FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Superfícies Interativas com Kinect

Rui Miguel Costeira Alves da Costa

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Orientador: Eurico Manuel Elias de Morais Carrapatoso (PhD)

Co-orientador: António Abel Vieira de Castro (PhD)

22 de Julho de 2013

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

MIEEC - MESTRADO INTEGRADO EM ENGENHARIA ELETROTÉCNICA E DE COMPUTADORES **2012/2013**

A Dissertação intitulada

"Superfícies Interativas com Kinect"

foi aprovada em provas realizadas em 22 Julho 2013

o júri

Presidente Professor Doutor Miguel Fernando Paiva Velhote Correia
Professor Auxiliar do Departamento de Engenharia Eletrotécnica e de Computadores da Faculdade de Engenharia da Universidade do Porto

Professora Doutora Ana Maria Perfeito Tomé
Professora Associada do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro

Professor Doutor Eurico Manuel Elias Morais Carrapatoso
Professor Auxiliar do Departamento de Engenharia Eletrotécnica e de Computadores da Faculdade de Engenharia da Universidade do Porto

Professor Doutor António Vieira de Castro
Professor Adjunto do ISEP-IPP

O autor declara que a presente dissertação (ou relatório de projeto) é da sua exclusiva autoria e foi escrita sem qualquer apoio externo não explicitamente autorizado. Os resultados, ideias, parágrafos, ou outros extratos tomados de ou inspirados em trabalhos de outros autores, e demais referências bibliográficas usadas, são corretamente citados.

Autor - Rui Miguel Costeira Alves da Costa

Faculdade de Engenharia da Universidade do Porto

Resumo

O objetivo principal deste trabalho era conceber e realizar uma interface tátil com recurso ao sensor Kinect, que permitisse o controlo do sistema operativo Windows 8, interagindo diretamente numa superfície, como uma parede ou um quadro, que não esteja dotada de nenhum meio tecnológico. Podiam transformar-se, assim, grandes superfícies em enormes ecrãs táteis, dotando as típicas projeções de uma forma de interação com as mesmas, através da mãos.

Para tal, iniciou-se o trabalho estudando o processo de aquisição de dados através do sensor Kinect, que permite obter informação referente à profundidade de cada pixel da imagem captada, imagem de cor RGB e ainda um rastreio do esqueleto do utilizador do sistema. Analisou-se depois como todos os dados sincronizados e obtidos em tempo real, se pode realizar o processo de segmentação das mãos do utilizador, uma vez que pretende-se reduzir a essa parte do corpo a interação com a superfície. Essa segmentação é, então, realizada em termos de profundidade, com a obtenção da localização da mão do individuo e definindo limites que permitiram realizar um binarização da mesma, eliminando todos os artefactos da imagem que não sejam parte integrante da mão. Contudo, o facto de junto à superfície essa segmentação não produzir resultados satisfatórios, levou a que se realizasse uma segmentação relativamente à cor de pele, convertendo a imagem RGB para o espaço de cor HSV e posteriormente com a definição de limites de cada componente deste espaço, se detetasse apenas a mão do utilizador através da cor da sua pele. Conseguiu-se assim, obter com a interseção destas duas segmentações, a mão perfeitamente definida e livre de qualquer artefacto que não pertença à mesma.

Com a identificação da mão, definiu-se um processo de extração de características, com a utilização do *wrapper* Emgu CV, que permitissem obter mais informação relativamente à mão, como o centro da sua palma e a localização dos dedos, conseguindo-se, posteriormente, perceber onde se encontra cada parte em termos de profundidade.

Tendo toda a informação reunida, foi necessário processá-la de forma a se perceber que ação está a ser realizada e se está, ou não, a ser realizado algum tipo de toque ou interação com a projeção. Para isso, desenvolveu-se uma janela de calibração, que permitiu definir, não só uma zona de toque, como também informar o sistema da localização da projeção na superfície.

Com a combinação de todos estes processos e de toda a informação relativa à mão, aos dedos e à superfície, desenvolveu-se um método de rastreio de toque na superfície, que permite ao sistema perceber se o toque aconteceu, ou não, e através das características da mão, como a sua orientação ou posição, realizar diferentes tipos de ações, como por exemplo, os típicos cliques esquerdo e direito do cursor.

Conseguiu-se, desta forma, dar ao utilizador capacidade de interagir diretamente com a projeção através do toque, sem nenhum tipo de dispositivo físico.

Abstract

The main objective of this work was the creation of a tactile interface with appeal to Kinect sensor, that allow the control of the operating system Window 8, interacting directly onto a surface, such as a wall or a table that is not endowed with any technological means. Could become, thus, large tactile surfaces in huge screens, giving the typical projections a form of interaction with them, through the hands.

To this end, it started the work by studying the process of acquisition of data through the sensor Kinect, which provides information regarding the depth of each pixel of the captured image, the image of RGB color and even a trace of the skeleton of the user of the system. It was analyzed after all synchronized data and obtained in real time, you can carry out the process of segmentation of the user's hands, since it is intended to reduce the this part of the body's interaction with the surface. This segmentation is then performed in terms of depth, with obtaining the location of the hand of the individual and setting limits that have made possible a by binarizing the same, eliminating all the image artifacts that are not an integral part of the hand. However, the fact that near the surface this segmentation does not produce satisfactory results, has led to performs a segmentation for the color of skin, converting the RGB image to the color space of HSV and subsequently with the definition of limits of each component of this space, detect only the user's hand through the color of their skin. We managed to get to the intersection of these two segmentations, the hand perfectly defined and free of any artifact that does not belong to them.

With the hand obtained, is started a process of extraction of characteristics, with the use of wrapper Emgu CV that would allow us to obtain more information regarding the hand, such as the center of the palm and the location of the fingers, achieving, subsequently, realize where is each party in terms of depth.

Having all the information gathered, was necessary process them in order to realize what action is to be performed and if it is, or not, to be carried out some kind of touch or interaction with the projection. For this reason, it was developed a calibration window which allowed define not only a zone of touch, as also inform the system of the location of the projection on the surface.

With the combination of all these processes and of all the information in relation to the hand, the fingers, and the surface obtained, it has developed a method for detection of the touch in surface, which allows the system realize if the touch has happened, or not, and by means of the characteristics of the hand, as its orientation or position, perform different types of actions, such as for example, the typical left and right click of the cursor.

It was, in this way, giving the user ability to interact directly with the projection through the touch, without any type of physical device.

Agradecimentos

Agradeço ao Professor Doutor Eurico Carrapatoso, bem como ao Professor Doutor António Castro pela orientação dada durante o desenvolvimento deste projeto. Pela ajuda e motivação dada ao longo do tempo que resultou nesta dissertação.

Agradeço, também, a todos os meus colegas e amigos, que comigo se cruzaram durante esta fase da minha vida e da minha aprendizagem, pelo seu contributo por mais pequeno que tenha sido, pois ajudaram-me a ser aquilo que sou hoje. Um especial obrigado, ao meu colega e amigo Miguel Correia, pela partilha de todos os seus conhecimentos, por estar sempre presente para ajudar em todos os momentos e, acima de tudo, pela sua amizade. Um especial agradecimento ainda para os meus amigos e colegas Nuno Soares, João Botelho e Rúben Guedes por todo o apoio e amizade. Ainda o meu agradecimento ao Eduardo Magalhães, pelo apoio dado no decorrer desta dissertação.

Um especial agradecimento à minha família, aos meus pais, ao meu irmão e ao meu avô, por todo o apoio incondicional, pela confiança depositada e por me proporcionarem todas as condições e saberes para hoje ser quem sou. À minha namorada e noiva, Ana Sousa, por todo o amor, apoio, ajuda e compreensão demonstrados durante todos estes ano e por todos os momentos de felicidade partilhados que me ajudaram e ajudam a ultrapassar todos os obstáculos.

Um muito obrigado a todos eles e a todas as pessoas que acreditaram em mim, pois sem eles não teria chegado até aqui.

Rui Costa

"I have been impressed with the urgency of doing.
Knowing is not enough, we must apply.
Being willing is not enough, we must do."

Leonardo da Vinci

Conteúdo

Lista de Figuras

Lista de Tabelas

Abreviaturas e Símbolos

API	Application Programming Interface
CMOS	Complementary Metal–Oxide–Semiconductor
DI	Diffused Illumination
DSI	Diffused Surface Illumination
FTIR	Frustrated Total Internal Reflection
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HSV	Hue, Saturation, Value
ITO	Indium Tin Oxide
LED	Light-emitting Diode
LLP	Laser Light Plane
NUI	Natural User Interface
OpenCV	Open Source Computer Vision Library
PDA	Personal Digital Assistant
RGB	Red Green Blue
RGB-D	Red Green Blue - Depth
SDK	Software Development Kit
WPF	Windows Presentation Foundation
XNA	XNA's Not Acronymed

Capítulo 1

Introdução

Desde o início da sua história e desde que começou a viver em sociedade, o Homem sentiu uma enorme necessidade de comunicar. Como tão bem sabemos, desde os primatas até ao presente a comunicação foi e continua a ser o fator mais importante da evolução humana.

Podemos, então, explicar o desenvolvimento da existência humana através de diferentes etapas no desenvolvimento da comunicação, desde o tempo dos hominídeos, era dos símbolos e sinais, em que a comunicação era feita através de gestos, sons e alguns sinais padronizados, sinais esses que eram passados através de gerações para que fosse possível viver em sociedade. Com o aparecimento do Cro-Magnon, dá-se início a uma nova era no desenvolvimento humano e da sua capacidade intelectual. Chamamos-lhe a era da Fala, que possibilitou ao ser humano a capacidade de se expressar e comunicar com os seus semelhantes. Mais tarde, com o aparecimento da escrita, deu-se início à alfabetização através da padronização do significado de diferentes representações pictóricas. Com os Sumérios a transformarem os sons em símbolos, foi dado o primeiro passo para a escrita fonética, permitindo a cada sociedade criar formas particulares de escrita e fala.

A humanidade assistiu ao longo de sua existência ao desenvolvimento de diferentes meios de comunicação, e essa mesma diversidade introduziu na sociedade novas formas de pensar e ver o mundo, permitindo o aparecimento de meios de comunicação em massa levando-nos a uma nova era na comunicação, que com o avançar da tecnologia se traduziu na era dos computadores, onde a diversidade está presente nos mais variados formatos. Desde então, o Homem com a sua evolução, saiu dos computadores gigantescos e chegou a pequenos dispositivos de comunicação portáteis, que disponibilizam as mais variadas funcionalidades.

Com o avançar da "era dos computadores" novos meios de comunicação começaram a ser adicionados ao quotidiano da população. A introdução dos mais variados mecanismos e acessórios de controlo de dispositivos veio abrir uma nova forma de comunicação. A comunicação homem-máquina, que se tornou parte intrínseca em muitos tipos de componentes eletrónicos, passou a ter um papel fundamental no mundo em que vivemos. Esta nova forma de comunicar continua a sua evolução introduzindo novas formas de interação, sem recurso a nenhum tipo de mecanismo ou acessório, possibilitando a manipulação natural de um sistema através do toque. O desenvolvimento deste tipo de tecnologia levou à criação de sistemas multitoque. Estes consistem numa

comunicação homem-computador por forma a permitir o seu controlo através do reconhecimento de múltiplos contactos, simultaneamente.

A evolução tecnológica que se verifica nos últimos anos traz para a sociedade novos paradigmas relativamente à interação homem-máquina, mergulhando a população em ambientes imersivos, dando primazia à desmaterialização, deixando para trás a dependência que existia com o uso de dispositivos físicos de controlo, tornando o contacto corpóreo e a interação humana o meio de comunicação primordial com dispositivos eletrónicos.

O crescente recurso a novas tecnologias para tornar a utilização de diferentes dispositivos num ato natural e fluido veio abrir um novo mundo na investigação e na engenharia, alargando novos horizontes na forma como se comunica com os mais variados sistemas e acessórios. Toda uma nova forma de interagir com uma máquina começou a ser ato quotidiano tornando o desenvolvimento de novas formas de interação multimodal um campo de interesse e em grande expansão. O aumento de dispositivos que tiram partido das mais variadas formas da comunicação humana, permite atualmente um conjunto de abordagens e propostas que satisfazem os mais diversos utilizadores, conferindo uma liberdade de movimentos mais natural, ultrapassando barreiras e limitações que a utilização e dependência de dispositivos físicos de controlo podem originar.

1.1 Caracterização do Tema

A crescente evolução e o desenvolvimento de dispositivos capazes de uma interação por parte do utilizador através de gestos e toque, trouxe à população uma nova forma de comunicar com a tecnologia. Contudo, quando queremos realizar uma apresentação ou realizar uma demonstração através de uma projeção por forma a existir uma maior área de visualização, perdemos por completo este tipo de interação e comunicação gestual e corporal.

As projeções são efetuadas para superfícies como paredes ou quadros, superfícies essas que não incorporam qualquer meio tecnológico no seu interior e não respondem às nossas ações. Estas projeções apenas podem ser controladas diretamente no computador, impedindo o utilizador de ser mover livremente e interagir de uma forma mais natural, necessitando de dispositivos físicos para proceder às ações pretendidas.

As superfícies interativas existentes atualmente, como quadros ou mesas interativas, são bastante dispendiosas e pouco móveis, uma vez que são volumosas o que torna difícil o seu transporte.

A introdução de novos sistemas operativos, capazes de suportar multitoque e desenvolvidos para esse tipo de interação, veio reforçar o tipo de comunicação homem-computador, pretendendo enraizar na população meios de comunicação corporais e gestuais, formas naturais de interagir com as máquinas e dotar as mesmas com este tipo de interfaces. Contudo, uma das principais lacunas nestes sistemas está relacionada com a sua ligação a projetores, eliminando por completo a interação gestual e corporal do utilizador. Neste projeto pretende-se assim suprimir esta lacuna, com a introdução deste tipo de comunicação e interação com as projeções.

1.2 Objetivos

Com este trabalho pretende-se desenvolver uma aproximação tecnológica à relação homem-máquina, com a introdução de novas formas de comunicação por parte do ser humano. A interação de uma forma natural e sem a limitação de estarmos presos a uma secretária ou a qualquer tipo de dispositivo físico para comunicarmos com o nosso computador torna-se um dos principais objetivos deste projeto, por forma a oferecer ao utilizador uma maior sensação de liberdade, imersão e capacidade de interação.

Com o recurso às tecnologias emergente na área da multimédia e da comunicação, como é o caso do Kinect, pretende-se criar um sistema imersivo e inteligente de interação para o utilizador controlar o seu computador em qualquer superfície, através de uma projeção e com tecnologia de toque. Assim, pretende-se proceder ao controlo do sistema operativo Windows 8, para através de uma projeção da imagem de um computador para uma superfície interagir com o sistema e com o computador diretamente na local onde está a ser realizada a projeção.

Assim, numa fase inicial, será importante estudar e compreender as limitações existentes na necessidade do uso de dispositivos físicos de controlo, para proceder à comunicação com um computador. Teremos de perceber bem quais as maiores dificuldades que advêm deste tipo de comunicação, quando estamos perante uma projeção e é pretendido realizar interação com a máquina, por forma a tornar este processo o mais simplificado possível para o utilizador. É importante o estudo da evolução das telas multitoque existentes no mercado, pois permitirá a compreensão do seu funcionamento, bem como deixar perceber as vantagens e desvantagens deste tipo de tecnologia, retirando com isto as melhores e mais completas abordagens para a implementação do sistema a desenvolver. É, também, importante estudar e compreender os gestos utilizados com maior frequência na comunicação por toque e por multitoque. Existindo uma padronização deste tipo de movimentos para a interação com os mais variados dispositivos, pretende-se perceber de que modo a sua integração no sistema a desenvolver pode facilitar a interação e adaptação do utilizador com este tipo de tecnologia. Por fim todo o *software development kit* (SDK) da *Kinect*, será estudado e aprofundado, permitindo o conhecimento de todas as potencialidades deste tipo de controlador, e qual a forma de retirar o maior partido desta tecnologia.

Pretende-se, então, criar um sistema que permita tirar partido do mais recente software existente no mercado, vocacionado para superfícies táteis e abrir ainda um novo mercado que permita a criação de sistemas específicos de interação que possam ser usados e direcionados para projeções e a sua capacidade de interação. Com recurso a uma projeção e através das mãos, pretende-se com este sistema transformar paredes ou quadros num enorme ecrã tátil.

1.3 Motivação

Este projeto tem como objetivo a transformação de diferentes superfícies num ecrã tátil, com recurso a uma projeção.

A motivação para elaborar este projeto surge do interesse do autor pela área da multimédia e pelo interesse em novas interfaces inteligentes com recurso às mais variadas formas de comunicação (multimodal). O gosto pela tecnologia e inovação contribuíram decisivamente para a escolha do tema. Nesse sentido, a possibilidade de desenvolver um projeto numa área em crescente expansão, como as interfaces inteligente com recursos às mais variadas formas de comunicação, contribuiu decisivamente na escolha desta tese de mestrado.

A possibilidade de permitir ao ser humano uma interação mais natural com um dispositivo, por forma a tornar a sua experiência mais imersiva e facilitada conferindo-lhe liberdade de movimentos sem perca de funcionalidade, contribuindo assim para o desenvolvimento tecnológico na área da computação, oferece uma motivação acrescida no desenvolvimento e elaboração deste projeto.

1.4 Estrutura do Documento

Este documento é constituído por seis capítulos principais.

