a

distância Euclidiana.

Tabela 4.10: Distância Manhattan de todas as juntas não normalizadas 1 quadro por segundo.

Gesto	Qtd. Usuários
1	2
2	0
3	0
4	1
5	1
6	1
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0

Na Tabela 4.11, demonstra-se a comparação dos gestos de todas as juntas superiores do corpo não normalizadas com 1 quadro por segundo no qual foram reconhecidos 7 gestos. Não se observa uma diferença de quantidade comparado a distância Euclidiana, porem se mantém a influência das juntas que tange o movimento.

Tabela 4.11: Distância Manhattan de todas as juntas superiores não normalizadas 1 quadro por segundo.

Gesto	Qtd. Usuários
1	1
2	0
3	0
4	0
5	0
6	1
7	0
8	3
9	1
10	0
11	0
12	0
13	1
14	0

Na Tabela 4.12, demonstra-se a comparação dos gestos de todas as juntas superiores do corpo não normalizadas com 30 quadros por segundo no qual foram reconhecidos 9 gestos.

Uma diferença na qual a quantidade comparada a distância Euclidiana é reduzida, os gestos no qual influenciam este valor são 5,6 e 13.

Tabela 4.12: Distância Manhattan de todas as juntas não normalizadas 30 quadros por segundo.

Gesto	Qtd. Usuários
1	1
2	4
3	0
4	0
5	0
6	1
7	0
8	1
9	1
10	0
11	0
12	0
13	1
14	0

Na Tabela 4.13, demonstra-se a comparação dos gestos de todas as juntas superiores do corpo não normalizadas com 30 quadros por segundo, no qual foram reconhecidos 8 gestos. Se observasse uma diferença de quantidade de decréscimo de apenas 1 gesto comparado a distância Euclidiana, porem os gestos reconhecidos de mantém ou se diferenciam em alguns casos.

Tabela 4.13: Distância Manhattan de todas as juntas superiores não normalizadas 30 quadros por segundo.

Gesto	Qtd. Usuários
1	1
2	4
3	0
4	0
5	0
6	1
7	0
8	1
9	0
10	0
11	0
12	0
13	1
14	0

Na Tabela 4.14, demonstra-se a comparação dos gestos de todas as juntas do corpo norma-

lizadas com 1 quadro por segundo no qual foram reconhecidos 4 gestos. Sem alteração com a quantidade na distância Euclidiana, devido à pouca variação de valores após a normalização.

Tabela 4.14: Distância Manhattan de todas as juntas normalizadas 1 quadro por segundo.

Gesto	Qtd. Usuários
1	0
2	0
3	2
4	0
5	1
6	1
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0

Na Tabela 4.15, demonstra-se a comparação dos gestos de todas as juntas superiores do corpo normalizadas com 1 quadro por segundo no qual foram reconhecidos 4 gestos. Sem alteração com a quantidade na distância Euclidiana, devido à pouca variação de valores após a normalização.

Tabela 4.15: Distância Manhattan de todas as juntas superiores normalizadas 1 quadro por segundo.

Gesto	Qtd. Usuários
1	0
2	0
3	2
4	0
5	1
6	1
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0

Na Tabela 4.16, demonstra-se a comparação dos gestos de todas as juntas do corpo norma-

lizadas com 30 quadros por segundo, no qual foram reconhecidos 6 gestos. Sem alteração com a quantidade na distância Euclidiana, devido à pouca variação de valores após a normalização.

Tabela 4.16: Distância Manhattan de todas as juntas normalizadas 30 quadros por segundo.

Gesto	Qtd. Usuários
1	1
2	0
3	0
4	1
5	1
6	2
7	0
8	0
9	0
10	0
11	0
12	1
13	0
14	0

Na Tabela 4.17, demonstra-se a comparação dos gestos de todas as juntas superiores do corpo normalizadas com 30 quadros por segundo, no qual foram reconhecidos 6 gestos. Sem alteração com a quantidade na distância Euclidiana, devido à pouca variação de valores após a normalização.

Tabela 4.17: Distância Manhattan de todas as juntas superiores normalizadas 30 quadros por segundo.

Gesto	Qtd. Usuários
1	1
2	0
3	0
4	1
5	1
6	2
7	0
8	0
9	0
10	0
11	0
12	1
13	0
14	0

Nestes experimentos em comparação nas distâncias Euclidiana e Manhattan foram observa-

das diferenças nas coordenadas não normalizadas, a distância de Manhattan é tipicamente mais tolerante a *outliers* que resolve em casos em que as coordenadas (x, y, z) variam muito devido a instabilidades do rastreamento.

Também se notou que dentre os 14 gestos treinados, apenas 8 de fato se tornaram relevantes os quais são (1,2,3,4,5,6,8 e 13), assim excluindo os gestos que tiveram erros de produção (9,10,11 e 14) discutidos anteriormente.

Demonstra-se através da Tabela 4.18, a matriz apenas considerando a soma dos gestos reconhecidos, dos 8 gestos relevantes. Se demonstrou que 30 quadros é de fato mais vantajoso para o reconhecimento de gestos, pois se obteve uma alta taxa de reconhecimento. Também foi observado que a escolha das juntas de alvo, deve ser uma característica na ferramenta a se considerar. Provamos que após a normalização das juntas, as distâncias Euclidiana e Manhattan não afetam o reconhecimento, mas em juntas não normalizadas há uma variância significativa, onde a distância de Manhattan se torna vantajosa.

Demonstra-se também, que a distância em que mais se reconheceu foi a Euclidiana, porém a normalização das juntas não se destacou em comparação as juntas não normalizadas, acreditamos que devido aos erros de medição que ocorreram durante o treinamento afetou-se a normalização, a normalização possui muitos cálculos e seus *inputs* devem estar em constância para uma melhor transformação continua a fim de se obter uma melhor descrição.

Tabela 4.18: Matriz dos 8 gestos relevantes

	Todas juntas[a]	Juntas superiores[a]	Todas juntas[b]	Juntas superiores[b]	Total
JNNE	3	6	10	9	28
JNNM	5	6	8	8	27
JNE	4	4	5	5	18
JNM	4	4	5	5	18
Total	16	20	28	27	91

JNNE = Juntas não normalizadas - Euclidiana; JNNM = Juntas não normalizadas - Manhattan; JNE = Juntas normalizadas - Euclidiana; e JNM = Juntas normalizadas - Manhattan.

[a] 1 quadro por segundo.

[b] 30 quadros por segundo.

Por fim, conforme é discutido na Produção do Gesto, demonstra-se que alguns erros podem ser gerados durante a modelagem do gesto como o erro de empenho e erro de medição, durante os experimentos deste trabalho notamos os dois tipos, neste sentido é interessante uma melhor análise para correção, ou mesmo ações que evitam estes tipos de erros.

Alguns pontos a se discutir é a amostra principal, no qual foi base de comparação no reconhecimento do gesto, durante o treinamento desta a amostra notou se erros de medição, ruídos onde as características dos gestos podem ter sido afetadas, talvez uma análise previa nas informações, um validador de características, ou mesmo na ferramenta de captura acrescentar um bloqueio no treinamento se durante o mesmo houve algum ruído. Outro ponto observado para uma melhor comparação, antes do treinamento definir as juntas de alvo e pesos que correspondem melhor ao gesto. Além de reduzir as características para um melhor desempenho, também se evita de juntas que são indiferentes ao movimento, sejam compostas na comparação.

4.3 Casos de Uso

Com o intuito de demonstrar a utilidade e contribuição deste trabalho, é demonstrado também o uso e aplicação das etapas do modelo de reconhecimento gestos em aplicações desenvolvidas em Casos de Uso ao longo deste trabalho.

4.3.1 Comitê de Ética

O desenvolvimento deste trabalho foi aprovado pelo Comitê de Ética em Pesquisa com Seres Humanos / UFSCar - CAAE 11319712.4.0000.5504 e apoiado pelo Centro de Ciências Exatas e de Tecnologia/UFSCar.

O presente estudo teve o parecer favorável pelo Comitê de Ética (CEP/UFSCar) pelo Nº 182.271/2013 para respeitar e cumprir as prerrogativas da resolução 196/96 da Comissão Nacional de Ética em Pesquisa (CONEP) que versa sobre ética em pesquisa com seres humanos.

Os participantes deste trabalho receberam e assinaram o termo de consentimento livre e esclarecido (TCLE) com todas as informações sobre o projeto, como: objetivo, procedimento da coleta de dados, resguardo da privacidade do participante e utilização dos dados para fins científicos.

4.3.2 A aplicação e seus Casos de Uso | I

A primeira aplicação se trata de uma ferramenta computacional denominada *RehabGesture* para avaliar a amplitude de movimento dos membros superiores no plano coronal e reconhecer gestos. O software representa uma alternativa de baixo custo à análise da amplitude de movimento dos membros superiores no plano coronal, estimula o processo de ensino-aprendizagem nas disciplinas relativas ao estudo do movimento humano e complementa o processo de reabilitação do membro superior de forma lúdica por meio da interação do usuário com o ambiente de RV através da visualização do seu corpo e esqueleto.

Os casos de uso identificados na aplicação são listados a seguir;

1. **Caso de uso 1**: Análise no plano lateral/coronal da amplitude dos movimentos da articulação do cotovelo de 0 (zero) graus de extensão até 145 graus de flexão
2. **Caso de uso 2**: Análise da amplitude dos movimentos de abdução da articulação glenoumeral (acrescido dos movimentos da cintura escapular) de até 180 graus de amplitude, partindo da posição ortostática.
3. **Caso de uso 3**: Reconhecimento do Gesto - Posição Inicial : Posição anatômica | Movimento : Abdução de 0 a 90º.
4. **Caso de uso 4**: Reconhecimento do Gesto - Posição Inicial : Abdução de 90º + Rotação posterior | Movimento : Flexão de cotovelo de 0 a 90º.
5. **Caso de uso 5**: Reconhecimento do Gesto - Posição Inicial : Posição anatômica | Movimento : Abdução de 90 a 180º.
6. **Caso de uso 6**: Reconhecimento do Gesto - Posição Inicial : Abdução de 90º + Flexão de cotovelo a 90º | Movimento : Rotação anterior de 0 a 90º.
7. **Caso de uso 7**: Reconhecimento do Gesto - Posição Inicial : Posição anatômica | Movimento : Abdução de 0 a 180º.

Os resultados desse Caso de Uso foram demonstrados nos trabalhos (BRANDAO et al., 2016), (BRANDÃO et al., 2015) e (BRANDÃO et al., 2014a), sendo alcançado o seu objetivo de aplicação.

4.3.3 A aplicação e seus Casos de Uso | II

A segunda aplicação se trata de uma aplicação computacional denominada "GestureChair" onde o usuário controla o personagem do jogo *KapMan* (jogo distribuído em licença de *software*-livre) com movimentos gestuais ao invés de usar o teclado, trazendo um ambiente lúdico e aspectos para determinar e analisar orientação espacial de pessoas. O controle do jogo com movimentos manuais rápidos para cima, baixo, direita ou esquerda (denominado *swipe*) através das mãos; a partir deste ponto o programa reconhece cada gesto e permite ao usuário controlar o jogo. Caso o usuário não fizer o movimento *swipe* rápido o suficiente, o programa não interpreta nenhum gesto prevenindo o reconhecimento de gestos indesejáveis, o que poderia tornar o controle do jogo inviável.

Os casos de uso identificados na aplicação são listados a seguir;

1. **Caso de uso 1**: Controlar o jogo através de gestos.
2. **Caso de uso 2**: Movimentar para cima - Reconhecimento do Gesto: Movimentação das mãos para cima
3. **Caso de uso 3**: Movimentar para baixo - Reconhecimento do Gesto: Movimentação das mãos para baixo
4. **Caso de uso 4**: Movimentar para esquerda - Reconhecimento do Gesto: Movimentação das mãos para esquerda
5. **Caso de uso 5**: Movimentar para direita - Reconhecimento do Gesto: Movimentação das mãos para direita

Os resultados desse Caso de Uso foram demonstrados no trabalho (BRANDÃO et al., 2014b), sendo alcançado o seu objetivo de aplicação.

4.3.4 A aplicação e seus Casos de Uso | III

A terceira aplicação se trata de uma aplicação computacional denominada "*GestureMaps*" de RV para interação, navegação e exploração de imagens panorâmicas em 360 graus da API do Google *Street View*. A segmentação da imagem do usuário é utilizada com o intuito de explorar a sensação mista. As ações de navegação e interação é feita através dos gestos corporais pré determinados, de modo que para avançar no mapa virtual é necessário que o usuário simule uma caminhada.