No capítulo 1, é realizada uma introdução e uma caracterização do tema que será estudado e abordado o problema que se pretende resolver. São apresentados os objetivos desde projeto e a motivação para a realização desta Tese de Mestrado.

No capítulo 2, apresenta-se o estado da arte relativo ao projeto que será desenvolvido. Inicialmente é estudada a interação homem-computador e de seguida são abordados os tipos de ecrãs e superfícies multitoque existente na atualidade.

No capítulo 3, apresenta-se a arquitetura utilizada para se proceder ao desenvolvimento do sistema, é feita uma introdução ao sensor Kinect e explicando todo o seu funcionamento, mais concretamente o das câmaras RGB e de infravermelhos e ainda o seu método de deteção do esqueleto. Posteriormente fez-se uma breve introdução à biblioteca de processamento de imagem utilizada para extração de características das mãos. Por fim apresenta-se a arquitetura de todo o sistema, com um diagrama explicativo.

No capítulo 4, são explicados todos os passos que levaram desenvolvimento do sistema. Inicialmente, são estudados os espaços de cor existentes, mais direcionados à deteção de cor de pele, para perceber qual o melhor método a utilizar quando se pretende este tipo de deteção. Posteriormente, todo o processo de deteção e rastreio das mãos é aprofundado e explicado, e ainda todos os métodos de deteção da superfície de toque, como se procedeu à deteção desses mesmos toques e se interagiu com o computador através os mesmos.

No capítulo 5, são analisados todos resultados obtidos e as decisões tomadas durante o processo de realização do sistema. Justifica-se a cada decisão tomada, o resultado final obtido e são analisados os vários resultados obtidos com os diferentes processos que foram pensados durante o desenvolvimento.

Por fim, no capítulo 6, são tiradas as conclusões ao projetos e analisados os resultados obtidos, de uma forma mais geral, bem como avaliando se os objetivos iniciais foram, ou não, totalmente compridos. É ainda proposto trabalho futuro, para que possa ser melhorado o sistema desenvolvido.

Capítulo 2

Estado da Arte

Neste capítulo será apresentado o estado da arte relativamente à área de realização deste projeto. Inicialmente são estudadas as formas de comunicação com o computador e a interação humano-computador, analisando o humano e as formas de interação existentes. São depois revistas diferentes tecnologias referentes ao tipo de superfícies multitoque existentes. Analisam-se os tipos de ecrãs e superfícies, o seu funcionamento e vantagens e desvantagens. Olha-se, ainda, de uma forma mais específica para os sistemas óticos e para os gestos que permitem interagir com este tipo de superfícies. No final apresenta-se um sumário deste capítulo.

2.1 Interação Humano-Computador

Interação Homem-computador (*Human-Computer Interaction - HCI*), pode ser definido como a disciplina relacionada com o design, a avaliação e a implementação de sistemas de computador interativos para uso humano e o estudo dos fenómenos à sua volta [CH92]. Esta definição não pode, contudo, ser entendida como a única ou a principal, uma vez que não existe uma definição geral para HCI, podendo ser definida de múltiplas formas. Existe sim um principio geral que é a forma como as pessoas usam os computadores para realizar um determinado trabalho. Podem assim ser criadas três classes principais analisando esta interação ao pormenor, as pessoas, o computador e a forma de realizar esse trabalho, mais concretamente a interação pessoa-computador.

Falar em HCI pode levar a pensar em apenas um homem e um computador. Mas HCI é muito mais que isso, podendo ser vista como uma ou mais pessoas a realizar determinadas tarefas, não apenas com um computador, mas com qualquer tipo de tecnologia existente. Por interação entende-se qualquer tipo de comunicação existente entre o(s) utilizador(es) e a tecnologia, de forma direta ou indireta, usada para realizar uma ação com um determinado objetivo.

HCI pode assim ser considerada uma área multi-disciplinar: Psicologia e a Ciência Cognitiva podem oferecer conhecimento cognitivo do utilizador, da sua perceção e, ainda, uma ajuda essencial na resolução do problema; Sociologia para ajudar o utilizador a perceber o contexto da interação; Ciência da Computação e a Engenharia para permitirem a construção de qualquer tecnologia; Capacidades de Negócio para venda do produto [DFAB04].

Falar em interação pode ser compreendido como a realização de uma ação. O modelo de Norman [Nor02], talvez o que mais influencia a HCI, uma vez que oferece uma proximidade com a nossa compreensão intuitiva da interação entre o homem e o computador [NAC+95], divide uma ação em sete fases diferentes, permitindo distinguir os diferentes processos que levam a que uma ação seja realizada com sucesso, sendo estas fases:

1. Estabelecer o objetivo;
2. Formar a intenção;
3. Especificar a sequência da ação;
4. Executar a ação;
5. Perceber o estado do sistema;
6. Interpretar o estado do sistema;
7. Avaliar o estado o do sistema no que diz respeito ao objetivo e às intenções.

Neste modelo, é pretendido que o utilizador formule um plano de ação para depois ser executado pelo computador, traduzindo-se as diferentes fases do modelo em ações que terão que ser seguidas e realizados por parte do utilizador, por forma a executar uma interação com o computador. Este ciclo interativo pode ser divido em duas fases principais, uma de execução e outra de avaliação.

2.1.1 Humano

O Humano, é sem dúvida a personagem central na HCI, uma vez que os computadores são feitos precisamente para prestar auxílio ao Homem. Um dos aspetos mais importantes neste tipo de interação, no que diz respeito ao Homem, é o processo necessário para existir uma interação do homem com o computador.

As ações necessárias em HCI podem ser comparadas com o modelo humano de processamento, descrito por Card, Morgan e Newell, o *Model Human Processor* [CMN86], e consistem em três sub-sistemas, o sistema percetual, motor e cognitivo. O sistema percetual recebe e manipula estímulos sensoriais, o motor controla as ações relativas a esses estímulos e, por fim, o sistema cognitivo onde está presente o processo necessário para a junção dos dois sistemas anteriores, podendo ser feita uma analogia entre o sistema humano e o sistema computacional, em que a informação é recebida, armazenada, processada e posteriormente enviada uma resposta ou despoletada uma ação.

A receção de informação por parte do ser humano é normalmente realizada através do sistema sensorial (visão, audição, tato, paladar e o olfato), provocando uma reação do nosso controlo motor, sendo tudo isto armazenado na nossa memória. Contudo quando se fala em HCI, é dada maior relevância aos sentidos da visão, audição e tato, podendo assim ser equiparado o computador com o ser humano e o seu processo de receção, manipulação e reação à informação recebida.

2.1.2 Interação com o Computador

O fenómeno do crescimento de ambientes virtuais baseados em sistemas computacionais, trouxe à área da tecnologia um sem número de dispositivos físicos capazes de interagir com os computadores, permitindo as mais variadas formas de interação. O desenvolvimento da tecnologia e o avançar da ciência trouxeram mudanças no que diz respeito à HCI, introduzindo nos computadores técnicas de interação baseadas nos sentidos humanos, como a fala e a visão, mas também através de gestos e rastreio corporal. Desmaterializando a comunicação homem-máquina deu-se um passo enorme na forma como interagimos com o computador, tornando essa experiência mais natural, quase imersiva, permitindo-nos tocar naquilo que realmente queremos.

A este dois métodos de interação com o computador, podemos dar o nome de Interfaces gráficas do utilizador (*Graphical User Interface - GUI*) e Interfaces naturais do utilizador (*Natural User Interface - NUI*).

As GUIs correspondem a interfaces baseadas na comunicação homem-computador que requerem dispositivos físicos de entrada de dados, como rato e teclado, para se proceder à interação com os elementos digitais apresentados. Este sistema usa uma combinação de tecnologias e dispositivos para fornecer ao utilizador várias formas de interação.

NUI é um tipo de comunicação homem-máquina intuitiva, que dispensa o uso de qualquer dispositivo, tornando esta comunicação invisível. Baseada na natureza humana, usa o reconhecimento do corpo, dos gestos ou dos sentidos como forma de comunicar com o computador, sendo as ações despoletadas por um simples gesto, ou palavra. Este tipo de comunicação e este tipo de interface oferecem uma interação mais natural ao utilizador, estando presente atualmente nos mais variados computadores e dispositivos tecnológicos.

2.2 Tecnologia de Superfícies Multitoque

A tecnologia multitoque, denominada vulgarmente pelo anglicismo *touchscreen*, apesar de apenas agora se estar a enraizar no quotidiano da população, com a proliferação desta tecnologia em telemóveis, tablets e computadores, existe há mais de 30 anos, desde a década de setenta. Desde então, múltiplas patentes foram registada para a construção deste tipo de superfícies, sendo algumas das mais importantes as de Johnson - *Touch actuable data input panel assembly* [Joh72], a de Kasday - *Touch position sensitive surface* [Kas84], ou a de Mallos - *Touch position sensitive surface* [Mal82].

Buxton, conhecido como um dos pioneiros da interação homem-computador e um dos primeiros investigadores desta tecnologia, reconhece [Bux12] que a primeira superfície multitoque foi criada em 1982, *Flexible Machine Interface* [Met82].

Mais recentemente a Apple, com o lançamento do primeiro iPhone, introduziu no mercado o uso dispositivos móveis com interação multitoque. O desenvolvimento desta tecnologia permitiu a criação de mais e melhores superfícies e dispositivos interativos. Com a Microsoft, em 2007, a apresentar a sua primeira versão de uma mesa multitoque, a *MS Surface* atualmente designada

como *PixelSense* [Mic12b], similar a *HoloWall* desenvolvida por Matsushita e Rekimoto, em 1997 [MR97], sendo este o ponto de partida para um maior desenvolvimento e banalização deste tipo de tecnologia e interação em superfícies.

Esta tecnologia tem a capacidade de detetar a realização de toques numa superfície e está geralmente associada ao toque através dos dedos num ecrã, dispensando, assim, o uso de periféricos de entradas de dados, que são substituídos pelos dedos. Através da deteção da localização de cada toque, este tipo de interação coloca de uma forma literal o controlo dos mais variados dispositivos eletrónicos nas mãos dos utilizadores, provocando uma desmaterialização e um contacto mais natural, com a máquina por parte do utilizador.

2.2.1 Tipos de Ecrãs e Superfícies Multitoque

A evolução da tecnologia multitoque tem permitido ao longo do tempo a construção de ecrãs mais precisos e sensíveis ao toque, com uma maior durabilidade e melhor tempo de resposta. Para a fabricação desde tipo de ecrãs têm sido usadas diferentes técnicas e métodos que permitem a sua evolução e desenvolvimento.

O contributo de Han em 2005, com o princípio *Frustrated Total Internal Reflection (FTIR)* [Han05], um sensor multitoque de alta sensibilidade e resolução por um baixo custo, baseado numa câmara, permitindo reduzir os custos de construção desta tecnologia, abriu a porta para o desenvolvimento deste tipo ecrãs e superfícies por parte das grandes empresas da área.

Atualmente os métodos mais usados para o fabrico de ecrãs e superfícies são: Resistivos, Capacitivos, de Superfície de Onda Acústica e Superfícies Óticas. Todos eles apresentam todas múltiplas vantagens e desvantagens no que que se refere ao toque, como será analisado abaixo.

2.2.1.1 Ecrãs Resistivos

Ecrãs resistivos são compostos por múltiplas camadas, duas das quais são as mais importantes neste tipo de ecrãs. Consistem em duas camadas condutivas revestidas por uma substancia transparente, o *Indium Tin Oxide (ITO)*, de forma a facilitar a condução elétrica e transformando-as em condutores. Uma muito fina e flexível na parte de cima, a que é tocada, e uma rígida de vidro em baixo, separadas por um pequeno espaço.

O funcionamento deste tipo de ecrã é feito quando a camada superior é pressionada, e sendo flexível, entra em contacto com a inferior e as duas passam a ser atravessadas por uma corrente elétrica, como ilustrado na figura 2.1.

A resistividade existente nestas duas camadas faz com que quando se tocam criem um circuito elétrico fechado, que em conjunto com a tensão (5V) aplicada, produz uma mudança no campo elétrico. A mudança é detetada e medida horizontal e verticalmente, sendo as suas coordenadas X e Y enviadas para o controlador do ecrã, que posteriormente envia os dados para o sistema operativo para serem processados. As medidas são obtidas com base em dois tipos de configurações:

- Na ***4-wire configuration***, a camada interna e externa da superfície são utilizadas para estabelecer o local de ocorrência do toque, com a camada superior a ter fios posicionados

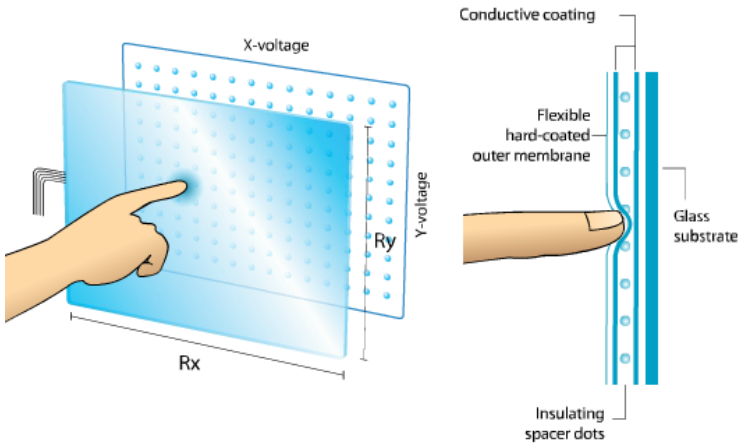


Figura 2.1: Ecrã multitoque baseado em tecnologia resistiva. Adaptada de [Ele].

verticalmente e a inferior horizontalmente. Uma tensão é aplicada em cada uma das camadas, sendo a posição de toque obtida quando o contacto entre elas existe. Este tipo de configuração é a mais comum nestas superfícies; no entanto apresenta uma maior margem de erro com o passar do tempo, relativamente à sua precisão.

- A ***5-wire configuration***, resolve o problema existente na configuração anterior, com o posicionamento vertical e horizontal de fios na camada inferior da superfície, com a camada superior a apresentar apenas num único fio. O seu funcionamento é equivalente à configuração *4-wire*, com a tensão a ser aplicada primeiro vertical e depois horizontalmente, sendo a posição de toque detetada com o contacto entre camadas.

Este tipo de tecnologia tem a vantagem de ter baixos consumos energéticos, poder ser operado tanto com os dedos, como com uma caneta, e ser bastante precisa Em sentido oposto está a claridade existente neste tipo de superfícies (cerca de 75(%) - 80(%)) e a sua durabilidade. É usado em dispositivos móveis, como PDAs ou câmaras digitais.

2.2.1.2 Ecrãs Capacitivos

Superfícies capacitivas (2.2) baseiam-se no conceito da capacidade elétrica que ocorre entre dois condutores, posicionados em locais próximos, com a capacidade a ser determinada pelo dielétrico existente entre a superfície de toque e os dedos do utilizador. As propriedades da pele humana, que contém eletrólitos condutores, tornam-se parte fundamental neste tipo de tecnologia.

A vantagem da utilização deste tipo de tecnologia é a sua clareza, precisão, durabilidade e fiabilidade. No entanto é dispendiosa a sua construção. São usadas em sistemas simples, como *kiosks* ou controlos industriais.

Este tipo de superfícies pode ser dividido em duas classes: Capacitivas de Superfície e Capacitivas Projetadas.

Capacitivos de Superfície

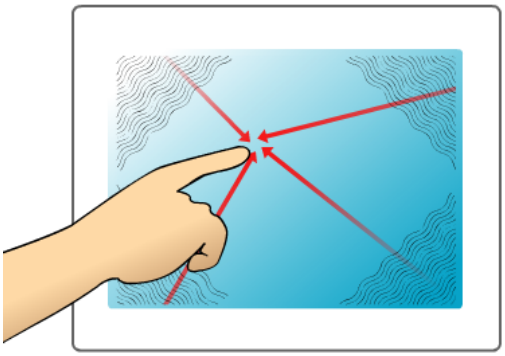


Figura 2.2: Representação de uma ecrã multitoque baseado em tecnologia capacitiva. Adaptada de [Ele].

Este tipo de ecrã é composto por um revestimento sobre uma camada de vidro capaz de armazenar cargas elétricas. Quando este tipo de ecrã é tocado, uma pequena quantidade de carga elétrica é transferida para o objeto de contacto.

O funcionamento deste tipo de ecrã é feito com a medição da variação da carga que é aplicada nos cantos do monitor. Apresentando inicialmente um campo elétrico constante por toda a camada condutora, quando tocado, uma carga é transferida da superfície para os dedos, resultando num direcionamento da corrente aplicada. Com o uso de sensores nos cantos do ecrã o valor da corrente transferida para os dedos é medida e as medições efetuadas são enviadas para o processador de controlo. Com base nas diferenças relativas de carga medidas em cada um dos cantos da superfície o sistema calcula com precisão o local onde foi realizado o toque no ecrã.

Contra esta tecnologia existe o facto de apenas ser possível realizar toques com objetos condutores, como os nossos dedos, podendo ainda ter um limite no reconhecimento do número de toques em simultâneo e o seu desempenho ser afetado pela existência de condutores nas proximidades da superfície.