1. **Caso de uso 1**: Controlar o jogo através de gestos.
2. **Caso de uso 2**: Andar para frente - Reconhecimento do Gesto: Com flexão de quadril e joelho correspondente ao deslocamento mínimo de 15 centímetros entre a posição inicial e final da patela.
3. **Caso de uso 3**: Direção de Navegação Esquerda ou Direita - Reconhecimento do Gesto: Movimento de rotação do tronco para direita ou esquerda, de acordo com a exploração geográfica desejada.

Os resultados desse Caso de Uso foram demonstrados nos trabalhos (BRANDÃO et al., 2013) , (BRASIL et al., 2014) e (BRANDÃO; BRASIL; TREVELIN, 2014), sendo alcançado o seu objetivo de aplicação.

4.3.5 A aplicação e seus Casos de Uso | IV

A quarta aplicação se trata de uma aplicação computacional denominada "*GesturePuzzle*" consiste de um jogo de quebra-cabeça em que a movimentação e encaixe das peças ocorre conforme os gestos do usuário. Para isso, as coordenadas 3D (tridimensionais) da mão do jogador são envidas para o jogo para que as peças sejam movimentadas e encaixadas nas devidas posições. A interface do jogo. No lado esquerdo da tela localiza-se a grade onde as peças devem ser encaixadas; na parte central estão as peças do quebra-cabeça embaralhadas. A mão do usuário é representada na interface por intermédio de uma metáfora gráfica - imagem no formato de uma mão.

Os resultados desse Caso de Uso foram demonstrados nos trabalhos (GUIMARÃES et al., 2011) e (BRANDÃO et al., 2014c), sendo alcançado o seu objetivo de aplicação.

1. **Caso de uso 1**: Para montar o quebra-cabeça, basta que o jogador posicione uma das mãos sobre uma das peças embaralhadas e a arreste para a grade de montagem.

2. **Caso de uso 2**:Se o usuário posicionar a mão sobre área da interface no qual se localiza o texto "próxima imagem", um novo jogo é inicializado, com uma nova imagem.

3. **Caso de uso 3**: Se o jogador posicionar uma das mãos sobre a área da interface cujo texto é "mudar música", uma nova música é tocada durante a execução do texto.

4.3.6 A aplicação e seus Casos de Uso | V

A quinta aplicação se trata de uma aplicação computacional denominada "*GestureChess*"o uso de gestos é utilizado como meio de interação com um jogo de Xadrez virtual, permitindo ao usuário explorar sua habilidade motora e de concentração visando otimizar a tática escolhida, e assim superar seu oponente. Tais ações são realizadas simultaneamente e exigem do usuário maior recrutamento muscular e atividade cognitiva. O xadrez é um jogo de estratégia e tática que demanda considerado desempenho cognitivo e, concomitantemente, mínima atividade motora do membro superior. Entretanto, apesar de mínima, os gestos motores relativos a prática do xadrez recruta grupos musculares constituintes da articulação do punho, cotovelo e ombro, sendo o último especialmente acometido de atrofia muscular durante o processo de senescência. Esta aplicação evidencia o conceito de dupla tarefa durante a execução do xadrez em ambientes virtuais imersivos, controlados a partir de gestos motores por meio de sensores de movimentos que permitem a IHC, oferecendo ao usuário uma interface de entrada que possibilita a manipulação do jogo de xadrez virtual, utilizando apenas gestos intuitivos, tais como pegar e soltar a peça.

1. **Caso de uso 1**: Manipulação de peças através de gestos.
2. **Caso de uso 2**: Pegar a peça com o gesto de *push*,
3. **Caso de uso 3**: Com a peça selecionada, onde ele efetuar o *push* novamente será onde a peça deverá ser movida, cria-se virtualmente coordenadas espaciais para o movimento das peças, as regras do Xadrez devem ser levadas em consideração.

Os resultados (DIAS et al., 2013a), (BRANDÃO et al., 2013) e (BRANDÃO; DIAS; TREVELIN, 2014), sendo alcançado o seu objetivo de aplicação.

Todos os Casos de Uso permitiram a melhor interpretação para o desenvolvimento do modelo de reconhecimento de gestos e também serviram de base de estudo e experiência na implementação da ferramenta GGGesture.

Desta forma cada tipo de reconhecimento de gesto pode ser aplicado como interface natural de entrada e gerar comandos dentro do seu contexto de aplicação foco deste trabalho. Diferentes objetivos são atingidos através da adaptação de gestos estáticos e dinâmicos à realidade de cada aplicação. Nos trabalhos citados utiliza-se os gestos dinâmicos do corpo como a caminhada e rotação do corpo como meio de interação que tem o intuito de prover a exploração virtual espacial e geográfica que pode ser aplicada em RV e na Desorientação Espacial. Já no contexto de atividade motora e cognitiva é aplicado a interface natural através do gesto onde o controle de um jogo de xadrez é realizado pelos movimentos contínuos da mão e a representação de um gesto de clique, também se aplica interface natural do gesto a reabilitação do membro superior que é composta por uma aplicação de RV onde se utiliza os movimentos das mãos e gestos denominados de *swipe*.

Capítulo 5

CONCLUSÕES

5.1 Conclusão

Neste trabalho foi proposto um modelo e uma ferramenta para reconhecimento de gestos do corpo através de um sensor de baixo custo e soluções de tecnologias abertas para ambientes convencionais. Para a implementação deste trabalho foram realizadas, pesquisas e análises bibliográficas dos trabalhos relativos a gestos, sensores visuais e algoritmos de reconhecimento de gestos.

Durante produção deste trabalho foram desenvolvidos estudos e Casos de Uso para validação das ferramentas e tecnologias a serem utilizadas. Descobriu-se extensas aplicações e lacunas para utilização das Interfaces Naturais em áreas da saúde e educação, além do grande potencial de conexões da Interface Natural com ambientes virtuais na RV. Os resultados produzidos neste trabalho impactam não somente na computação, mas também em diversas outras áreas do conhecimento, como educação e saúde como demonstrado.

Foi observado também que a inclusão de dispositivos e sensores visuais de baixo custo contribuíram para evolução da área de Interfaces Naturais, sendo assim muitas aplicações estão sendo propostas, mas limitando-se a um conjunto pré-definido de gestos e não permitindo a inclusão de novos ou a troca gestos o que dificulta a utilização e expansão de tal área.

Em cada trabalho de Caso de Uso, há uma definição e atribuição dos significados na classificação dos conjuntos de posturas e movimentos de forma única para cada aplicação. Neste sentido, foi proposto também uma ferramenta que permite aos próprios usuários definirem quais poses ou gestos que serão utilizados em suas aplicações.

Uma visão geral do Modelo Proposto deste trabalho é apresentada na Figura 3.1, sendo que as etapas de Aquisição e Modelagem do Gesto foram concluídas e a etapa de Interface Multi-

modal torna-se trabalho futuro. Apesar das dificuldades ao longo das fases deste trabalho, os objetivos foram alcançados, isto é, produzir um modelo de reconhecimento e implementação de uma ferramenta computacional. No modelo foram usados conjuntos de características que foram variantes a produção e observação do gesto em frente ao sensor, embora o modelo tenha sido validado para diferentes situações, não foi apresentado um bom resultado para casos de variação e oclusão visual na detecção das juntas de origem da biblioteca OpenNI, porém as taxas de erros estiveram dentro de intervalos que não comprometem a usabilidade e a sua aplicabilidade.

5.2 Produções Científicas e Tecnológicas da Pesquisa

Durante o desenvolvimento deste trabalho foram apresentados e publicados alguns estudos que foram redigidos para congressos, simpósios e livros, no qual trazem como base de metodologias e estudos de caso para este trabalho onde as colaborações diretas com os autores contribuíram para este trabalho. As produções desenvolvidas permitiram visualizar diferentes possibilidades para a utilização de gestos em diversos contextos e trouxeram contribuição para área acadêmica.

- Marca GestureCollection registrada no INPI: são um conjunto de ferramentas representadas por três aplicações que permitem a interação homem-computador através de gestos do motor (chamado GestureChess, GestureMaps e GesturePuzzle). Eles fornecem estímulos motores e cognitivos em situações de ensino-aprendizagem, reabilitação neuromuscular e entretenimento fisicamente ativo.

- BRANDAO, A. F., DIAS, D. R., CASTELLANO, G., PARIZOTTO, N. A.,& TREVELIN, L. C. RehabGesture: An Alternative Tool for Measuring Human Movement. Telemedicine and e-Health, v. 22, n. 7, p. 584-589, 2016.

- BRANDÃO, A. F.; PARIZOTTO, N. A. ; TREVELIN, LC . Controle de ambientes virtuais por interação gestual. 2015. . In: XVIII Simpósio Sesc de Atividades Físicas Adaptadas, 2015, São Carlos. XVIII Simpósio Sesc de Atividades Físicas Adaptadas, 2015. v. 18.

- BRANDÃO, A. F.; ALMEIDA, S. R. M. ; DIAS, DRC ; BRASIL, GJC ; TREVELIN, L. C. ; CASTELLANO, G. . Realidade Virtual associada a dupla tarefa e ao estudo do movimento humano. In: V Simpósio Internacional de Ciência do Desporto, 2015, Campinas. Anais do V Simpósio Internacional de Ciência do Desporto e VI Congresso de Ciência do Desporto, 2015.

- BRANDÃO, A. F.; PARIZOTTO, N. A. ; TREVELIN, L. C. . Gesture recognition for health and rehabilitation. In: IX Congresso Internacional de Educação Física e Motricidade Humana, 2015, Rio Claro. Motriz. Rio Claro: Unesp, 2015.

- BRANDÃO, A. F.; PARIZOTTO, N. A. ; TREVELIN, L. C. . RehabGesture for support in kinesiology class. In: IX Congresso Internacional de Educação Física e Motricidade Humana, 2015, Rio Claro. Motriz. Rio Claro: Unesp, 2015.

- BRANDÃO, A. F.; ALMEIDA, S. R. M. ; MIN, L. L. ; PARIZOTTO, N. A. ; CASTELLANO, G. ; TREVELIN, L. C. . Software for Measuring the Amplitude of Arm Movement in Stroke Patients. In: 2nd BRAINN Congress, 2015, Campinas. BRAINN Congress, 2015.

- BRANDÃO, A. F.; PARIZOTTO, N. A. ; TREVELIN, L. C. . Luz Estruturada Aplicada ao Reconhecimento de Gestos para Interação Humano Computador Fisicamente Ativa: GestureCollection. In: 67ª Reunião Anual da Sociedade Brasileira para o Progresso da Ciência (SBPC), 2015, São Carlos. Anais/Resumos da 67ª Reunião Anual da SBPC, 2015. v. 67.

- BRANDÃO, A. F.; BRASIL, GJC ; DIAS, DRC ; ALMEIDA, S. R. M. ; CASTELLANO, G. ; TREVELIN, LC . Realidade Virtual e Reconhecimento de Gestos Aplicados às Áreas de Saúde. Tendências e Técnicas em Realidade Virtual e Aumentada, v. 4 p. 33-48, 2014; ISSN/ISBN: 21776776.

- PIEMONTE, MEP ; BRANDÃO, A. F. . Tecnologia pode ser aliada da saúde. Espaço Aberto USP, Cidade Universitária - USP, p. 13 - 17, 05 nov. 2014.

- BRANDÃO, A. F.; DIAS, DRC ; BRASIL, GJC ; ALMEIDA, S. R. M. ; MIN, L. L. ; CASTELLANO, G. ; TREVELIN, LC . Virtual reality for motor and cognitive stimulation. In: 1ST Congress CEPID BRAINN, 2014, Campinas. Anais CEPID, 2014.

- SOARES, M.C. ; MANZINI, M. G. ; FERRIGNO, ISV ; TREVELIN, LC ; PARIZOTTO, N. A. ; BRANDÃO, A. F. . Avaliação de um aplicativo de Realidade Virtual para reabilitação de ombro: percepção de profissionais da saúde. In: XVIII Semana de Estudos em Terapia Ocupacional da UFSCar, 2014, São Carlos. XVIII Semana de Estudos em Terapia Ocupacional da UFSCar, 2014.

- BRANDÃO, A. F.; BRASIL, GJC ; DIAS, DRC ; TREVELIN, LC . Realidade Virtual Aplicada ao Transtorno do Movimento. In: III Jornada de Estudos da Doença de Parkinson, 2014, Rio Claro. JEDP, 2014.

- BRANDÃO, A. F.; BRASIL, GJC ; DIAS, DRC ; TREVELIN, LC . Software para Análise da Amplitude de Movimento dos Membros Superiores no Plano Coronal. In: V Congresso de Ciência do Desporto e IV Seminário Internacional de Ciência do Desporto, 2014, Unicamp. Anais CCD2014, 2014. v. 1.

- BRANDÃO, A. F.. Jogos da reabilitação. Revista Pesquisa FAPESP, Tecnociência, , v. 213, p. 12 - 12, 25 nov. 2013.