Capacitivos Projetados

Este tipo de ecrã é mais dispendioso de ser construido, contudo, apresenta uma resistência mecânica superior. O funcionamento desta tecnologia é explicado por Rekimoto como sendo uma rede muito fina de fios de microfone, instalados entre duas camada protetoras e vidro [Rek02]. Nesta tecnologia é aplicada uma carga a todas as linhas e colunas existentes e medida a sua capacidade, de forma a encontrar perturbações.

Através da condutividade existente na pele humana, quando a superfície é tocada, é medida a capacidade elétrica entre o dedo e a rede de sensores, através dos valores medidos de corrente transferidos para os dedos pelos sensores existentes nos cantos da superfície é calculada a posição de toque no ecrã horizontal e verticalmente.

Os ecrãs projetados, sendo muito semelhantes aos de superfície, apresentam uma transmissão de luz bastante superior. Oferecem, ainda, a possibilidade de não utilização de apenas objetos condutivos para interação direta e permitem uma maior capacidade de reconhecimento multitoque.

2.2.1.3 Superfície de Onda Acústica

Este tipo de tecnologia é um das mais avançados na atualidade para a construção de ecrãs multitoque. Baseia-se em dois transdutores, posicionados horizontal e verticalmente junto do ecrã e um refletor que é colocado sobre o mesmo, conceito proposto por Adler and Desmares [AD86]. O seu funcionamento baseia-se no envio de um sinal elétrico por parte do controlador para o transdutor de emissão, este, por sua vez, converte o sinal em ondas ultrassónicas, sendo transmitidas e alinhadas ao longo da extremidade do ecrã. O transdutor recebe posteriormente as ondas refletidas convertendo-as num sinal elétrico que depois é passado ao controlador. Quando um toque acontece as ondas são absorvidas pelos dedos atenuando o sinal medido pelo recetor. Este é posteriormente processado de forma a ser obtido o ponto de contacto.

Este tipo de superfície permite obter uma imagem com uma maior resolução e claridade, apresentando uma durabilidade bastante elevada, uma vez que todo o ecrã é feito de vidro. Tem, no entanto, a desvantagem de reconhecer apenas dedos, luvas ou objetos macios e de ser uma tecnologia bastante dispendiosa. É recomendado para ambientes exteriores.

2.2.1.4 Superfícies baseadas em Sistemas Óticos

Sistemas óticos são a forma mais usada para a construção de mesas e superfícies multitoque, sendo a abordagem para este tipo de ecrãs a utilização de um emissor de infravermelhos e de uma câmara modificada por forma a captar a luz emitida.

Ecrãs óticos são compostos por uma matriz X-Y de emissores *Light Emitting Diodes (LED)* de infravermelhos e um par de foto-detetores em torno do ecrã. Perturbações no sistema de feixes de LEDs são detetados e passados para o controlador do sistema, por forma a localizar a origem das mesmas, que resultam do toque no ecrã.

As principais técnicas utilizada nesta tecnologia multitoque são essencialmente *Frustrated Total Internal Reflection (FTIR)*, *Diffused Illumination (DI)*, *Diffused Surface Illumination (DSI)* e *Laser Light Plane (LLP)*, sendo este tipo de superfície abordado de uma forma mais específica na secção seguinte deste capítulo.

2.2.2 Sistemas Óticos

Esta tecnologia, como referido na secção 2.2.1.4, é a mais usada na construção de mesas e superfícies multitoque de maior dimensão. Não apresenta limites no número de pontos de contacto e pode interagir e reconhecer objetos que estejam direta ou indiretamente sobre a superfície.

Baseiam-se num sistema constituído por um emissor de infravermelhos, tirando partido do facto de este tipo de luz ser invisível para o olho humano, e ainda uma câmara adaptada com um filtro de forma a captar apenas este tipo de luz. Permite, com isto, a captação de imagens de alta resolução com um número elevado de *frames* por segundo, proporcionado uma interação em tempo real de forma muito eficaz e precisa. Os elementos apresentados no ecrã de toque, para serem alvo de interação por parte do utilizador, são realizados sobre a forma de projeções diretamente na superfície de toque.

A construção deste tipo de superfícies inclui diversos componentes que quando incorporados no mesmo sistema resultam em mesas multitoque. Estas são de uma forma geral iguais, independentemente do tipo de superfície a ser criada, sendo constituídas por: emissores de luz infravermelha, câmaras, filtros, ecrãs de projeção e superfícies compatíveis com silício.

Com este sistema de feixes de luz infravermelha e a sua visualização por parte da câmara existente na parte inferior da superfície, cada toque no ecrã irá originar um "ponto brilhante" que quando captado pela câmara, será processado e permitirá aos sistema detetar o ponto de contacto do utilizador com a superfície.

2.2.2.1 Tipos de Superfícies Óticas

Este tipo de sistemas apresenta múltiplas abordagens, no que se refere ao posicionamento do emissor infravermelho e a forma como este é transmitido e visualizado na câmara existente, sendo elas *Frustrated Total Internal Reflection (FTIR)*, *Diffuse Illumination (DI)*, *Diffused Surface Illumination (DSI)* e *Laser Light Plane (LLP)*.

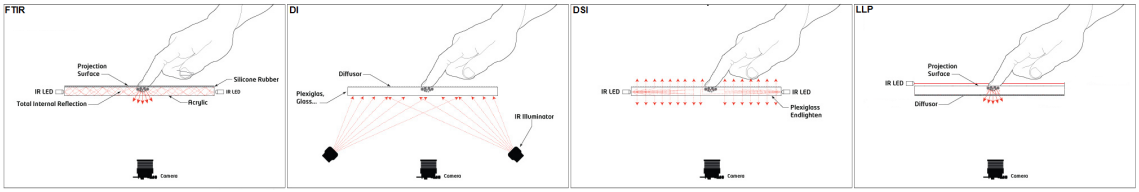


Figura 2.3: Tipos de superfícies óticas. Retiradas de [Rot05].

Frustrated Total Internal Reflection

A *Frustrated Total Internal Reflection*, redescoberta com Han em 2005 [Han05], pode ser vista como o ponto de partida para os sistemas multitoque existentes na atualidade. Este tipo de sistema funciona com base no princípio *total internal reflection* que acontece quando um feixe de luz atinge a extremidade de um meio e encontra outro com um menor índice de reflexão. A deteção do toque depende do ângulo de incidência da luz. Se o ângulo da luz refletida não for demasiado acentuado, a superfície não está a ser tocada, podendo-se assumir a existência de uma *total internal reflection*; contudo, quando existe um toque no sistema o feixe de luz é refletido com um ângulo bastante acentuado resultando no que pode ser chamado *fustrated total internal reflection*.

Este sistema baseia-se, então, na reflexão ótica total interna dentro de uma superfície com um grande índice de reflexão. Geralmente são construídos com um painel acrílico transparente e no seu interior possuem um conjunto de emissores LED infravermelhos.

Quando o acrílico é tocado, a luz infravermelha é refletida, criando um "ponto brilhante", captado de uma forma muito precisa pela câmara que está posicionada na parte inferior da superfície. As coordenadas horizontal e vertical deste "ponto brilhante" são obtidas e processadas, sendo detetado, desta forma, o ponto preciso onde a superfície foi tocada.

Vantagens:

- Não é necessário um caixa fechada;
- Toques apresentam um contraste forte;
- Permite variação de pressão de toque;

Desvantagens:

- Necessita de uma moldura de LEDs, soldados;
- Requer um superfície apropriada (com silício);
- Não reconhece objetos;
- Não pode ser usada uma superfície de vidro.

Diffused Illumination

A *Diffused Illumination* é uma tecnologia muito semelhante à FTIR, no entanto, neste tipo de superfície o emissor de infravermelhos é posicionado atrás da superfície de projeção, tal como a câmara de captação da luz emitida. Este posicionamento permite ao sistema detetar não só toques na superfície, como também objetos existentes na área da mesma. Isto deve-se ao facto de a luz ser difundida por toda a superfície, permitindo encontrar formas através da mancha que vai provocar. Todo o processo de deteção de toque é semelhante ao existente no método FTIR.

Vantagens:

- Não necessita de uma superfície apropriada, funcionando com uma projeção;
- Pode ser usada qualquer material transparente, incluindo vidro;
- Não necessita de uma moldura de LEDs;
- Sem soldaduras;
- Instalação simples;
- Consegue rastreio de objetos e dedos sobre a superfície.

Desvantagens:

- Difícil obter uma iluminação uniforme;
- Toques no ecrã apresentam pouco contraste;
- Possibilidade de haver "falsos toques";
- Necessita de uma caixa fechada.

Diffused Surface Illumination

A *Diffused Surface Illumination* é mais uma tecnologia semelhante ao FTIR; contudo, este tipo de superfície funciona através de uma distribuição uniforme da luz infravermelha emitida por toda a sua área. Nesta tecnologia, Tim Roth [Rot05] propõe o uso de um acrílico especial que funciona como um conjunto de espelhos que refletem a luz emitida de uma forma uniforme por todo o ecrã. Todo o processo de deteção de toque é semelhante ao existente no método FTIR.

Vantagens:

- Não necessita de uma superfície apropriada;
- Pode alternar facilmente, entre modo DI e FTIR;
- Consegue rastreio de objetos e dedos sobre a superfície;
- Sensível à pressão;
- Iluminação constante por toda a superfície.

Desvantagens:

- Utiliza um tipo de acrílico especial mais caro;
- Toques no ecrã apresentam pouco contraste.

Laser Light Plane

A *Laser Light Plane* é uma tecnologia, como todas as outras existentes neste tipo de superfície, baseada na FTIR. Este tipo de ecrã usa para a deteção do toque um emissor de infravermelho que é colocado na parte superior da superfície, paralelamente ao acrílico. Quando tocado este feixe de luz infravermelha é interrompido, existindo uma dispersão de luz que é captada pela câmara existente, para posteriormente ser processado. As coordenadas horizontais e verticais são encontradas e passadas ao sistema.

Vantagens:

- Não necessita de uma superfície apropriada;
- Pode ser usado qualquer material transparente, incluindo vidro;
- Não necessita de uma moldura de LEDs;
- Não necessita de uma caixa fechada;
- Instalação simples;
- Mais barata que outras técnicas.

Desvantagens:

- Não consegue detetar objetos;
- Não é sensível a pressão de toque;
- Pode existir oclusão de toques, se apenas forem usados 1 ou 2 lasers.

2.2.2.2 *Pipelines* para Superfícies Óticas

O rastreamento de toques em superfícies envolve um *pipeline* de operadores de processamento de imagem que transformam a câmara de captação de luz infravermelha num dos principais intervenientes neste tipo de ecrãs.

***Pipeline* do rastreamento FTIR**

Neste tipo de superfícies a realização do rastreio do toque é realizado através de um pré-processamento da imagem captada pela câmara existente no sistema para a remoção de partes

inalteradas a cada *frame* da imagem, subtraindo o *frame* atual com o seu anterior, conseguindo, desta forma obter alterações na imagem.

Posteriormente, são procuradas regiões brilhantes na imagem pré-processada, através de um *connected component algorithm* [HW90], resultando a região brilhante encontrada por este algoritmo na zona da superfície que foi tocada.

Um pós-processamento é efetuado, uma vez mais com a comparação de diferentes *frames* da imagem ao longo do tempo à procura de toques semelhantes. No final é realizada um relação entre as coordenadas da câmara e as da superfície.

***Pipeline* do rastreamento DI**

Neste tipo de superfícies a realização do rastreamento é mais complexo, relativamente ao FTIR, uma vez que este tipo de ecrã deteta não só os toques na superfície como também objetos na sua proximidade. Assim, o *pipeline* que é seguido neste rastreio apresenta uma divisão depois do pré-processamento inicial da imagem captada para a remoção das partes inalteradas. Assim, a imagem pré-processada divide-se em duas fases, uma para a localização de objetos na proximidade, que são apresentados como manchas e outra para os toques na superfície, onde é adicionado um filtro passa-alto, por forma a detetar os pontos brilhantes resultantes do algoritmo aplicado para a deteção dessas regiões.

Um pós-processamento é efetuado com a comparação de diferentes *frames*, ao longo do tempo e do espaço, para a localização de toques semelhantes, resultando numa associação entre os dedos e as mãos na superfície de toque.

2.2.3 Gestos em Interações Multitoque

Esta é uma das áreas mais importante no que ao desenvolvimento de interfaces deste tipo diz respeito. O toque na superfície ou ecrã não só passa a indicar o ponto de interesse do utilizador, como também pode invocar ou iniciar uma ação, passando a ser o principal meio de interação entre o utilizador e o objeto com o qual se pretende interagir.

Este tipo de interação envolve um complexo sistema que pode incorporar uma multimodalidade, como a visão, o tato ou a audição. Nestes sistemas o *feedback* recebido após a realização de uma ação torna-se fundamental para o utilizador, sendo muito importante providencia-lo, por forma a perceber quando essa ação é, ou não, ativada. O seu tempo de resposta é outro dos fatores mais importantes, uma vez que o retardar das ações por parte do sistema pode eliminar por completo a fluidez do sistema e retirar ao utilizador a sensação de interação direta com os objetos.

Este tipo de sistema consegue desenvolver um tipo de interação mais natural e fluída para o utilizador, juntando os movimentos físicos da natureza humana com um sistema de computador, de forma a facilitar o seu uso. A fluidez existente atualmente deve-se ao facto de existir na área dos gestos multitoque, um conjunto de movimentos que são implementados na quase totalidade deste tipo de sistemas interativos, que recorrem à tecnologia de toque, pretendendo-se na sua conceção movimentos que sejam previsíveis e consistentes com os movimentos naturais do Homem, para o tipo de ações mais usadas quando se interage nestas superfícies ou ecrãs.

As interações multitoque estão presentes nos mais variados tipos de NUIs. Para a introdução desta tecnologia é desenvolvido um sistema gestual 2.4, que pode ser dividido em três partes essenciais: o sensor, o comparador e o atuador.

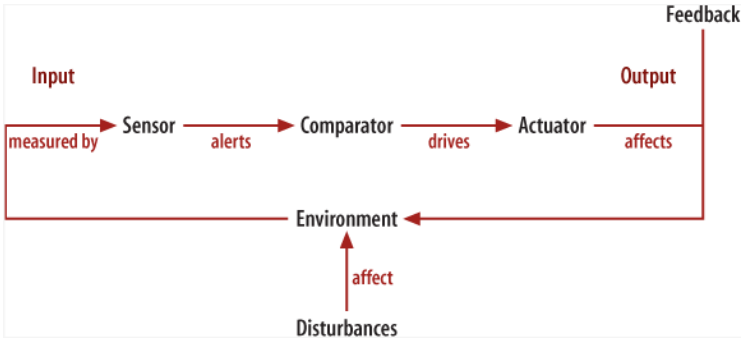


Figura 2.4: Sistema gestual. Retirada de [Saf08].

O sensor é normalmente um dispositivo ou um componente eletrónico capaz de detetar mudanças no ambiente em que está inserido. Para interação através de gestos, normalmente são usados sensores de pressão, luz, proximidade, acústicos, movimento, orientação, entre outros. Não sendo o gesto ou o tipo de gesto reconhecido pelo sensor, não é possível passar à fase seguinte deste, nem completar o ciclo de processamento que irá resultar no *feedback* transmitido ao utilizador. O sensor torna-se assim a parte principal destes sistemas.

Uma vez o gesto detetado, a informação é passada para o que podemos chamar comparador. Nesta fase é comparado o estado atual do processo com o estado anterior ou com o objetivo do sistema. Posteriormente, é feito um julgamento sobre o procedimento a ser executado por parte do dispositivo.

As decisões tomadas no comparador são depois passadas para o atuador na forma de um comando, que, como o nome indica, irá atuar sobre o sistema, de forma a realizar a ação que o gesto efetuado pretendia despoletar.

2.2.3.1 Atributos dos Gestos

Em interfaces interativas com multitoque, é importante perceber as características de um gesto por forma a despoletar uma ação. Para isto, são definidos um conjunto de atributos que se tornam semelhantes em dispositivos ou superfícies desde tipo:

Presença - é o mais básico desses atributos, uma vez que alguma coisa tem que estar presente, para existir um resposta ou haver um reação;

Duração - todos os gestos são realizados durante um determinado tempo, sendo este medido com o cálculo do primeiro impacto ou movimento até ao término do mesmo;

Posição - onde o gesto está a ser feito; em HCI é determinada pela localização no espaço bi ou tridimensional e representado em coordenadas x, y (e z);

Movimento - em alguns sistemas é muito importante uma vez que é ele que dá início a uma ação;

Pressão - em superfícies multitoque pode desempenhar um papel importante pois pode estar na origem da realização, ou não, de uma ação;

Tamanho - largura e altura podem ser combinadas para medições ou distinção entre um toque de um dedo e um objeto;

Orientação - torna-se importante para saber a direção do utilizador enquanto está a realizar um gesto ou toque; tem que ser determinada usando pontos fixos;

Inclusão de objetos - em sistemas mais complexos o reconhecimento de objetos permitem interação dos mesmo com o sistema; outros sistemas "reconhecem"os objetos como sendo uma extensão do corpo do utilizador;

Número de pontos/combinações de toque - em sistemas NUI permite realizar reconhecimento multitoque e múltiplos contactos simultâneos por forma a ser realizada uma única ação;

Sequência auxilia no uso de combinações multitoque para a realização de ações.

Existe ainda um conjunto de movimentos que se referem a determinadas ações, como é o caso do *tap*, *double tap*, *drag*, *flick*, *pinch*, *spread* e *rotate*, que estão uniformizadas em todos os dispositivos multitoque existentes. Contudo, muitos gestos utilizados não são tão intuitivos por não estarem presentes no quotidiano do utilizador, requerendo uma aprendizagem da sua parte.