- TREVELIN, LC ; BRANDÃO, A. F. ; DIAS, DRC ; BRASIL, GJC . Aplicativos de computador podem ser utilizados para estimular exercícios físicos e auxiliar em tratamentos. Universidade Federal de São Carlos - serviços, Notícia, 22 out. 2013.

- TREVELIN, LC ; BRANDÃO, A. F. ; DIAS, DRC ; BRASIL, GJC . Trabalho de pesquisadores do PPGCC e PPGBiotec recebe menção honrosa em Conferências sobre Saúde. SAOCARLOSOFICIAL, Notícias, 04 set. 2013.

- BRANDÃO, A. F.; PARIZOTTO, N. A. ; TREVELIN, L. C. . Pesquisa da UFSCar que desenvolveu aplicativos que estimulam exercícios físicos e auxiliam em tratamentos recebe menção honrosa. UFSCar Notícias, UFSCar website.

- GNECCO, B. B., DIAS, D. R. C., BRASIL, G. J. C., & GUIMARÃES, M. P. Desenvolvimento de Aplicações com Interface Natural de Usuário e Dispositivos PrimeSense como Meio de Interação para Ambientes Virtuais. Tendências e Técnicas em Realidade Virtual e Aumentada, v. 3, p. 89-103, 2013; ISSN/ISBN: 21776776.

- GNECCO, B. B. ; DOMINGUES, R. G. ; BRASIL, G. J. C. ; DIAS, D. R. C. ; TREVELIN, L.C . Estratégias Mistas de Mecanismos para Imersão em Modelos de Interação. Tendências e Técnicas em Realidade Virtual e Aumentada, v. 3, p. 104-120, 2013; ISSN/ISBN: 21776776.

- BRANDÃO, A. F.; SOARES, M.C. ; BRASIL, GJC ; DIAS, DRC ; FABBRO, A. C. ; DUARTE, A. C. G. O. ; TREVELIN, LC . Prevenção de atrofia muscular da articulação glenoumeral por meio de atividade física adaptada à realidade virtual e reconhecimento de gestos. In: Simpósio SESC de Atividades Físicas Adaptadas, 2013, São Carlos. Anais do evento, 2013.

- DIAS, D. R. C., BRANDAO, A. F., BRASIL, G. J. C., GUIMARAES, M. P., BREGA, J. R. F., & TREVELIN, L. C. Gesture Chess - Interface Natural de Usuário na Atividade Motora e Cognitiva. In: WRVA 2013, 2013, Jataí. Anais do WRVA, 2013.

- BRANDÃO, A. F.; BRASIL, GJC ; DIAS, DRC ; TREVELIN, LC . Paradigm Shift in Human Interaction with Virtual Reality Environments. In: VIII Congresso Internacional de Educação Física e Motricidade Humana e XIV Simpósio Paulista de Educação Física, 2013, Rio Claro. VIII Congresso Internacional de Educação Física e Motricidade Humana e XIV Simpósio Paulista de Educação Física, 2013. ISSN/ISBN: 1980-6574.

- BRANDÃO, A. F.; DIAS, DRC ; BRASIL, GJC ; TREVELIN, LC . Gesture Chess: Tarefa Dupla em Ambiente Virtual. In: Conferências USP: Determinantes Sociais de

Saúde e Ações Interprofissionais, 2013, Ribeirão Preto. Anais do DSSAI, 2013. v. 1. p. 1-1.

- BRANDÃO, A. F.; BRASIL, GJC ; DIAS, DRC ; TREVELIN, LC . Gesturemaps: Perspectivas para a desorientação espacial. In: IV Colóquio Internacional de Gerontologia, 2013, Ribeirão Preto. Medicina (USP.FMRP), 2013. ISSN/ISBN: 0076-6046.

- FABBRO, A. C. ; TREVELIN, LC ; BRASIL, GJC ; DIAS, DRC ; SOARES, M.C. ; BRANDÃO, A. F. . Kinect: aprendendo e desenvolvendo aplicações para o hardware. In: 10ª Jornada Científica e Tecnológica da UFSCar, 2013, São Carlos. 10ª Jornada Científica e Tecnológica da UFSCar, 2013.

- BRANDÃO, A. F.; DIAS, DRC ; BRASIL, GJC ; SOARES, M.C. ; TREVELIN, LC . Atividade Motora, Cognição e Realidade Virtual. In: IV Simpósio de Biotecnologia da UFSCar, 2013, São Carlos. IV Simpósio de Biotecnologia da UFSCar, 2013.

- SOARES, M.C. ; MANZINI, M. G. ; BRASIL, GJC ; DIAS, DRC ; FABBRO, A. C. ; TREVELIN, LC ; BRANDÃO, A. F. . Avaliação da eficácia da Realidade Virtual para Reabilitação do Ombro. In: IV Simpósio de Biotecnologia da UFSCar, 2013, São Carlos. IV Simpósio de Biotecnologia da UFSCar, 2013.

- GNECCO, B. B., DIAS, D. R. C., BRASIL, G. J. C., & GUIMARÃES, M. P. Desenvolvimento de Interfaces Naturais de Interação usando o Hardware Kinect. Tendências e Técnicas em Realidade Virtual e Aumentada, v. 2, p. 37-62, 2012 ISSN/ISBN: 21776776.

- GUIMARÃES, M. P., MARTINS, V. F., TREVELIN, L. C., & BRASIL, G. J Um Modelo de Processo de Desenvolvimento de Interfaces de Gesto: Definição e um Estudo de Caso. In: XXXVII Conferencia Latino-americana de Informática (XXXVII CLEI), 2011, Quito

- BRASIL, G. J. C.; GUIMARÃES, M. P. . Um jogo de quebra-cabeça para interface natural de usuário. In: XIX Congresso de Iniciação Científica da UFSCar, 2011, São Carlos. XIX Congresso de Iniciação Científica da UFSCar,2011.

5.3 Trabalhos Futuros

O modelo e ferramentas desenvolvidos neste trabalho representam os primeiros passos, porem foram encontradas algumas tarefas que podem resultar em uma robustez neste trabalho, que são:

- No modelo de reconhecimento de gestos permitir um classificador através do agrupamento de unidades de gestos de vários usuários diferentes ou do mesmo usuário em que representam o mesmo significado.
- Reduzir o vetor características para maior desempenho de velocidade e otimização no modelo de reconhecimento de gestos
- Implementar na ferramenta GGGesture a transformação das coordenadas no plano do corpo para o ângulo zero, assim todos os usuários estarão de face ao sensor.
- Implementar na ferramenta GGGesture a análise em tempo real do reconhecimento de gestos por meio de um limite de tempo apropriado para cada gesto a ser comparado.
- Implementar na ferramenta GGGesture uma API para enviar eventos de reconhecimento de gestos a outras aplicações
- Gerar um conjunto de gestos maior para testar mais a capacidade da ferramenta e do modelo proposto.

REFERÊNCIAS

AGGARWAL, J.; PARK, S. Human motion: modeling and recognition of actions and interactions. In: *3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings. 2nd International Symposium on*. [S.l.: s.n.], 2004. p. 640–647.

AGGARWAL, J.; RYOO, M. Human activity analysis: A review. *ACM Comput. Surv.*, ACM, New York, NY, USA, v. 43, n. 3, p. 16:1–16:43, abr. 2011. ISSN 0360-0300. Disponível em: <http://doi.acm.org/10.1145/1922649.1922653>.

AKLEMAN, E.; CHEN, J. Generalized distance functions. In: *Shape Modeling and Applications, 1999. Proceedings. Shape Modeling International '99. International Conference on*. [S.l.: s.n.], 1999. p. 72–79.

BAE, S.-H. et al. Tangible nurbs-curve manipulation techniques using graspable handles on a large display. In: *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*. New York, NY, USA: ACM, 2004. (UIST '04), p. 81–90. ISBN 1-58113-957-8. Disponível em: <http://doi.acm.org/10.1145/1029632.1029646>.

BAUDEL, T.; BEAUDOUIN-LAFON, M. Charade: remote control of objects using free-hand gestures. *Communications of the ACM*, ACM, v. 36, n. 7, p. 28–35, 1993.

BERNSEN, N. O. Defining a taxonomy of output modalities from an hci perspective. *Computer Standards & Interfaces*, Elsevier, v. 18, n. 6, p. 537–553, 1997.

BESL, P. J. Active, optical range imaging sensors. *Machine vision and applications*, Springer, v. 1, n. 2, p. 127–152, 1988.

BOBICK, A. F. Movement, activity, and action: The role of knowledge in the perception of motion. *Royal Society Workshop on Knowledge-based Vision in Man and Machine*, v. 352, p. 1257–1265, 1997.

BOGIN, B.; VARELA-SILVA, M. I. Leg length, body proportion, and health: a review with a note on beauty. *International journal of environmental research and public health*, Molecular Diversity Preservation International, v. 7, n. 3, p. 1047–1075, 2010.

BRANDÃO, A. F. et al. Gesturemaps: perspectivas para a desorientação espacial. *Revista da Faculdade de Medicina de Ribeirão Preto e do Hospital das Clínicas da FMRP-USP*, ISSN 0076-6046, v. 46, n. Supl 4, p. 27, 2013.

BRANDÃO, A. F. et al. Software para análise da amplitude de movimento dos membros superiores no plano coronal. In: ISBN 978-85-99688-19-9. *V Congresso de Ciência do Desporto*. [S.l.], 2014. p. 172.

BRANDÃO, A. F. et al. Realidade virtual e reconhecimento de gestos aplicada as áreas de saúde. *Tendências e Técnicas em Realidade Virtual e Aumentada*, v. 1, n. 4, p. 33–48, 2014.

BRANDÃO, A. F. et al. *GesturePuzzle*. 2014. BR Patent App. BR 51 2014 001378 2.

BRANDÃO, A. F. et al. *RehabGesture*. 2015. BR Patent App. BR 51 2014 001376 6.

BRANDÃO, A. F.; BRASIL, G. J. C.; TREVELIN, L. C. *GestureMaps*. 2014. BR Patent App. BR 51 2014 001376 6.

BRANDÃO, A. F. et al. Gesturechess: tarefa dupla em ambiente virtual. In: ISBN 978-85-64922-03-7. *Conferências USP sobre Determinantes Sociais de Saúde e Ações Interprofissionais*. [S.l.], 2013. p. 62.

BRANDAO, A. F. et al. Rehabgesture: An alternative tool for measuring human movement. *Telemedicine and e-Health*, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 22, n. 7, p. 584–589, 2016.

BRANDÃO, A. F.; DIAS, D. R. C.; TREVELIN, L. C. *GestureChess*. 2014. BR Patent App. BR 51 2014 001377 4.

BRASIL, G. J. et al. Natural user interface applied for spatial disorientation and movement disorder: Gesturemaps. In: ISBN 978-85-87937-20-9. *Workshop de Realidade Virtual e Aumentada*. [S.l.], 2014. p. 72–76.

CASTRO, A. A. M. de; PRADO, P. P. L. do. Algoritmos para reconhecimento de padrões. *Revista Ciências Exatas*, v. 8, n. 2002, 2002.

CORPORATION, M. A. Motion analysis corporation, http://www.motionanalysis.com. 2014. Disponível em: <http://www.motionanalysis.com>.

CORRADINI, A. Dynamic time warping for off-line recognition of a small gesture vocabulary. In: IEEE. *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 2001. Proceedings. IEEE ICCV Workshop on*. [S.l.], 2001. p. 82–89.

CORRADINI, A. Real-time gesture recognition by means of hybrid recognizers. In: *Gesture and sign language in human-computer interaction*. [S.l.]: Springer, 2002. p. 34–47.

DIAS, D. R. et al. Gesture chess - interface natural de usuário na atividade motora e cognitiva. In: *Workshop de Realidade Virtual e Aumentada*. [S.l.: s.n.], 2013. p. 115–120.

DIAS, D. R. C. et al. Desenvolvimento de aplicações com interface natural de usuário e dispositivos primesense como meio de interação para ambientes virtuais. In: *Tendências e Técnicas em Realidade Virtual e Aumentada*. [S.l.: s.n.], 2013. v. 3. ISSN 2177-6776.

DIX, A. *Human-computer Interaction*. Pearson/Prentice-Hall, 2004. ISBN 9780130461094. Disponível em: <http://books.google.com.br/books?id=IuQxui8GHDcC>.

DONG, L.; WU, J.; CHEN, X. A body activity tracking system using wearable accelerometers. In: *Multimedia and Expo, 2007 IEEE International Conference on*. [S.l.: s.n.], 2007. p. 1011–1014.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. [S.l.]: John Wiley & Sons, 2012.