2.3 Sumário

Neste capítulo foi realizado um estudo das temáticas necessárias para o desenvolvimento deste projeto: a interação-humano computador, os tipo de ecrãs e superfícies. Abordaram-se as formas de interação entre o Ser Humano e os computadores, para compreender como é que o Homem comunica com a tecnologia, com o sistema sensorial humano e os sistemas percetual, motor e cognitivo a serem os principais meios que permitem a interação com os sistemas GUI e NUI existentes.

Foi feita uma descrição dos diferentes tipos de ecrãs existentes nos dispositivos que permitem multitoque, como são o caso dos Resistivos, Capacitivos e Onda Acústica, descrevendo os seus métodos de funcionamento e as principais vantagens e desvantagens que cada uma destas tecnologias representa.

Foram também estudados, de uma forma mais aprofundada, os Sistemas Óticos, a tecnologia existente nas superfícies e mesas interativas. Os diferentes tipos de superfícies óticas, FTIR, DI, DSI e LLP, o que as diferencia, os seus métodos de rastreio de toque, de construção e as vantagens e desvantagens de cada um destes sistemas foram também objeto de estudo. São abordados, ainda, de forma sucinta os *pipelines* de funcionamento dos sistemas FTIR e DI.

Por fim, um estudo dos gestos utilizados neste tipo de interações foi também realizado, permitindo perceber os atributos necessários para a implementação dos gestos e toques nas interações multitoque.

Capítulo 3

Ferramentas e Arquitetura

Neste capítulo pretende-se apresentar a arquitetura e as ferramentas necessárias para a criação do sistema que se pretende desenvolver. O funcionamento do sistema de entrada de dados que será utilizado, o sensor Kinect, será descrito bem como os três tipos de dados que serão adquiridos para o sistema proposto: imagem de cor, informação de profundidade e rastreio do esqueleto. São, ainda, dados alguns exemplos de projetos que foram desenvolvidos com o sensor, de forma a perceber a enorme flexibilidade e potencial do Kinect. De seguida é explicada a biblioteca de processamento de imagem que será utilizada para se proceder à extração de características das mãos do utilizador, que serão obtidas e segmentadas num pré-processamento. É descrita a arquitetura de todo o sistema realizado, com uma explicação das diferentes fases do processo de desenvolvimento. Por fim, apresenta-se um sumario deste capítulo.

3.1 Microsoft Kinect

O Kinect é um sensor de movimento desenvolvido pela *Microsoft*, apresentado em 2010, inicialmente para a a sua consola de jogos *Xbox 360* e na atualidade aperfeiçoado para o seu sistema operativo, o Windows. Tem, como função principal, a produção de dados tridimensionais [WA12]. O processo de aquisição de dados denomina-se RGB-D e pode ser entendido como a captação de uma imagem com cor, sendo realizada uma medição da sua profundidade, com técnicas de luz estruturada.

Construído com um estilo semelhante ao de uma *webcam*, este dispositivo veio introduzir no mundo dos vídeo jogos a possibilidade de um controlo mais natural, com gestos, voz e através de movimentos corporais por parte do utilizador das consolas, e na atualidade dos computadores. Remove, assim, qualquer tipo de dispositivo físico da interação homem-computador.

O sensor Kinect permite ao computador "sentir" diretamente a terceira dimensão do jogador ou do ambiente, a profundidade. Reconhece a fala, o andar e o aproximar e afastar do utilizador. Consegue ainda entender os seus movimentos, traduzindo-os de forma a serem interpretados pelo ambiente virtual apresentado [Zha12].

A interação natural que o desenvolvimento e construção deste dispositivo trouxe para o quotidiano, criando um mundo de oportunidades para a área da computação e da multimédia, permite afirmar que a construção do Kinect veio revolucionar a forma como as pessoas jogam e as suas experiências de entretenimento [Zha12]. O Kinect tornou-se, assim, um avanço enorme na tecnologia computacional, na interação homem-computador e na área do processamento de imagem, oferecendo dados de natureza diferente à que existia até à data da sua criação, combinado geometria com atributos visuais [CLV12]. Tornou também possível a sua inclusão nas mais variadas áreas, como a ciência computacional, a engenharia eletrotécnica e a robótica.

A interação humano-computador através do Kinect, de uma forma fluída e natural, com recurso à natureza humana, ao seu corpo e a fluidez de movimentos existentes, gestos e voz, trouxe um elemento chave para o entendimento da linguagem humana por parte do computador. Adicionou, também, um nível de complexidade bastante elevado ao processamento computacional, uma vez que inicialmente é necessário o computador perceber o que o utilizador está a fazer antes de poder apresentar uma resposta [Zha12].

3.1.1 O Sensor Kinect

O sensor Kinect consiste, de uma forma geral, numa câmara RGB, num sensor de profundidade, num conjunto de microfones e num acelerómetro, como pode ser observado na figura 3.1, de forma a captar o movimento tridimensional e realizar reconhecimento facial e vocal. [Mic12a]

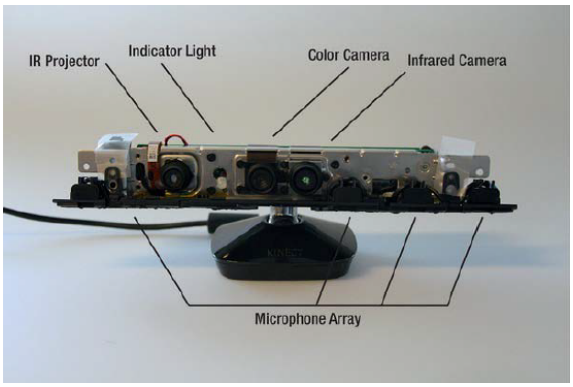


Figura 3.1: Componentes do Kinect para Windows, Retirada de [WA12].

Mais especificamente, a câmara RGB presente suporta uma resolução de 640×480 pixels a 30fps, até 1280×960 pixels a 12fps, com um filtro de cor *Bayer* [Bay76] e é usada para a captação das imagens de cor, enquanto que a câmara usada para medição de profundidade, é de infravermelhos e suporta uma resolução máxima de 640×480 pixels a 30fps.

O sensor de profundidade [FSMA08] é constituído por um projetor de luz infravermelha combinado com uma câmara de infravermelhos, que é um sensor *Complementary Metal–Oxide –Semiconductor (CMOS)* monocromático, capaz de obter imagens tridimensionais em todas as condições de luz ambiente. O feixe de luz emitido pelo projetor de infravermelhos é passado por uma grade de difração, que faz com que a luz emitida se transforme em pequenos pontos, que

são depois captados pela câmara. Sendo a profundidade medida pela distancia entre o sensor e o objeto.

Quatro microfones, com cancelamento de ecos e supressão de ruído, posicionados de forma a conseguir obter uma captação sonora de todo o ambiente onde se encontra o dispositivo e obter um direcionamento da fonte sonora.

Existe, ainda, um acelerómeto com 3 eixos, configurado para uma variação 2G, em que G representa a aceleração devido à gravidade, possibilita assim determinar a cada instante a orientação do sensor Kinect.

O Kinect apresenta também, um campo de visão em forma de pirâmide o que incorpora algumas limitações. Permite o reconhecimento de objetos ou utilizadores de forma mais precisa entre os 40cm e os 4m, como pode visto na figura 3.2, tendo um ângulo de visão 57º graus na horizontal, 43º na vertical.

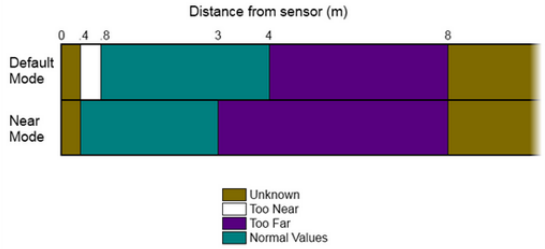


Figura 3.2: Campo de visão do Kinect. Retirada de [fWT12].

3.1.2 Kinect - *Software Development Kit (SDK)*

O SDK do Kinect é um conjunto de bibliotecas específicas para o desenvolvimento de aplicações que usem o sensor Kinect como dispositivo de entrada de dados. Com isto, é possível o desenvolvimento aplicações *Windows Presentation Foundation (WPF)*, *XNA's Not Acronymed (XNA)*, como os mais variados tipos de aplicações que tirem partido de reconhecimento corporal e vocal.

Atualmente, com o lançamento do Kinect para Windows foi lançado um novo SDK, melhorado e mais preciso, incluindo melhorias substanciais no que diz respeito ao reconhecimento e rastreio do esqueleto, ao reconhecimento de objetos próximos do sensor com a introdução de um *Near Mode*, que permite rastreio mais próximo do sensor. Atualizações e melhorias da *Application Programming Interface (API)*, assim como resolução de problemas existentes relativamente a áudio, tempos de execução e estabilidade.

3.1.3 Processamento da Imagem em Profundidade

O Kinect tira partido do sensor de profundidade existente para através de técnicas de processamento de imagem realizar a deteção de formas e objetos presentes numa imagem. Consegue,

ainda, rastrear os esqueletos dos utilizadores e encontrar objetos, realizando a distinção entre os seus utilizadores e os objetos circundantes.

Como referido na secção 3.1.1, o sensor de profundidade do Kinect funciona através da emissão de um feixe de luz por parte de um emissor de infravermelhos, que é passado por uma grade de difração e resulta em pontos que são captadas pela câmara, como observado na figura 3.3.

Figura 3.3: Pontos captados pela câmara de infravermelhos do Kinect. Retirada de [Pli12].

A geometria relativa existente entre o emissor e a câmara de infravermelhos, tal como o padrão de pontos infravermelhos projetados, são conhecidos. Através disto, ao observar um ponto numa imagem e combiná-lo com um ponto no padrão do projetor, torna-se possível a construção dos objetos numa representação tridimensional, usando triangulação e, ainda, a medição da distância entre o objeto e o sensor, através de uma triangulação da informação referida. Sendo o padrão de pontos relativamente aleatório, a combinação entre a imagem obtida pela câmara e o padrão do emissor pode ser feito através de técnicas de processamento de imagem, como é o caso da correlação cruzada [Zha12].

Esta técnica de processamento de imagem resulta em imagens como a da figura 3.4, em que diferentes profundidades são representadas por diferentes valores de cinzento, sendo os tons mais escuros os mais próximos do sensor. Algumas zonas, porém, aparecem a preto, existindo devido ao facto de o Kinect apresentar limitações no seu campo de observação, podendo os locais representados a preto estar demasiado próximos ou afastados do sensor . Podem, ainda, existir devido a serem locais ou objetos com pouca capacidade de reflexão de luz infravermelha, ou por estarem como sombras, e não estarem ao alcance do emissor de infravermelhos.

Figura 3.4: *Frame*da imagem de profundidade do Kinect. Retirada de [WA12].

A imagem em profundidade pode ser representada por 16bits, onde os três primeiros bits correspondem ao índice do utilizador que está a ser detetado, sendo zero se este pixel não pertencer a um individuo, e os restantes representam a profundidade do pixel correspondente, figura 3.5.

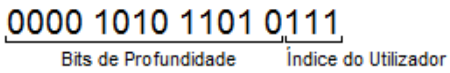


Figura 3.5: Bits de um pixel de uma imagem em profundidade. Adaptada de [WA12].

3.1.4 Rastreio do Esqueleto

Rastreio do esqueleto é o processamento da informação da profundidade de uma imagem, para estabelecer o posicionamento das articulações do esqueleto humano [WA12], permitindo extrair informação relativamente precisa de onde se posiciona cada parte do corpo num espaço tridimensional, como, por exemplo, cabeça e mãos em cada instante.

Este rastreio é realizado através de uma segmentação da imagem de profundidade através de classificações *per-pixel* do corpo humano e, de seguida através do posicionamento das articulações, é calculada uma aproximação do centro de massa. No final, mapeia-se as articulações aproximadas com as articulações do esqueleto considerando continuidade temporal e o conhecimento dos dados do esqueleto anterior.

É possível obter-se, com este método de rastreio, uma representação aproximada das diferentes partes do corpo humano e as suas articulações, como observado na figura 3.6 sendo cada articulação representada pelo seu posicionamento, ou seja, as suas coordenadas no espaço tridimensional.

Este processo permite um rastreio preciso de cada movimento de cada parte do corpo de uma forma individual, traduzindo os gestos realizados em interações definidas num determinado sistema e tornando o objetivo principal deste processo, o reconhecimento em tempo real de todas as coordenadas no espaço tridimensional de cada articulação do corpo humano.

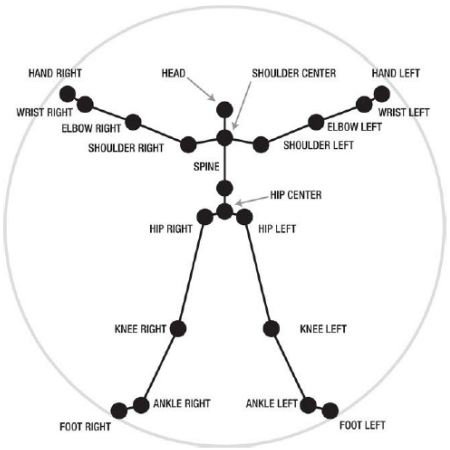


Figura 3.6: Esqueletização das articulações reconhecidas pelo Kinect. Retirada de [WA12].

O sensor Kinect, tem então a capacidade de proceder à deteção de um máximo de seis utilizadores em simultâneo, contudo, apenas dois desses seis têm a possibilidade de interação direta e em simultâneo, com os programas desenvolvidos. Permitindo o rastreio até 20 articulações no seu modo normal (10 superiores e 10 inferiores), e até um máximo de 10 articulações (as superiores), com o *Near mode* ativo.

3.1.5 Gestos e Rastreio das Mãos

Os gestos são uma parte fundamental para o Kinect, assim como os *clicks* são para os GUIs e os *taps* para as interfaces multitoque [WA12]. O gesto, o principal meio de comunicação entre o Homem e o computador, pode mesmo ser definido, segundo Kurtenbach, como: *"O movimento do corpo que possui a informação. Acenar um adeus é um gesto. Carregar na tecla de um computador não é um gesto porque o movimento do dedo no seu caminho para a tecla não é importante nem tem significado. Tudo o que importa é a tecla premida"* [KH90].

As pessoas podem interagir com o Kinect com o seu corpo e mais especificamente com as mãos, por isso saber o seu posicionamento e perceber o que estão a realizar é parte importante para o desenvolvimento de qualquer aplicação neste sistema. Com esta base o Kinect foi desenvolvido e pensada para serem utilizados gestos para a realização das mais diversas interações e ações, desenvolvendo uma nova forma de comunicação do homem com o computador. Contudo, a Kinect não inclui um detetor de gestos de raiz, sendo os gestos implementados e definidos por quem desenvolve aplicações.

3.1.6 Projetos desenvolvidos com Kinect

O Kinect tem sido usado nas mais variadas áreas, nomeadamento na medicina, no cinema, na engenharia ou na robótica, sendo usado também para fins lúdicos, educativos e comerciais. Esta abrangência que o sensor Kinect oferece nas mais diversas disciplinas torna este dispositivo um elemento chave no desenvolvimento tecnológico na atualidade. A sua capacidade de interação com o corpo e mãos, ou através da voz, oferece nas áreas mencionadas uma desmaterialização que permite em muitos casos uma redução de custos e simplificação dos sistemas, tornando-os fáceis de usar.

Esta diversidade é um efeito desejado por parte da Microsoft, como pode ser comprovado com o lançamento do Kinect para Windows para uso comercial e, também com o lançamento de um programa de apoio à realização de aplicações que usem o seu sensor, o *Microsoft Kinect Accelerator* [Biz12]. Este programa apresenta alguns dos melhores trabalhos existentes na atualidade, em diversas áreas de desenvolvimento, como são:

Freak'n Genius

A aplicação permite a animação de personagens, em tempo real, usando o Kinect. Esta aplicação é uma ferramenta de desenvolvimento digital que permite a qualquer pessoa a criação de vídeos animados. Os utilizadores podem, com este sistema, escolher entre um grupo de cenas ou personagens e através do sensor Kinect usar movimentos corporais e

reconhecimento facial e vocal para animarem as personagens e interagirem entre si, por forma a realizarem um vídeo ou uma animação [Gen12].

GestSure Technologies

Este projeto consiste numa interface gestual e de toque para salas de operações. Esta interface permite aos cirurgiões controlar e manipular imagens médicas dos pacientes sem saírem do bloco operatório. Assim, se o cirurgião necessitar de consultar qualquer tipo de exame médico, pode observá-lo e controlá-lo à distância com as suas mãos, com a utilização do sensor Kinect.

Tendo os membros constituintes desta equipa grande experiência, na área da visão computacional, no desenvolvimento de interfaces e em cirurgia, foi criado um sistema que é facilmente integrado num bloco operatório e permite um auxílio eficaz e rápido ao cirurgião [Tec12].

IKKOS

O objetivo deste sistema é permitir aprendizagem de movimentos usando neuroplastia, através de treino. Este sistema, tem como finalidade ensinar atletas, através do treino dos seus cérebros com o fornecimento detalhado de imagens de rastreio corporal. Pretende-se com a repetição de vídeos de rastreio de movimento treinar os cérebros para aprendizagem rápida de movimentos [IKK12].