FARHADI-NIAKI, F.; GHASEMAGHAEI, R.; ARYA, A. Empirical study of a vision-based depth-sensitive human-computer interaction system. In: ACM. *Proceedings of the 10th asia pacific conference on Computer human interaction*. [S.l.], 2012. p. 101–108.

FELDMAN, R. S.; RIMÉ, B. *Fundamentals of nonverbal behavior*. [S.l.]: Cambridge University Press, 1991.

FELS, S. S.; HINTON, G. E. Glove-talk: A neural network interface between a data-glove and a speech synthesizer. *Neural Networks, IEEE Transactions on*, IEEE, v. 4, n. 1, p. 2–8, 1993.

FERREIRA, A. de H.; FERREIRA, M.; ANJOS, M. dos. *Novo dicionário Aurélio da língua portuguesa*. Editora Positivo, 2009. Disponível em: <http://books.google.com.br/books?id=mMm6twAACAAJ>.

FUJIYOSHI, H.; LIPTON, A. J. Real-time human motion analysis by image skeletonization. In: IEEE. *Applications of Computer Vision, 1998. WACV'98. Proceedings., Fourth IEEE Workshop on*. [S.l.], 1998. p. 15–21.

FURNISS, M. Motion capture. *posted at http://web. mit. edu/mit/articles/index—furniss. html on Dec*, v. 19, 1999.

GIORGINO, T. et al. Computing and visualizing dynamic time warping alignments in r: the dtw package. *Journal of statistical Software*, v. 31, n. 7, p. 1–24, 2009.

GNECCO, B. B. et al. Desenvolvimento de interface naturais de interação usando o hardware kinect. In: *Tendências e Técnicas em Realidade Virtual e Aumentada*. [S.l.: s.n.], 2012. v. 2. ISSN 2177-6776.

GOKTURK, S. B.; YALCIN, H.; BAMJI, C. A time-of-flight depth sensor-system description, issues and solutions. In: IEEE. *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*. [S.l.], 2004. p. 35–35.

GUIMARÃES, M. et al. Um modelo de processo de desenvolvimento de interfaces de gesto: Definiçao e um estudo de caso. In: *XXXVII Conferencia Latinoamericana de Informática (CLEI)*. [S.l.: s.n.], 2011. v. 1, p. 378–390.

GUIZZO, E. Ieee robotics blog. *Hands-On With the Next Generation Kinect: PrimeSense Capri. http://spectrum. ieee. org/automaton/robotics/robotics-hardware/handson-with-the-next-generation-kinect-primesense-capri*, 2013.

GUNA, J. et al. An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 14, n. 2, p. 3702–3720, 2014.

HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.l.]: Elsevier, 2011.

HECHT, J. Photonic frontiers: Gesture recognition: Lasers bring gesture recognition to the home. *Laser Focus World*, p. 1–5, 2011.

HEWETT, T. T. et al. *ACM SIGCHI Curricula for Human-Computer Interaction*. New York, NY, USA, 1992.

HICKSON, M.; STACKS, D. *NVC, Nonverbal Communication: Studies and Applications*. W.C. Brown Publishers, 1985. ISBN 9780697003133. Disponível em: <http://books.google.com.br/books?id=flktAAAAMAAJ>.

INC, A. Siri - your wish is its command. Maio 2014. Disponível em: <http://www.apple.com/ios/siri/>.

INC, G. Google now - just the right information at just the right time. Maio 2014. Disponível em: <http://www.google.com/landing/now/>.

INC, M. Cortana - the most personal smartphone assistant. Maio 2014. Disponível em: <http://www.windowsphone.com/en-us/features-8-1>.

JAIMES, A.; SEBE, N. Multimodal human–computer interaction: A survey. *Computer vision and image understanding*, Elsevier, v. 108, n. 1, p. 116–134, 2007.

JUNIOR, J. A. et al. Estudo da influência de diversas medidas de similaridade na previsão de séries temporais utilizando o algoritmo knn-tsp. Universidade Estadual do Oeste do Parana, 2012.

KENDON, A. Gesticulation and speech: Two aspects of the process of utterance. *The relationship of verbal and nonverbal communication*, v. 25, p. 207–227, 1980.

KENDON, A. Do gestures communicate? a review. *Research on language and social interaction*, Taylor & Francis, v. 27, n. 3, p. 175–200, 1994.

KIM, D.; SONG, J.; KIM, D. Simultaneous gesture segmentation and recognition based on forward spotting accumulative hmms. *Pattern Recognition*, Elsevier, v. 40, n. 11, p. 3012–3026, 2007.

KWON, D. Y. *A design framework for 3D spatial gesture interfaces*. Tese (Doutorado) — ETH Zurich, 2008.

LIEBLING, D.; MORRIS, M. R. Kinected browser: depth camera interaction for the web. In: ACM. *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*. [S.l.], 2012. p. 105–108.

LIU, W. Natural user interface-next mainstream product user interface. In: *2010 IEEE 11th International Conference on Computer-Aided Industrial Design&Conceptual Design 1*. [S.l.: s.n.], 2010. v. 1, p. 203–205.

LUCIANI, A. et al. A basic gesture and motion format for virtual reality multisensory applications. *arXiv preprint arXiv:1005.4564*, 2010.

MCNEILL, D. *Hand and mind: What gestures reveal about thought*. [S.l.]: University of Chicago Press, 1992.

MCNEILL, D.; LEVY, E. Conceptual representations in language activity and gesture. ERIC, 1980.

MOESLUND, T. B.; GRANUM, E. A survey of computer vision-based human motion capture. *Computer vision and image understanding*, Elsevier, v. 81, n. 3, p. 231–268, 2001.

MORAN, T. P. The command language grammar: a representation for the user interface of interactive computer systems. *International Journal of Man-Machine Studies*, v. 15, n. 1, p. 3 – 50, 1981. ISSN 0020-7373. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0020737381800223>.

MORETTIN, P. A.; TOLOI, C. *Análise de séries temporais*. [S.l.]: Blucher, 2006.

NEFIAN, A. V. et al. Dynamic bayesian networks for audio-visual speech recognition. *EURASIP J. Appl. Signal Process.*, Hindawi Publishing Corp., New York, NY, United States, v. 2002, n. 1, p. 1274–1288, jan. 2002. ISSN 1110-8657. Disponível em: <http://dx.doi.org/10.1155/S1110865702206083>.

NEWMAN, J. et al. Ubiquitous tracking for augmented reality. In: *Mixed and Augmented Reality, 2004. ISMAR 2004. Third IEEE and ACM International Symposium on*. [S.l.: s.n.], 2004. p. 192–201.