Jintronix

Pretendeu-se com este trabalho desenvolver um sistema médico para o auxilio na reabilitação física e cognitiva. Este sistema, usa o Kinect e um par de luvas para rastreio do movimento corporal, captando o movimento do paciente e permitindo a sua interação com um ambiente virtual para a realização de exercícios específicos [Jin12].

Skanect

Este sistema consiste num scanner tridimensional *low-cost*, usando o Kinect para capturar modelos em três dimensões de pessoas, objetos ou locais, criando assim imagens digitais. Permite o rastreio de todo o ambiente ou das pessoas de uma forma rápida e barata [Man12].

Styku

Este projeto é um sistema de rastreio corporal para realização de um provador virtual de roupa. Este sistema usa o Kinect para realizar um rastreio corporal, realizando um modelo tridimensional do mesmo. Isto vai permitir a adaptação de uma peça de roupa ao utilizador, através de indicações fornecidas pelo sistema, sendo realizado um ajuste automático da peça ao corpo. Com isto o modelo será produzido de acordo com as especificações corporais de quem pretende realizar a compra *online* [Sty12].

UBI

A aplicação consiste num sistema que permite a transformação de superfícies em ecrãs multitoque. O sistema consiste numa aplicação capaz de conter em si diversas aplicações, que

estão presentes no computador, e permite uma interação com elas nas superfícies onde está a ser projetada a imagem, sendo capaz de reconhecer se a mão do utilizador está apontada, a tocar ou a passar na projeção [Int12].

VoxieBox

O objetivo desta aplicação é oferecer uma maneira simples de construir imagens e vídeos em três dimensões para a utilização por parte de artistas, indústria cinematográfica ou designers de jogos. Este sistema usa o Kinect para a construção de modelos tridimensionais para posteriormente serem inseridos em trabalhos realizados em três dimensões [Vox12].

3.2 Emgu CV

O Emgu CV é, essencialmente, uma enorme biblioteca de funções de processamento de imagem, escrito inteiramente em C#. Mais concretamente o Emgu CV consiste num *wrapper* da biblioteca OpenCV (Open Source Computer Vision Library). Permite a implementação de funcionalidades do OpenCV através do *Visual Studio Windows Forms Application* em linguagens de programação como .NET e C#.

O OpenCV foi projetado especialmente para eficiência computacional e têm enorme foco em aplicações em tempo real, que utilizam processamento de visão por computador. Foi desenvolvido em C/C++ otimizado e permite tirar partido de processamento multi-core. Confere, ainda, um enorme grau de abstração da programação que requer este tipo de processamento.

A inclusão desta biblioteca no desenvolvimento do sistema permite a utilização de técnicas de processamento de imagem para a extração de características relativas às mãos do utilizador, uma vez que o sensor Kinect não apresenta rastreio dos dedos e não permite a extração de informação dos mesmos.

3.3 Arquitetura do Sistema

Como referido anteriormente, este trabalho pretende transformar uma grande superfície num ecrã tátil. Através do rastreio e deteção das mãos do utilizador e da extração de informação de profundidade e localização de uma superfície, pretende-se criar um sistema que satisfaça esse objetivo.

Para isso, foi proposta a arquitetura na imagem 3.7, que diz respeito ao sistema pensado e que será desenvolvido, dividindo-se em cinco processos principais: a aquisição de dados, o pré-processamento, a extração de características, o classificador e o atuador.

A aquisição de dados corresponde às entradas do sistema e resulta na captura de imagens de cor RGB, na obtenção de informação de profundidade e no rastreio do esqueleto do utilizador, sendo nestas componentes que se irá encontrar grande parte da informação necessário para realizar os restantes processos. Toda esta informação é obtida através dos sensores Kinect, que permitem em simultâneo a conjugação destas três componentes. Após a recolha da informação é realizado um

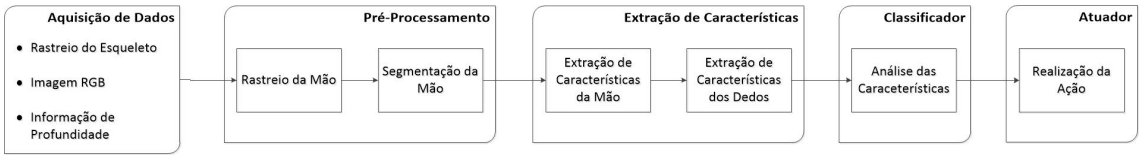


Figura 3.7: Diagrama do sistema desenvolvido.

pré-processamento recorrendo-se ao esqueleto para obtenção da localização e rastreio da mão em tempo real, para posteriormente se proceder à segmentação da imagem nos pixeis correspondentes à sua posição, através da deteção da cor da pele e relativamente à sua profundidade em cada instante, eliminando, assim, artefactos que não sejam parte integrante da mesma.

A obtenção da imagem segmentada permite a extração de características e informações relevantes e é uma das partes principais no sistema. Nesta componente a mão é analisa e são retiradas características como o seu contorno, a sua área convexa e a localização do centro da mão, a sua palma, os dedos e as suas extremidades, sendo retirada a informação relativa às suas profundidades . Os dados são depois processados e analisadas as suas características através de algoritmos, classificando cada ação de uma forma diferente, permitindo perceber em tempo real as ações que estão a ser realizadas por parte do utilizador e, ainda, se o mesmo pretende, ou não, atuar no sistema. É depois passada ao atuador a informação recolhida e analisada anteriormente de forma a que seja comparada a informação processada com a da superfície de toque, despoletando a ação correspondente à pretendida pelo utilizador, traduzindo-se na saída do sistema, ou seja, a interação com a superfície de toque e a resposta do computador.

3.4 Sumário

Neste capítulo abordou-se a arquitetura do sistema desenvolvido, com maior foco na fase de aquisição de dados e da extração de características. Foi inicialmente descrito o processo de funcionamento do Kinect, com todo o processo de aquisição de dados através da câmara RGB que permite resoluções até 1280x960 a 12fps. A obtenção da informação de profundidade obtida através dos 11 bits correspondentes de cada pixel da imagem e a associação com o utilizador a ser realizado pelos 3 primeiros bits, dos 16 que representam cada pixel. A aquisição de dados referentes ao rastreio do esqueleto do utilizador, oferecendo o sensor um capacidade de deteção até um máximo de 6 esqueletos. Foram ainda descritos alguns dos principais programas desenvolvido atualmente com o Kinect.

Posteriormente foi possível compreender como se consegue extrair informação das mãos do utilizador, depois da segmentação das mesmas, com recurso a um *wrapper* para a linguagem de programação C#, o Emgu CV, que permite o uso de técnicas de processamento de imagem da biblioteca OpenCV.

Por fim, a arquitetura do sistema foi explicada, podendo-se facilmente observar que se trata de uma típica arquitetura de um sistema de visão computacional. Com uma fase de aquisição

de dados por parte do sensor Kinect, uma de segmentação da zona de interação pretendida, um módulo de extração de características, um classificador que analisa essas mesmas características e um atuador que se traduz na saída e ações do sistema desenvolvido.

Capítulo 4

Desenvolvimento

Neste capítulo pretende-se abordar todo o desenvolvimento realizado para a criação do sistema proposto para esta tese de mestrado. Inicialmente, os espaços de cor existentes serão alvo de estudo, percebendo-se as vantagens e desvantagens de cada um, para a sua utilização em técnicas de deteção de cor de pele.

Será abordado o processo de deteção e reconhecimento das mãos do utilizador, que leva à extração de características presentes, e todo o processo de análise das mesmas. Posteriormente será analisada a forma de controlar o computador através de uma projeção, com a descrição de todos os processos que permitem diferenciar um toque, ou não, na superfície e descritos os métodos de toque e interação com a superfície. No final apresenta-se um sumário do capítulo.

4.1 Espaços de Cor para Reconhecimento de Cor de Pele

O primeiro cientista a provar que a sensação da luz branca era o resultado da existência simultânea de luzes de várias matizes foi Isaac Newton, através de uma simples experiência sendo que consistiu em fazer incidir um feixe de luz branca sobre um prisma, resultando, e a luz emergente era constituída por um conjunto de diferentes matizes de cor, semelhantes a um arco-íris (figura 4.1).

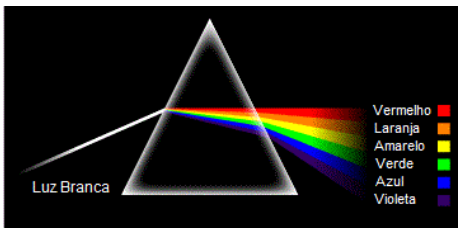


Figura 4.1: Prisma representativo da teoria da decomposição da luz branca de Isaac Newton.

Segundo Foley *et al.* [FDF90], espaço de cor é um sistema tridimensional de coordenadas, onde cada eixo se refere a uma cor primária. A quantidade desta necessária para reproduzir uma determinada cor é a atribuição de um valor sobre o eixo correspondente.

Colorimetria e computação gráfica deram origem a muitos espaços de cor, com diferentes propriedades, sendo muitos deles atualmente aplicados em métodos de deteção da cor de pele em sistemas de visão computacional, tirando partido da separação das componentes da luminância e da crominância. A primeira, é a componente principal de uma imagem que contém a informação do brilho (tons de cinzento) e permite distinguir a sua nitidez e qualidade. Quanto à crominância, esta é a componente que agrega as cores de uma imagem e lhe confere o colorido.

4.1.1 Espaço RGB

Um dos espaços de cor mais utilizados em imagens é o RGB. Este espaço é composto pelas três cores primárias *red* (vermelho), *green* (verde) e *blue* (azul) que quando misturadas com intensidades diferentes produzem cores. É um modelo de cor aditivo, ou seja, na ausência de luz ou de cor surge a cor preta, enquanto que a mistura ou adição das três componentes RGB, com igual intensidade produz a cor branca.

Figura 4.2: Modelo aditivo. Adaptado de [Ram10].

Para a obtenção de uma determinada cor neste espaço é usado normalmente um intervalo pré-especificado de 0 a 255, sendo a cor preta obtida pela combinação RGB (0,0,0) e a cor branca pela combinação (255,255,255). Este espaço de cor pode ser ilustrado por um cubo (o chamado cubo RGB)(figura 4.3), onde nas extremidades estão as cores primárias, as secundárias e ainda o preto e o branco.

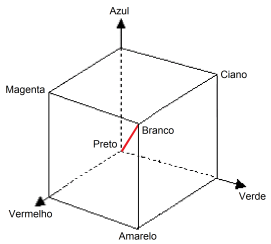


Figura 4.3: Cubo RGB. A linha vermelha representa os vários níveis de cinzento. Adaptado de [Con].

Numa imagem, cada pixel tem o seu próprio valor RGB. Este valor é tipicamente representado por três bytes, uma para cada componente da cor, sendo todos encapsulados num valor inteiro. Assim, uma vez que cada componente da cor é armazenado num byte, é possível obter 2^n cores, em que n representa o número de bits de cada componente. Ou seja, cada componente pode apresentar $2^8 = 256$ intensidade de cor diferentes, o que representa mais de 16 milhões de cores.

Este modelo de cor possui, contudo, uma grande desvantagem. Não suficientemente bom para a definição de cores com base no sistema percetual de visão humano. Isto é, não oferece garantias de que cores próximas no espaço RGB sejam próximas em termos de perceção visual, tornando difícil para sistema de visão computacional determinar se uma determinada cor é de interesse, ou não.

4.1.2 Espaço HSV

Espaços baseados em matiz e saturação (Hue e Saturation) foram apresentados quando houve uma necessidade de se especificar as propriedades das cores numericamente e é caracterizado por ser uma transformação não-linear do espaço de cor RGB . Este espaço mostra valores baseados em ideias de *Hue* (matiz), *Saturation* (saturação) e *Value* (valor). A matiz está relacionada com a cor em si e define a cor dominante da área, permitindo a diferenciação entre cores diferentes. A saturação mede a pureza da cor e é o que permite distinguir entre cores visualmente parecidas, por exemplo, enquanto se pode definir o vermelho como uma cor pura e primária, o rosa pode ser definido com vermelho com alguma quantidade branco. Por fim o valor refere-se à luminância da cor, ou seja, o brilho, permitindo distinguir o claro do escuro em cada cor da matiz.

O facto de os seus componentes serem de fácil perceção e de existir uma separação entre as propriedades da luminância e da crominância, torna este espaço de cor bastante utilizado em trabalhos de segmentação de cor de pele [ZSQ99]. Contudo, a descontinuidade da matiz e o elevado processamento no tratamento da luminância das imagens pode ser prejudicial quando se pretende realizar técnicas de deteção de pele.

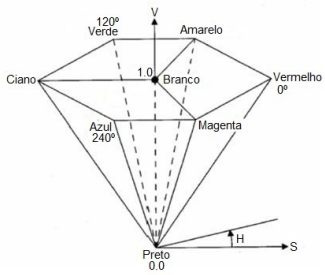


Figura 4.4: Hexagono HSV. Adaptado de [Mon].

O espaço de cor HSV pode ser ilustrado através de uma derivação do cubo RGB, mas é representado por uma pirâmide com uma base hexagonal, figura 4.4, com o ângulo H, que varia entre 0 e 360 graus, a dizer respeito ao eixo horizontal, determinando a matiz de cor desejada. A distância

perpendicular do centro até à extremidade, determina a saturação *S*. A distância vertical determina a luminância ou valor *V*, variando ambas entre 0 até 1.

A transformação do espaço de cor RGB para o espaço de cor HSV pode ser efetuado pelas equações, 4.1, 4.2 e 4.3 [MSL01]:

$$H = \arccos \frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{((R-G)^2+(R-B)(G-B))}} \tag{4.1}$$

$$S = \frac{max(R,G,B)-min(R,G,B)}{max(R,G,B)}, max(R,G,B) \neq 0 \tag{4.2}$$

$$V = max(R,G,B) \tag{4.3}$$

Existem, ainda, alguma variantes deste espaço de cor, como o espaço HSL (*Hue*, *Saturation*, *Lightness*) e HSI (*Hue*, *Saturation*, *Intensity*), que podem se representados por pirâmides de duplas hexagonais e duplas tetraédricas, respetivamente. Não há no entanto diferença significativa entre os atributos matiz e saturação comparativamente ao espaço de cor HSV.

4.1.3 Espaço YCbCr

O espaço YCbCr, uma versão em escala e deslocamento do espaço de cor YUV (Y - luminância e U e V - crominâncias), é um sinal codificado RGB não-linear, muito utilizado pelas televisões e em trabalhos de compressão de imagem. O *Y* representa a luminância, ou seja, a informação entre o preto e o branco, e *Cr* e *Cb* dizem respeito à crominância da imagem, a cor. Sendo ambos representados por 8 *bits* a luminância apresenta um intervalo de valores entre os 16 e os 235, enquanto que as componentes da crominância entre os 16 e os 240 [Spa].

A transformação do espaço de cor RGB para o espaço YCbCr é realizado pelas equações (referentes à *ITU-R(International Telecommunication Union - Radiocommunication Sector) Recommendation BT.601* [Uni11]), 4.4, 4.5 e 4.6 [MSL01]. A luminância é calculada pela luminância RGB não linear, criada através de pesos na somas das componentes RGB, com a crominância a ser obtida através da subtração da componente vermelha (R) e azul (B) com a luminância.

$$Y = 0.299R + 0.587G + 0.114B \tag{4.4}$$

$$Cb = \frac{B-Y}{2-2\times 0.1145} \tag{4.5}$$

$$Cr = \frac{R-Y}{2-2\times 0.2989} \tag{4.6}$$

A simplicidade na transformação e separação explicita dos componentes da luminância e da crominância tornam este espaço de cor bastante utilizado nas técnicas de deteção de cor de pele.

4.2 Deteção e Reconhecimento das Mãos

Inicialmente, vamos abordar o desenvolvimento relativo à aquisição de dados por parte do sistema. Sabendo que sistemas de interação baseados em visão por computador não utilizam dispositivos de rastreio físicos, mas apenas captação de imagens por parte de câmaras como a única fonte de informação, e sendo este um sistema desse tipo, é usado o sensor Kinect para se realizar a captação de imagens em tempo real e proceder à aquisição de dados para o sistema. Este sensor permitiu a conjugação de dois métodos de captura de imagens, como a captação de imagens RGB, a utilização de câmaras de infravermelhos para informação de profundidade. E um método de rastreio, neste caso da totalidade do esqueleto humano, que permite ao sistema a deteção e identificação do utilizador que pretende realizar a interação. A sincronização existente entre as três partes do sensor permite uma deteção e uma capacidade de processar a informação obtida praticamente em tempo real, obtendo-se uma latência quase impercetível entre o movimento do utilizador e a resposta do computador.

Com toda a informação a ser obtida em tempo real, é possível, com recurso ao rastreio do esqueleto humano e à informação de profundidade obtida pelos sensores infravermelhos presentes no Kinect, saber a posição e a profundidade em relação ao sensor de cada articulação do corpo. Contudo, para o desenvolvimento deste sistema a informação que tem maior relevância é a das mãos, pois com o seu rastreio e deteção tem-se o foco principal no que diz respeito ao utilizador que está a realizar a interação com o sistema.

4.2.1 Pré-Processamento

Com toda a informação a ser obtida através dos métodos de aquisição referidos anteriormente, é realizado um pré-processamento às imagens obtidas pelas câmaras. Através da ativação do *Color Stream* e do *Depth Stream* são obtidas imagens RGB e de profundidade, respetivamente, de resolução 640×480 a 30fps, e do *Skeleton Stream* por forma a ser obtido um rastreio e uma deteção corporal, mais concretamente da mão do utilizador, em tempo real.

Com o Kinect a permitir o rastreio em tempo real de até 6 esqueletos e a interação de no máximo dois deles, foi necessário recorrer a um método de seleção por forma a perceber qual das pessoas que podem ser detetadas pelo sensor é realmente o utilizador do sistema. Assim sendo, como o sistema é desenvolvido para a interação em qualquer superfície, é processada a informação de profundidade de todos as pessoas que possam estar no ângulo de visão das câmaras do sensor e selecionado aquele que estará a uma maior distância do sensor como o utilizador do sistema, uma vez que será aquele mais próximo da projeção.

Com a obtenção do utilizador do sistema passa-se ao rastreio e à obtenção de informação relativa a essa mesma pessoa, principalmente no que se refere ao posicionamento da sua mão, isto é, a respetiva localização e profundidade, de forma a ser processada separadamente do restante esqueleto, que não terá qualquer interação direta com o sistema, tornado-se, assim, dispensável o seu processamento.

Com o rastreio e a deteção da mão do utilizador a ser realizado por parte do *Skeleton Stream* do Kinect, procede-se à sua segmentação relativamente à cor de pele e à profundidade da mão, apenas nos pixeis correspondentes à sua posição. Definiu-se uma região de interesse variável, através do envio do pixel correspondente à articulação do esqueleto referente à mão, para uma classe de pré-processamento (Anexo A.1), de forma a esta região conter no seu interior a mesma. Permitindo a obtenção e separação da mão do restante esqueleto mantendo sempre as mesmas proporções, apesar de alguma perda na sua definição, uma vez que é realizado um escalonamento dessa região de forma a que a área de interesse seja diretamente proporcional à distancia entre a mão e o sensor. Com a região da mão definida é realizada a extração de todos os pixeis pertencentes à mão através da segmentações referidas, eliminando, assim, todo o ruído que possa existir e que não seja parte integrante da mão do utilizador. Contudo, a realização da segmentação em profundidade quando a mão se encontra demasiado próxima da parede resulta na junção da mão com a mesma, obtendo-se essa junção como resultado final, tornando impossível a distinção entre a mão e a superfície.

Então, para a obtenção de uma só imagem segmentada e apenas com a mão do utilizador presente na região de interesse, é necessário realizar a junção das duas imagens segmentadas obtidas anteriormente. Para isso, é aplicado um operador lógico de interseção (*AND*) que permite obter como resultado a interseção das duas imagens, ou seja, $I = C \cap P$, em que I representa a imagem resultante da junção das duas imagens segmentadas, C a imagem segmentada pela cor da pele e P a imagem segmentada através da profundidade. Conclui-se, assim, todo este processo que resulta na deteção, rastreio e segmentação da mão do utilizador do sistema.

4.2.1.1 Segmentação das Mãos pela Cor da Pele

Para a segmentação das mãos pela cor da pele, temos que assumir que os restante elementos que estarão nas proximidades da mão durante a captação da imagem têm uma cor distinta da cor da pele; neste caso para o sistema assumiu-se uma gama de valores de saturação e pigmentação para cidadãos caucasianos (cor da pele do preponente).

Com esta pequena condicionante satisfeita, realiza-se a criação da região de interesse referida anteriormente e que diz respeito à localização da mão em tempo real. Para isso, é criada uma classe de pré-processamento que permite obter apenas os pixeis correspondentes à região de interesse, eliminando os restantes pixeis da imagem captada pela câmara RGB.

Esta classe é inicializada com o envio do tamanho da região de interesse para a segmentação da mão, assim como a máxima profundidade existente no *Depth Stream* para se proceder ao redimensionamento automático da região da mão, de forma a manter sempre as mesma proporções. Posteriormente, é envida para processamento a localização do ponto referente à mão, relativamente ao seu posicionamento em relação à câmara RGB e que é obtido através do *Skeleton Stream* do Kinect. Através desta informação, é possível obter o tamanho da região da mão, assim como o seu ponto central, que é o ponto do esqueleto que é definido como a mão. Contudo, este processo apresenta a condicionante da localização da mão mais o tamanho da região não exceder a largura ou altura da imagem que está a ser obtida pela câmara, ou seja, os 640×480 pixels de resolução da imagem, sendo o rastreio perdido caso isso aconteça.

Com o processo de deteção e rastreio, assim como a definição da região de interesse a ser realizado em tempo real em cada *frame* da imagem captada, passa-se para o processo de extração de informação referente a essa mesma região. Para isso, é enviada para a classe de pré-processamento da mão o *Color Stream* e os dados referentes à imagem de cor e onde se encontra a informação de cada pixel.

Tendo toda a informação reunida e pronta para ser processada, é realizada então uma segmentação da mão do restante corpo, como observado na imagem 4.5(a), percorrendo cada *frame* da imagem na sua totalidade e coletando a informação e os dados referentes apenas à região de interesse, das 3 componentes de cor da imagem (R, G e B), por forma a posteriormente se aplicar a deteção da cor de pele apenas nessa região, eliminando com isto pontos sem interesse e reduzindo o tempo de processamento do sistema.

(a) Região da mão segmentada relativo às componentes RGB.

(b) Região da mão segmentada em termos de cor de pele aplicando o operador morfológico de dilatação.

Figura 4.5: Segmentação da mão pela cor de pele.

Com a segmentação da imagem relativamente às componentes RGB realizada, torna-se necessário proceder à deteção da mão na sua totalidade, eliminando todas as componentes presentes na região de interesse que não sejam pertencentes à mão. Para isso, converteu-se inicialmente a imagem RGB para o espaço de cor HSV, através da classe de conversão da biblioteca Emgu CV, por forma a tornar a deteção menos sensível a variações na iluminação, uma vez que está mais relacionada com a deteção de cor de pele [ATD01]. Realiza-se uma busca por um gama de matiz, saturação e brilho, entre os 0 e 60, os 58 e 173 e os 89 e 229 respetivamente, de forma a obter uma imagem binarizada, com a pele humana a ser definida como o objeto de interesse da imagem e os pixeis constituintes da mão do utilizador a serem representados a branco, enquanto os restantes ficam a preto, como observado na figura 4.5(b). É ainda aplicado, depois da binarização, um operador morfológico de dilatação, com um círculo de raio igual a dois pixeis, como elemento estruturante, permitindo, então eliminar qualquer pequena cavidade que possa existir no interior da mão binarizada e suavizar a mesma.

4.2.1.2 Segmentação das Mãos em Profundidade

Para a segmentação das mãos em profundidade, é realizada a criação da região de interesse que diz respeito à localização da mão em tempo real. Para isso é usada, uma vez mais, a classe de pré-processamento que permite obter apenas os pixeis correspondentes à região de interesse, eliminando os restantes captados pela câmara de profundidade.

Esta classe é inicializada com o envio do tamanho da região de interesse para a segmentação da mão, assim como a máxima profundidade existente no *Depth Stream* para se proceder ao redimensionamento automático da região da mão, de forma a manter sempre as mesma proporções. Posteriormente, é envida para processamento a localização do ponto referente à mão, relativamente ao seu posicionamento em relação à câmara de profundidade, que é obtido através do *Skeleton Stream* do Kinect. Através desta informação é possível calcular o tamanho da região da mão, assim como o seu ponto central, que é o ponto do esqueleto que é definido como a mão. Contudo, é importante satisfazer a condição dos limites da imagem para não ser perdido o rastreio da mão, referida anteriormente na segmentação por cor da pele.

Com o processo de deteção e rastreio, assim como com a definição da região de interesse a serem realizadas em tempo real a cada *frame* da imagem captada, passa-se para o processo de extração de informação referente a essa região. Para isso, são enviados para a classe de pré-processamento da mão o *Depth Stream* e os dados referentes à imagem onde se encontra a informação de cada pixel.

Tendo toda a informação reunida e pronta para ser processada, é realizada uma segmentação da mão do restante corpo, como observado na imagem 4.6(a), percorrendo cada *frame* da imagem em profundidade na sua totalidade e coletando a informação e os dados referentes apenas à região de interesse, por forma a posteriormente se detetar a profundidade de cada ponto da região, eliminando, com isto, pontos sem interesse e reduzindo o tempo de processamento do sistema.

(a) Região da mão segmentada pela profundidade relativa da mão.

(b) Região da mão segmentada e binarizada pela profundidade relativa da mão.

Figura 4.6: Segmentação da mão pela cor de pele.

Com a segmentação da mão em profundidade realizada, é necessário então, proceder à distinção da mão das outras componentes presentes que não façam parte da mesma. Para isto, procede-se à binarização da mão em profundidade, definindo tudo que se encontra nas proximidades da sua profundidade, 15% à frente e atrás da sua localização, como parte integrante da mão, sendo esta

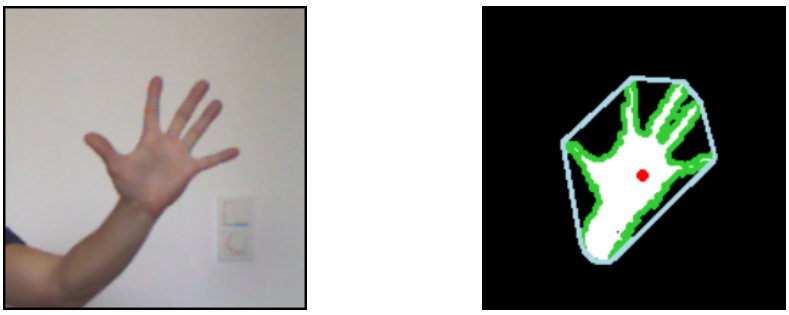
representada a branco e assumindo-se como a mão do utilizador. A restante zona será a parte dispensável da imagem e o fundo, e a sua a representação é feita a preto, como observado na imagem 4.6(b).

4.2.2 Deteção dos Dedos e Palma da Mão

Com a mão segmentada e binarizada, torna-se importante perceber de uma forma mais precisa que a do *Skeleton Stream* onde se encontra posicionada a mão e ainda detetar os dedos do utilizador em tempo real. Sendo necessário uma forma de detetar esses locais procedeu-se à sua deteção a partir da imagem segmentada e binarizada, extraindo as características importantes da mão do utilizador (Anexo A.3).

A realização desta deteção tira partido das funcionalidades da biblioteca referida anteriormente, Emgu CV, para proceder inicialmente à deteção dos contornos da mão, resultando este processo num contorno da imagem binarizada obtida através da segmentação. Com os contornos realizados, avança-se para a deteção da área convexa da mão do utilizador através da definição da região que contém todos os segmentos dos contornos da mesma, permitindo definir uma região ainda mais aproximada da localização da mão em cada instante.

Este processo permite, através da análise da área convexa da imagem, obter todos os defeitos existentes na mesma, ou seja, as cavidades onde não está presente a mão, mas que fazem parte da área convexa da imagem (figura 4.7).



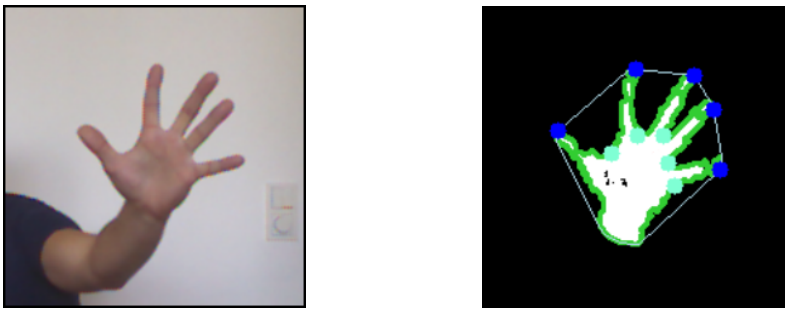
(a) Mão antes da segmentação e extração das características.

(b) Área convexa da mão (azul claro), os seus contornos (verde) e o centro da mão (vermelho).

Figura 4.7: Extração de características da mão do utilizador.

Obtém-se com esta analise o ponto de inicio de cada defeito, o seu fim e a sua profundidade, figura 4.8. Contudo, nem todos os defeitos são relevantes para o processamento, realizando-se uma busca pelos mais importantes, que são os que apresentam uma maior distância entre o seu ponto de profundidade e o seu fim, sendo estes definidos como os dedos do utilizador.

Com a obtenção dos defeitos importantes presentes na área convexa da mão, através do método descrito anteriormente, é realizado uma nova área convexa, desta vez referente aos ponto de profundidade obtidos através dos defeitos da área anterior. Assumindo-se que o centro do maior



(a) Mão antes da segmentação e extração das características.

(b) Dedos do utilizador, com ponto do fim do defeito (azul escuro) e ponto relativo à profundidade do defeito (azul claro).

Figura 4.8: Extração de características dos dedos do utilizador.

circulo que caiba no interior dessa nova área como o ponto mais preciso do centro da mão, a sua palma, como pode ser observado na figura 4.7(b).

Esta deteção e estas características tornam-se ainda mais importantes para posteriormente ser possível conseguir detetar diferentes ações do utilizador com análise dos defeitos e ainda através de um aumento do número dos defeitos importantes, diminuindo a distância de busca dos mesmo, definir a zona de toque principal com uma maior precisão.

4.3 Reconhecimento da Zona de Toque

O reconhecimento da zona de toque é onde se tomam as grandes decisões no sistema, é neste ponto que as características e a informação recolhida anteriormente durante o processo de reconhecimento e rastreio das mãos, irá ser processada e interpretada. Neste processo são analisadas todas as características obtidas referentes à mão, como a sua localização ou a sua forma, permitindo ao sistema saber se uma ação está a ser realizada, e se sim, que tipo de ação o utilizador pretende. Este processo permite ao sistema compreender e interpretar as ações do utilizador.

Para o reconhecimento ser feito torna-se necessário ainda a análise não só das mãos do utilizador, como também da superfície onde a projeção está a ser realizada. Só assim é possível ao sistema perceber onde e quando está a ser realizada uma determinada ação e posteriormente interpretar isso, permitindo perceber o que está a ser realizado e em que local.

Todo este reconhecimento torna, assim, possível traduzir o ponto de contacto do utilizador com uma superfície em informação que será interpretada por um computador e que permitirá realizar as ações básicas existentes em qualquer superfície de reconhecimento de toque.

4.3.1 Criação da Janela de Calibração

A criação da janela de calibração pode ser interpretada como o início da interação do utilizador com o sistema, uma vez que é através desta seleção que o sistema irá saber o local onde a projeção

está a ser realizada e o tamanho da mesma. Esta é a forma do sistema perceber onde estará cada ponto da imagem e permitir detetar o local de toque na superfície. É com este processo que o computador consegue perceber se a mão está , ou não, em contacto com a superfície.

A calibração é, então, efetuada logo após o programa ser posto a correr no computador e o Kinect estar posicionado em frente à superfície de projeção. Sendo inicialmente apresentada ao utilizador a imagem RGB que o sensor está a capturar, tendo a projeção que estar na sua totalidade a ser observada e o mais paralela à superfície possível. É importante que se tente que a projeção se localize no centro da imagem, para não se correr o risco de perda de rastreio das mãos num determinado local da projeção, devido à limitação do reconhecimento das mãos na proximidade das bordas da imagem capturada, referida anteriormente. É também de extrema importância que não se encontrem objetos ou pessoas em frente a superfície de projeção durante a criação da janela de calibração, por forma a que o sistema funcione corretamente.

Este processo (Anexo A.2) é então iniciado pela seleção dos quatros cantos da projeção na imagem que está a ser capturada, começando pelo ponto superior esquerdo da projeção e seguindo a orientação horária do relógio. Permiti-se, assim, ao sistema perceber onde está localizada a projeção e formando com isto um retângulo ao seu redor, figura 4.9.

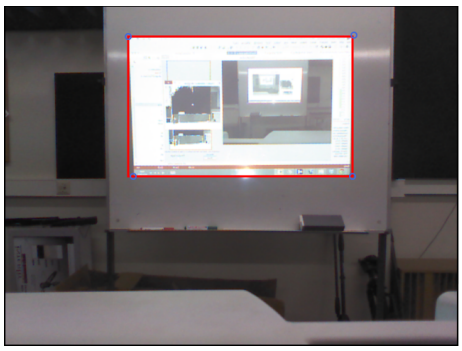


Figura 4.9: Janela de calibração após seleção dos cantos da projeção.

Contudo, não é através da criação do retângulo à volta de projeção que é realizada a criação da janela de calibração, sendo esta meramente um apoio visual ao utilizador, sendo importante para o sistema a localização dos seus cantos. Após a seleção dos quatro cantos da projeção e de guardada a sua localização é necessário ao sistema perceber a que distância se encontra cada pixel da imagem do sensor, de forma a perceber se o toque está ou não a ser efetuado. Este processo não pode, contudo, ser realizado através de um único *frame*, uma vez que a câmara de profundidade é demasiado irregular e sensível a variações de luz, o que não permite obter distâncias iguais para o mesmo pixel em *frames* consecutivos.

Para isso, é realizado num curto período de tempo uma captura de valores por parte do sensor de profundidade relativamente a cada pixel da imagem, resultando a sua média durante esse período numa distância mais precisa.