NGUYEN, K. D. et al. A wearable sensing system for tracking and monitoring of functional arm movement. *Mechatronics, IEEE/ASME Transactions on*, v. 16, n. 2, p. 213–220, April 2011. ISSN 1083-4435.

NORMAN, D. A. *The design of everyday things: Revised and expanded edition*. [S.l.]: Basic books, 2013.

NORMAN, D. A.; NIELSEN, J. Gestural interfaces: a step backward in usability. *interactions*, ACM, v. 17, n. 5, p. 46–49, 2010.

OVIATT, S.; COHEN, P. Perceptual user interfaces: Multimodal interfaces that process what comes naturally. *Commun. ACM*, ACM, New York, NY, USA, v. 43, n. 3, p. 45–53, mar. 2000. ISSN 0001-0782. Disponível em: <http://doi.acm.org/10.1145/330534.330538>.

PAVLOVIC, V. et al. Visual interpretation of hand gestures for human-computer interaction: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 19, n. 7, p. 677–695, 1997.

PIEGL, L.; TILLER, W. *The NURBS book*. [S.l.]: Springer Science & Business Media, 2012.

POSDAMER, J.; ALTSCHULER, M. Surface measurement by space-encoded projected beam systems. *Computer graphics and image processing*, Elsevier, v. 18, n. 1, p. 1–17, 1982.

PRATES, R. O.; BARBOSA, S. D. J. Avaliação de interfaces de usuário–conceitos e métodos. In: *Jornada de Atualização em Informática do Congresso da Sociedade Brasileira de Computação, Capítulo*. [S.l.: s.n.], 2003. v. 6.

PREECE, J. et al. *Human-computer interaction*. [S.l.]: Addison-Wesley Longman Ltd., 1994.

PUSTKA, D.; KLINKER, G. Dynamic gyroscope fusion in ubiquitous tracking environments. In: *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. Washington, DC, USA: IEEE Computer Society, 2008. (ISMAR '08), p. 13–20. ISBN 978-1-4244-2840-3. Disponível em: <http://dx.doi.org/10.1109/ISMAR.2008.4637317>.

QUEK, F. K. Toward a vision-based hand gesture interface. In: *Virtual Reality Software and Technology Conference*. [S.l.: s.n.], 1994. p. 17–29.

RABINER, L. R.; JUANG, B.-H. Fundamentals of speech recognition. PTR Prentice Hall, 1993.

RAUTARAY, S. S.; AGRAWAL, A. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, v. 43, n. 1, p. 1–54, 2015. ISSN 1573-7462. Disponível em: <http://dx.doi.org/10.1007/s10462-012-9356-9>.

RECTOR, M.; TRINTA, A. R. *Comunicação do corpo*. [S.l.]: Ática, 1999.

SAFFER, D. *Designing gestural interfaces: Touchscreens and interactive devices*. [S.l.]: "O'Reilly Media, Inc.", 2008.

SAKOE, H.; CHIBA, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, IEEE, v. 26, n. 1, p. 43–49, 1978.

SALVADOR, S.; CHAN, P. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, IOS Press, v. 11, n. 5, p. 561–580, 2007.

SATO, T. d. O.; HANSSON, G.-Å.; COURY, H. J. C. G. Goniometer crosstalk compensation for knee joint applications. *Sensors*, Molecular Diversity Preservation International, v. 10, n. 11, p. 9994–10005, 2010.

SENIN, P. Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, Citeseer, v. 855, p. 1–23, 2008.

SHNEIDERMAN, B.; PLAISANT, C. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. ADDISON WESLEY Publishing Company Incorporated, 2010. ISBN 9780321537355. Disponível em: <http://books.google.com.br/books?id=2CfROgAACAAJ>.

SHOTTON, J. et al. Real-time human pose recognition in parts from single depth images. In: *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE Computer Society, 2011. (CVPR '11), p. 1297–1304. ISBN 978-1-4577-0394-2. Disponível em: <http://dx.doi.org/10.1109/CVPR.2011.5995316>.

SHOTTON, J. et al. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, ACM, v. 56, n. 1, p. 116–124, 2013.

STIEFMEIER, T. et al. Combining motion sensors and ultrasonic hands tracking for continuous activity recognition in a maintenance scenario. In: IEEE. *Wearable Computers, 2006 10th IEEE International Symposium on*. [S.l.], 2006. p. 97–104.

TAHA, A. et al. Skeleton-based human activity recognition for video surveillance. *International Journal of Scientific & Engineering Research*, v. 6, n. 1, p. 993–1004, 2015.

TECHNOLOGIES, F. D. 5dt data gloves. Maio 2014. Disponível em: <http://www.5dt.com>.

TECHNOLOGY, A. Tracking 3d worlds. 2014.

TOU, J. T.; GONZALEZ, R. C. Pattern recognition principles. 1974.

TRUCCO, E.; VERRI, A. *Introductory techniques for 3-D computer vision*. [S.l.]: Prentice Hall Englewood Cliffs, 1998.

VALLI, A. The design of natural interaction. *Multimedia Tools and Applications*, Springer, v. 38, n. 3, p. 295–305, 2008.

WATKINS, M. P.; PORTNEY, L. *Foundations of clinical research: applications to practice*. [S.l.]: Pearson/Prentice Hall, 2009.

WEI, T.; QIAO, Y.; LEE, B. Kinect skeleton coordinate calibration for remote physical training. In: *Proceedings of the International Conference on Advances in Multimedia (MMEDIA)*. [S.l.: s.n.], 2014. p. 23–27.

WEISE, T. et al. Online loop closure for real-time interactive 3d scanning. *Computer Vision and Image Understanding*, Elsevier, v. 115, n. 5, p. 635–648, 2011.

WIGDOR, D.; WIXON, D. *Brave NUI world: designing natural user interfaces for touch and gesture*. [S.l.]: Elsevier, 2011.

WINKLER, M. B. et al. Automatic camera control for tracking a presenter during a talk. In: IEEE. *Multimedia (ISM), 2012 IEEE International Symposium on*. [S.l.], 2012. p. 471–476.

YOUNG, T. Y. *Handbook of pattern recognition and image processing (vol. 2): computer vision*. [S.l.]: Academic Press, Inc., 1994.

ZHAO, L.; BADLER, N. I. Synthesis and acquisition of laban movement analysis qualitative parameters for communicative gestures. 2001.