O sistema inicia, assim, uma classe de processamento das *frames* obtidas pelo sensor de profundidade. Este processo realiza-se através da análise das 200 *frames* posteriores ao início do

processo, com o sistema a percorrer cada uma das *frames* na sua totalidade, em largura e altura, e guardando a informação referente à profundidade de cada pixel da imagem numa matriz de igual dimensão, 640×480. Com a obtenção de todas as profundidades, realiza-se uma média das mesma, para permitir ao sistema perceber com maior precisão a que distância se encontra cada pixel na imagem, criando uma matriz de calibração normalizada, ou seja, um matriz das médias das profundidades de cada pixel num curto espaço de tempo, apresentando-se de seguida uma imagem em de profundidade que diz respeito à imagem de calibração obtida, figura 4.10, e que pode ser observada pelo utilizador, para garantir que a mesma não apresenta irregularidades. Durante o processo de calibração caso o sistema detete alguma sombra ou algum ponto de profundidade não seja detetado, é apresentada uma mensagem que informa o utilizador de que o processo de calibração pode apresentar erros.

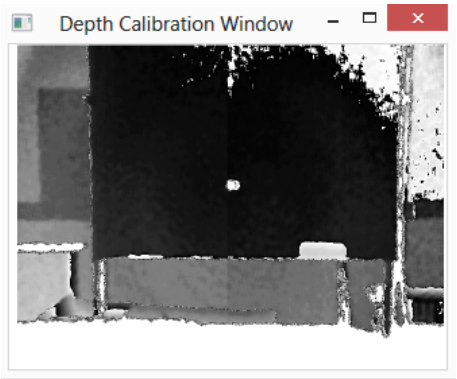
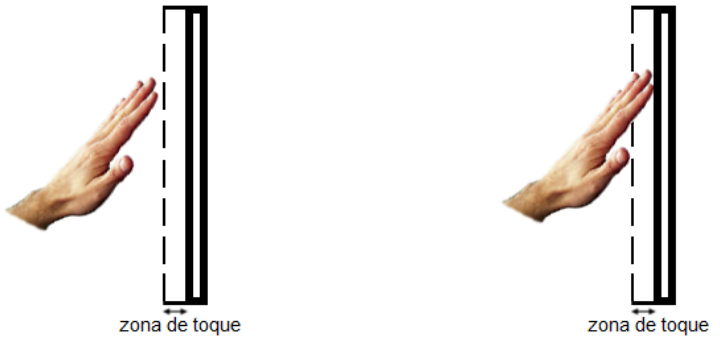


Figura 4.10: Exemplo de imagem de profundidade de uma superfície obtida depois de realizada a calibração.

4.3.2 Deteção de Toques na Superfície

Com a calibração realizada o utilizador do sistema pode começar a interagir com a superfície. Contudo, quando a superfície é tocada é necessário detetar a realização dos toque e a sua localização. Para isso é importante a análise das características da mão e dos dedos extraídas depois do processo de segmentação da mão, tal como a localização da janela de calibração e de todos os pixeis que estão presentes no seu interior. É ainda importante referir a definição de toque no sistema, ou seja, a forma do utilizador interagir com o sistema desenvolvido. Assim, para a realização de um toque na superfície o utilizador terá que ter a sua mão numa posição totalmente aberta, preferencialmente, ou semi-aberta (como por exemplo dois dedos esticados), como observado nas figuras 4.13(a) e 4.13(b), para um tipo de interação, e com a mão aberta e o polegar em extensão lateral para outro, como observado na figura 4.13(c).

Inicialmente, é preciso que o sistema compreenda onde é a parede e se a mão do utilizador se encontra longe ou perto da mesma. Para isso, recorre-se à janela de calibração obtida anteriormente. Utilizando a matriz de distâncias calculadas durante o processo de calibração passa-se à definição de uma zona junto da parede que é definida como a zona de toque na superfície, como



(a) Sem Toque na Superfície.

(b) Toque na Superfície.

Figura 4.11: Zona de toque na superfície.

observado na figura 4.11. Esta zona é obtida através do cálculo de um pequena percentagem dos valores das distâncias obtidas na matriz de calibração. Ou seja, percorre-se toda a matriz de calibração retirando ao valor das distâncias 4% da distância obtida durante o processo de calibração, obtendo-se com isto, uma pequena área de dimensões muito reduzidas que permite ao sistema saber que quando um dedo está para lá dessa distância está em contacto com a superfície.

A zona de toque fica então definida como sendo a zona do espaço existente entre a janela de calibração e a distância calculada anteriormente, que está presente entre os valores selecionados como os cantos da projeção no passo inicial da calibração. Pretende-se com este processo tornar como área de toque e deteção apenas o local exato onde está a ser realizada a projeção.

Com esta área definida passa-se para um novo passo na deteção dos toques na superfície. Este processo (Anexo A.3) resulta da análise da extração de características da mão e dos dedos e ainda da área de toque definida com o auxílio da janela de calibração. Para a deteção de toque na superfície torna-se importante perceber inicialmente qual a direção em que a mão se encontra, bem como recorrer ao esqueleto rastreado através do Kinect usando-se os pontos referentes à mão e ao pulso obtidos pelo *Skeleton Stream* que permitem saber em tempo real qual a orientação de cada uma das articulações. Este método resulta do utilizador se encontrar sempre a alguma distância do sensor Kinect. Com a obtenção das coordenadas de cada um desses pontos torna-se possível calcular através de uma simples operação matemática (equação 4.7), o ângulo resultante destas duas articulações em cada instante.

$$\Theta = \arctan\left(\frac{wrist_y - hand_y}{wrist_x - hand_x}\right) \times \frac{180}{\pi} \tag{4.7}$$

O cálculo deste ângulo permite através da definição de limites (tabela 4.1), determinar em que direção a mão se encontra, ou seja, saber em tempo real se a mão do utilizador está voltada para cima, para baixo, para a esquerda ou para a sua direita. Assim, se por exemplo, o ângulo entre a mão e o pulso for de 170^{o}, a mão estará a apontar para a direita.

Com a realização deste último passo, toda a informação está disponível e pronta a ser tratada e processada de forma a ser executado o algoritmo de deteção de toques na superfície. Assim, para a

Tabela 4.1: Limites definidos para a orientação da mão em função do ângulo entre o ponto do esqueleto do pulso e da mão.

Posição da Mão	**Limites (º)**
Cima	Se $\Theta \geq -160$ e $\Theta \leq -45$
Baixo	Se $\Theta \geq 45$ e $\Theta \leq 135$
Esquerda	Se $\Theta > -45$ e $\Theta < 45$
Direita	Se $\Theta > 135$ ou $\Theta < -160$

sua criação inicialmente é necessário processar a informação obtida através da deteção da palma da mão do utilizador em tempo real obtida anteriormente. Dando início a este processo, é transferido o ponto obtido durante a segmentação da mão para a imagem RGB que está a ser capturada e onde será realizada a deteção. Posteriormente, são usadas a localização e a profundidade desse mesmo ponto para obter de forma mais exata o local onde se encontra a mão do utilizador em cada instante, através de um método ligeiramente mais preciso que o ponto da mão obtido através do *Skeleton Stream* (este é demasiado irregular na sua localização). Através deste método de deteção é possível ter a certeza que o centro da mão do utilizador é sempre o ponto a ser utilizado.

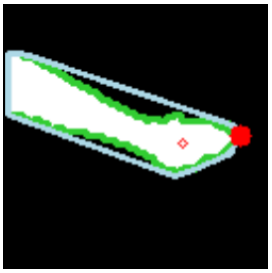
Este método, apesar de permitir obter uma boa precisão relativamente ao ponto central da mão quando são realizados movimentos, varia em demasia a localização e a sua profundidade, não sendo estável o suficiente para que a movimentação da mão possa ser processada de forma mais precisa. Para se proceder a uma suavização deste ponto, de forma a torná-lo estável o suficiente para o seu rastreio ser realizado, cria-se um *array* que guarda as últimas oito posições do centro da mão, permitindo com isto calcular uma média da sua localização num curto espaço temporal, assim como obter o valor mínimo de profundidade da mão nesse mesmo intervalo de tempo. Este método torna o rastreio do centro da mão muito mais suave, eliminando a oscilação de valores que existia anteriormente.

Contudo, apenas o centro da mão não é suficiente para o sistema perceber se o utilizador está realmente a tocar na superfície com os seus dedos, tornando o próximo passo essencial neste mesmo processo, ao utilizar os pontos obtidos anteriormente na segmentação da mão e passando para a imagem de cor captada as características que dizem respeito à localização dos dedos. É ainda neste ponto que se torna importante perceber a orientação da mão, de forma a detetar o local onde está a ser realizado o contacto com mais exatidão. Este método começa inicialmente por proceder à deteção da orientação da mão, sendo essa informação recebida e processada de forma a permitir ao sistema realizar uma busca do ponto principal de toque da mão com a superfície. Este ponto é definido como sendo o local que o sistema detetará se o contacto é ou não realizado e está definido como o ponto mais distante da mão do utilizador no sentido da sua orientação. Ou seja, se a mão estiver virada para a direita, o ponto principal de contacto com o sistema que será analisado será o ponto mais à direita da mão do utilizador.

Com o sistema a conseguir saber qual o ponto que deve detetar, sabendo a orientação da mão, são, então, percorridos todos os pontos obtidos na deteção dos dedos, realizando-se uma busca pelo ponto mais distante no sentido da orientação. Esse ponto, é definido, como referido anteriormente,

(a) Mão para cima.

(b) Mão para baixo.



(c) Mão para a direita.

Figura 4.12: Ponto principal de toque da mão na superfície (vermelho) em função da orientação.

por ponto principal de toque na superfície (figuras 4.12(a), 4.12(b) e 4.12(c)). Contudo, a oscilação referida anteriormente na deteção do ponto central da mão acontece de igual forma com os dedos. Por isso, é realizada também uma suavização desses pontos através do método atrás enunciado para suavização da palma da mão e, ainda, uma deteção da sua profundidade mínima nas últimas oito *frames*. Obtém-se assim um ponto muito mais estável e preciso.

Com a obtenção dos dois pontos acima identificados para a deteção do toque na superfície, é neste momento que o sistema percebe se esta a ser, ou não, realizada alguma ação, sendo neste ponto que se conjugam todos os passos anteriormente descritos e desenvolvidos. Resumindo, é pois detetado o posicionamento do centro da mão e o ponto de toque principal do utilizador em tempo real e a sua profundidade a cada instante, e realizada uma comparação desses valores com os valores calculados anteriormente, que dizem respeito à zona de toque na superfície, tal como a localização da janela de calibração e a profundidade de cada pixel da mesma. Caso seja detetada a presença da mão e dos dedos nessa mesma zona é realizado o processamento de deteção do tipo de ação que está a ser executada pelo utilizador e, sendo esta uma ação de interação, é passada ao computador toda a informação referente ao tipo de ação pretendida e a sua localização na superfície.

4.3.3 Controlo do Computador

Com todas as características das mãos e dedos extraídas e processadas e a deteção de toques na superfície realizada, é necessário o sistema interpretar as ações realizada pelo utilizador e enviar essa mesma informação ao computador de forma a obter um resultado e uma resposta a essas ações. Para isso, é então necessário traduzir e interpretar cada gesto e cada toque em *inputs* de dados específicos que correspondam à ação que está a ser realizada pela utilizador, na posição em que a mão toca a superfície.

4.3.3.1 Conversão das Coordenadas de Toque

Quando o toque é realizado todo o seu processamento e o processo de localização do mesmo é obtido em função da imagem capturada pela câmara de cor, sendo os dados obtidos referentes à localização desse ponto num determinado pixel da imagem visualizada. Assim, a resolução

obtida através da câmara RGB está definida para 640×480 mas a resolução do monitor/projeção do computador a ser utilizado sofre variações, dependendo do monitor que está a ser utilizado, sendo a sua relação variável de acordo com a resolução definida pelo utilizador para esse monitor.

Este diferença de proporções leva a que seja necessária uma conversão do ponto de toque detetado para o seu valor real, que diz respeito à resolução que está a ser utilizada para a projeção que está a ser realizada. Ou seja, é necessário fazer um mapeamento das coordenadas de toque para coordenadas no monitor real que está a ser utilizado, que no caso será a do projetor. O ponto de toque, neste processo não será no entanto a coordenada referente à resolução da imagem, mas sim o ponto de toque no interior da janela de calibração, sendo o canto superior esquerdo da projeção o ponto (0,0) e o inferior direito o definido durante o processo de calibração.

Por fim, é então mapeado o ponto de toque na superfície para o seu valor real através das expressões, 4.8 e 4.9, para as coordenadas *(x,y)* do monitor real que está a ser utilizado.

$$X_{real} = \frac{Monitor_{width}}{ProjectionX_{tl} - ProjectionX_{dr}} X_{toque} \tag{4.8}$$

$$Y_{real} = \frac{Monitor_{height}}{ProjectionY_{tl} - ProjectionY_{dr}} Y_{toque} \tag{4.9}$$

Sendo X_{real} e Y_{real} as coordenadas *(x, y)* do monitor real, X_{toque} e Y_{toque} as coordenadas *(x, y)* obtidas com o toque na superfície projetada, $Monitor_{width}$ e $Monitor_{heigth}$ a resolução atual do monitor/projetor em largura e altura, $ProjectionX_{tl}$ e $ProjectionY_{tl}$, $ProjectionX_{dr}$ e $ProjectionY_{dr}$ as coordenadas *(x, y)* selecionadas durante a calibração e que se referem ao canto superior esquerdo e inferior direito, respetivamente.

4.3.3.2 Movimentação e Ações do Cursor

Com a deteção do toque realizada e as coordenadas de toque obtidas, é agora necessário converter as ações do utilizador em movimentos ou ações, traduzindo-as a ações do cursor, ou ações que provocam mudanças no ambiente apresentado. A realização destas ações permite a ligação entre o utilizador e o sistema, permitindo a este último interpretar a intenção do utilizador com todo o processamento atrás descrito e traduzir essas ações em *inputs* de dados permitem o controlo através do toque da projeção que está a ser realizada.

A ligação do sistema ao computador (Anexo A.4) e para a introdução de *inputs* que permitam a interação direta com a projeção foi realizada através da injeção de pontos de toque por filas de mensagens do sistema operativo utilizado, Windows 8, permitindo a simulação de toques na superfície através de uma ligação à API de *input* de toques do Windows. É ainda feita uma ligação ao cursor do computador, permitindo o seu controlo de movimentos e a realização das suas operações, clique esquerdo e direito, assim como a realização de *scrolls*.

Todas estas operações são realizadas, como mencionado anteriormente através do envio de mensagens por fila. Para isto é necessário realizar uma importação para o sistema do *user32.dll*, uma biblioteca de vínculo dinâmico do Windows, sendo a mesma usada como parte da interface do sistema operativo permitindo a manipulação do ambiente do computador.

Então, é necessário o envio de mensagens especificas que permitam ao sistema operativo perceber que operação esta a ser realizada e em que local, sendo para isso usado um conjunto de *flags* (tabelas 4.2 e 4.3) que contém a ligação às diferentes ações pretendidas, juntamente com a localização do toque obtida anteriormente.

Tabela 4.2: Mensagens e *Flags* principais referentes à injeção de toques no sistema operativo [Mic12c].

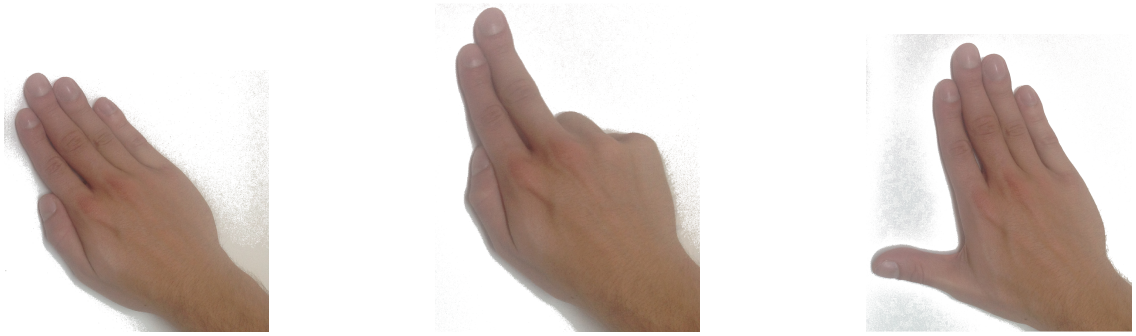
Pointer Flag	**Código**	**Ação**
NONE	0*x*00000000	*Default*
DOWN	0*x*00010000	Significa que o ponto de contacto está num estado para baixo, isto é, está em contacto com a superfície
UPDATE	0*x*00020000	Significa que não existe mudança de estado, isto é, de ação
UP	0*x*00040000	Significa que o ponto de contacto está num estado para cima, isto é, não há contacto com a superfície
WHEEL	0*x*00080000	Significa que está a existir um *scroll* vertival

Tabela 4.3: Mensagens e *Flags* principais referentes às ações e movimentos do cursor

Mouse Event	**Código (Hexadecimal)**	**Ação**
Move	0*x*01	Permite a movimentação do cursor para a posição em que é realizado o contacto com a superfície
Left Down	0*x*02	Realiza um clique esquerdo do rato realizando um ação, isto é, o botão esquerdo do rato é premido
Left Up	0*x*04	Significa o fim da ação do clique esquerdo, isto é, o botão esquerdo do rato deixa de ser premido
Right Down	0*x*08	Realiza um clique direito do rato realizando um ação, isto é, o botão direito do rato é premido
Right Up	0*x*10	Significa o fim da ação do clique direito, isto é, o botão direito do rato deixa de ser premido

Com toda a informação definida e todo o processamento realizado, é importante perceber qual a ação que o utilizador pretende e por fim associá-la à *flag* correspondente, para enviar a mensagem correta ao sistema operativo e ser realizada a ação certa. Torna-se importante perceber que ação da mão está associada a cada ação do cursor.

Assim, quando a mão do utilizador se encontra totalmente estendida e paralela à superfície de toque, com todos os dedos juntos ou dois em extensão (figura 4.13(a) e 4.13(b)), é realizada uma ação de toque na superfície, ou seja, um clique esquerdo do rato na posição onde se encontra o ponto principal de toque. Quando a mão do utilizador se encontra com o polegar estendido, ou seja perpendicular à mão (figura 4.13(c)), é realizada a ação associada ao clique direito do rato, mas apenas caso a mão esteja na mesma posição durante um segundo. Este mesmo gesto permite, também, realizar as ações de *scroll*, caso a mão seja movimentada para cima ou para baixo, durante a sua execução.



(a) Mão paralela à superfície e dedos todos juntos.

(b) Mão paralela à superfície e dois dedos em extensão.

(c) Mão paralela à superfície e polegar aberto.

Figura 4.13: Exemplo dos gestos que realizam as ações do cursor.

Toda a movimentação do cursor é realizada independentemente de como a mão seja apresentada, contudo as ações não são realizadas. É importante referir, que todas estas ações apenas têm lugar quando o ponto principal de toque, está na chamada zona de toque do sistema.

Neste ponto dá-se então o final de todos os processos anteriormente referidos, com o culminar de todo o processamento realizado a ser transferido para o computador e o seu sistema operativo, resultando as ações do utilizador em ações computacionais e permitindo o controlo do ambiente apresentado e projetado na superfície.

4.4 Sumário

Neste capítulo foi abordado todo o desenvolvimento realizado, todos os processos e decisões que levaram ao resultado final do sistema. A implementação do sistema foi assim alvo de uma análise mais profunda, sendo descrito o processo de deteção e rastreio da mão e o do controlo do computador.

Foram analisados os espaços de cor mais utilizados em técnicas de deteção de cor da pele: os espaços RGB, HSV e YCbCr. Foram explicadas e analisadas as vantagens e desvantagens do uso de cada um destes espaços, para o fim pretendido, e a forma de os a partir do espaço de cor o RGB.

Na deteção e rastreio da mão abordaram-se os dois métodos de segmentação desenvolvidos para permitir obter apenas a mão do utilizador e realizar uma análise apenas dessa parte do corpo, reduzindo o tempo de processamento e melhorando o tempo de resposta do sistema. Foi feita uma segmentação através da cor da pele e da profundidade a que se encontra a mão em relação ao Kinect. Este processo necessitou da extração de características das mãos, mais concretamente da obtenção da localização dos dedos e do centro da mão.

Para o controlo do computador abordou-se a criação de uma janela de calibração, que inclui todas as profundidades normalizadas da imagem captada pelo sensor livre de pessoas ou artefactos na área de projeção, define-se uma zona de toque limitada, através da escolha de limites de profundidade. Foi ainda descrito todo o processo de deteção de toques da mão na superfície de

projeção, que permite ao utilizador interagir diretamente com a imagem que está a ser projetada e o processo de definição do ponto de toque na superfície.

Por fim, foi feita a descrição de todo o processo de mapeamento do ponto de toque na superfície para coordenadas da projeção e a deteção de ações por parte do utilizador, como o posicionamento das mãos relativamente à projeção, para as operações básicas de interação e a definição dos gestos específicos que permitem essa mesma interação.

Foi assim descrito todo o processo que leva à deteção e rastreio da mão do utilizador e o processo de levantamento de características presentes na mesma, para posteriormente os gestos e movimentos realizados serem interpretados e permitirem uma interação com a imagem projetada do computador.

Capítulo 5

Análise de Resultados

Neste capítulo pretende-se analisar todas as decisões tomadas durante o desenvolvimento do sistema e os resultados obtidos. Inicialmente aborda-se a deteção e rastreio das mãos, analisando a fiabilidade e capacidade das mãos do utilizador serem detetadas com precisão, e ainda a capacidade de se extrair informação das mesmas de uma forma que posteriormente seja possível interagir com o sistema. Faz-se ainda uma análise à zona de reconhecimento de toque, com incidência na janela de calibração desenvolvida e no ponto de toque principal com essa mesma zona. No final apresenta-se um sumário deste capítulo.

5.1 Deteção e Reconhecimento das Mãos

Inicialmente tentou-se fazer o rastreio do esqueleto em diferentes posições (de lado, de costas e de frente), mas rapidamente se chegou à conclusão que o método a utilizar para o sistema teria que ser de frente para o sensor, uma vez que nas outras posições, a oclusão provocada pelo esqueleto, que foi desenvolvido para se interagir de frente, tornava o rastreio do mesmo demasiadamente irregular e grande parte das vezes errado, não estando o sensor preparado para essas oclusões e para se interagir com o mesmo nessas posições. Não é, ainda, possível ao sensor detetar se o utilizador está de frente ou virado de costas, de forma a se inverter a direção em que se pretende realizar as interações, caso se esteja de costas.

O processo de reconhecimento das mãos do utilizador passou por várias fases até ser possível o levantamento de informação e permitir serem retiradas características das mesmas, como o seu posicionamento, o centro da sua palma, a localização dos seus dedos e toda a informação da profundidade de cada uma destas partes.

Durante todo o processo de desenvolvimento todos os módulos de processamento e classes desenvolvidas foram testadas, a fim de comprovar a viabilidade destas e o processo de deteção foi analisado para assegurar a finalidade pretendida para o sistema desenvolvido.

Para este processo realizou-se uma segmentação da mão do utilizador. Para isto recorreu-se ao *Skeleton Stream* do Kinect, que permite obter a articulação referente à mão do utilizador, e dessa forma proceder à segmentação de uma área de interesse que contenha apenas a mão. Este processo

resultou numa deteção e rastreio da mão do utilizador com bastante rigor e precisão, que permite a definição da área de interesse, obtendo-se a localização da mão em tempo real e com relativa precisão.

Contudo, este sistema de rastreio têm a condicionante da precisão do *Skeleton Stream* sofrer perturbações quando determinada parte do corpo do utilizador (a sua articulação), se encontra demasiado próxima de uma superfície, como uma parede, ou mesmo quando a toca. Esta condicionante resulta numa perda do rastreio do braço e da mão do utilizador, com as localizações retornadas pelo *Skeleton Stream* a serem erradas e a segmentação a ser realizada num local errado, levando a que a mão não esteja presente na área de interesse definida, o que pode levar a que o sistema não funcione corretamente. Esta limitação é intrínseca ao sensor Kinect, não sendo possível a sua total correção. Tentou-se ao máximo, durante o desenvolvimento, que ao acontecerem estas perturbações, não sejam interpretadas como ações pelo sistema.

5.1.1 Segmentação das Mãos pela Cor de Pele e Profundidade

Para a segmentação das mãos com base na cor de pele, o mais importante a ser analisado é a capacidade de através da imagem segmentada em termos de cor, se proceder a uma deteção dos pixels que apresentam informação de cor correspondente à gama que foi selecionada, eliminando toda a informação que não seja relevante e não seja parte integrante do utilizador do sistema, para proceder a uma correta binarização da mão do individuo.

Foi implementado um método de deteção baseado na cor da pele do proponente, caucasiano, sendo selecionada uma gama de valores relativos ao espaço de cor HSV e YCbCr, [0:60, 58:173, 89:229] e [16:200, 131:185, 80:135] respetivamente, de forma a proceder a uma deteção apenas dos pixels presentes na imagem, que apresentem valores no interior desta gama. Estes processos de deteção resultaram na imagem 5.1.

(a) Mão segmentada antes da deteção da cor da pele através do espaço de cor HSV.

(b) Mão segmentada depois da deteção da cor da pele através do espaço de cor HSV.

(c) Mão segmentada antes da deteção da cor da pele através do espaço de cor YCbCr.

(d) Mão segmentada depois da deteção da cor da pele através do espaço de cor YCbCR.

Figura 5.1: Deteção da cor de pele.

Como pode ser observado facilmente, o processo de deteção através do espaço de cor YCbCr resulta numa imagem pouco nítida e explícita da mão do utilizador, apresentando demasiados artefactos que não fazem parte da mesma, principalmente resultantes da superfície que se encontra

em fundo. Pelo contrário, o espaço de cor HSV apresenta resultados bastante bons, resultando numa binarização bastante aproximada da mão em termos de cor de pele, eliminando por completo toda a superfície que se encontra como fundo da imagem capturada. O espaço de cor RGB não foi alvo de avaliação para o desenvolvimento deste método de deteção, uma vez que não é possível a separação da luminância e da crominância nos pixels da imagem, não sendo ainda, como referido anteriormente no capítulo 4.1, um método utilizado neste tipo de deteção, na grande maioria dos casos, pois levaria à apresentação de piores resultados de deteção.

Estes dois métodos de deteção, apresentam, no entanto, uma grande sensibilidade à luminosidade existente no espaço onde está a ser utilizado o sistema, resultando em zonas erradamente detetadas ou não detetadas, caso a luminosidade seja reduzida. As sombras existentes, ou a diminuição da cor da pigmentação do utilizador causam demasiadas alterações e restringem a deteção correta de apenas a mão.

Este sistema de deteção tem ainda a condicionante de, na região em que se encontra a mão, não poderem existir objetos ou quaisquer artefactos que sejam de cor aproximadamente igual à da cor da pele. Essa existência torna o objeto integrante do individuo e sendo representada com o valor binário 1, posteriormente ao processo de deteção, resulta numa deteção errada.

Analisando estes dois espaços de cor para a deteção de cor da pele, concluiu-se que o espaço de cor HSV apresentou melhores resultados nos mais variados ambientes de iluminação e nas superfície utilizadas durante o processo de testes do sistema, sendo este espaço adotado para este método de deteção na implementação do sistema desenvolvido.

Para a segmentação relativamente à profundidade, é importante um correta deteção da profundidade da articulação resultante do *Skeleton Stream*, para detetar o posicionamento correto da mão. Este processo resulta numa correta binarização da mão do utilizador, obtendo-se uma imagem livre de artefactos irrelevantes e apenas com a mão do utilizador como produto final.

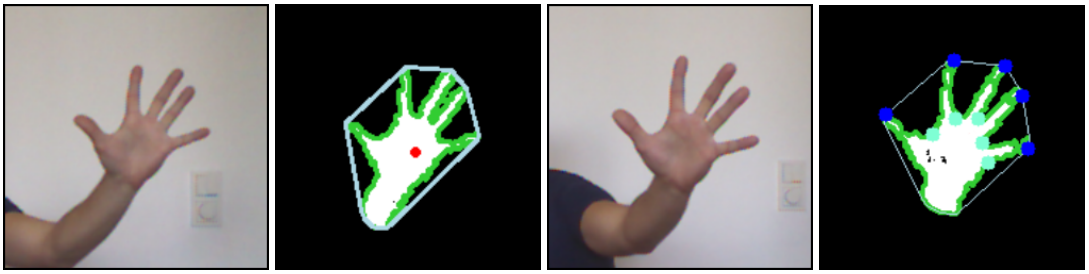
Contudo, este processo de segmentação com a mão demasiado próxima, ou mesmo em contacto com uma superfície faz com que a informação resultante da profundidade, transforme a mão em parte integrante da superfície, resultando a binarização numa fusão da mão com a superfície, obtendo-se uma imagem totalmente branca. Isto é, todos os pixeis da região são interpretados erradamente como sendo a mão do utilizador.

Para solucionar esta limitação e se obter uma melhor segmentação da mão do utilizador do sistema, adotou-se a utilização de um operador lógico de interseção, que permite uma análise dos dois processos de segmentação e obtendo-se uma só segmentação da mão, que é a interseção dos dois processos. Estes dois processos tornam-se, assim, complementares e permitem uma melhor e mais correta segmentação da mão do utilizador.

5.1.2 Extração de Características

Para o processo de extração de características da mão, utilizou-se como referido o *wrapper* Emgu CV. Estas características dizem respeito aos dedos do utilizador e ao centro da sua mão, sua palma, uma vez que era necessário perceber a localização de cada uma destas partes.

Assim, iniciou-se o processo de extração, obtendo-se o contorno da mão, a sua área convexa e por consequência os defeitos da convexidade. Através da definição de métricas, juntamente com a análise dos defeitos contidos na área convexa da mão, procedeu-se à deteção e extração da localização dos dedos e da palma da mão do utilizador. Este método resultou numa boa aproximação dessa localização, como pode ser observado na figura 5.2.



(a) Mão antes da segmentação e extração das características.

(b) Área convexa da mão (azul claro), os seus contornos (verde) e o centro da mão (vermelho).

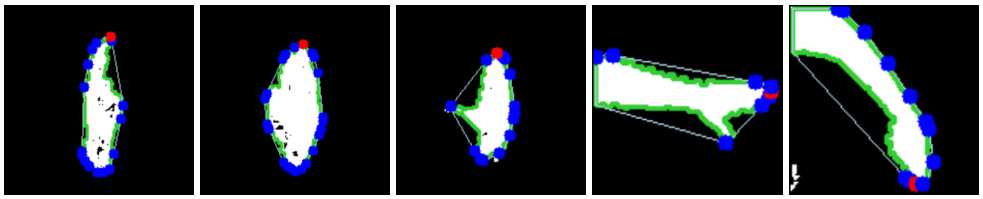
(c) Mão antes da segmentação e deteção do posicionamento dos dedos.

(d) Dedos do utilizador, com ponto do fim do defeito (azul escuro) e ponto relativo à profundidade do defeito (azul claro).

Figura 5.2: Extração de características da mão do utilizador.

No entanto, é importante a obtenção de um maior número de pontos de deteção, uma vez que os gestos utilizados necessitam que a mão se encontre em extensão, com os dedos juntos, o que torna este método pouco capaz, resultando, por vezes em pontos de deteção que podem não dizer respeito à ponta dos dedos, tornando a interação com o sistema ineficiente e errada.

Foi necessário, então, a alterar a métrica de pesquisa e análise dos defeitos da área convexa. Foi limitada a busca dos segmentos que apresentam maior comprimento, para deteção dos dedos, para ser possível detetar um maior número de pontos. Este método permite, também, definir com maior precisão o ponto mais distante com recurso à orientação da mão.



(a) Mão distante da superfície e posição de interação principal, dois dedos em extensão.

(b) Mão distante da superfície e posição de interação principal, todos dedos em extensão.

(c) Mão distante da superfície e posição de interação secundária, polegar aberto.

(d) Mão em contacto com a superfície e posição de interação principal.

(e) Mão em contacto com a superfície e posição de interação secundária.

Figura 5.3: Extração de características da mão e deteção de múltiplos pontos dos dedos do utilizador, com ponto principal de toque a vermelho.

A imagem 5.3 mostra o resultado da diminuição do comprimento na busca dos segmentos para pontos importantes a serem detetados, resultando o ponto mais distante no sentido da orientação

da mão, como o ponto de toque principal no sistema. A orientação da mão é calculada através do ângulo formado entre as articulações da mão e do pulso, resultantes do *Skeleton Stream*, com definição de limites para os quais a mão se encontra numa determinada orientação (tabela 4.1). Este sistema permite uma correta deteção da orientação da mão em relação ao corpo.

Foi, ainda, necessário obter as características referentes à mão do utilizador que digam respeito à posição dos dedos e mais especificamente do polegar. Para isso, analisou-se a profundidade dos defeitos da área convexa e a sua distância relativamente ao ponto de fim do defeito, e aplicou-se métricas que permitem o cálculo da distância entre estes dois pontos, conseguiu-se detetar os maiores defeitos existentes, independentemente da posição em que se encontra a mão. Esta análise permite saber se a mão está ou não aberta ou se o polegar está ou não em extensão de uma forma relativamente precisa.

5.2 Reconhecimento da Zona de Toque

O processo de reconhecimento da zona de toque, tal como a deteção realizada da mão do utilizador, passou por um processo de levantamento de informação de características da superfície, como a profundidade de cada pixel presente na imagem capturada e a localização da projeção.

Todo o processo de desenvolvimento bem como todas as classes foram testadas a fim de se comprovar a viabilidade do módulo de deteção desenvolvido, de forma a permitir um correta interação do utilizador com a superfície, sendo todos os toques realizados corretamente em termos de localização e de tipo de interação pretendida.

Inicialmente, para este processo, procedeu-se à criação de uma janela de calibração que permite ao sistema detetar onde está a ser realizada a projeção e obter, desta forma, a sua localização. Retirou-se, também, a profundidade de cada pixel existente na imagem captada pelo *Depth Stream*, o que permite criar uma imagem em profundidade de toda a superfície de toque e de todo o campo de visualização do Kinect.

Para este processo foi realizada a captura do *frame* correspondente ao fim da seleção da projeção na imagem RGB por parte do utilizador, e retirados todos os valores de profundidade correspondentes. Contudo, o sensor Kinect tem variações de precisão bastante grandes, principalmente com o aumento da distância entre o sensor e o objeto que está a captar [KE12], levando a que com apenas uma *frame* não seja possível obter a distância correta a que se encontra a superfície. Assim, procedeu-se a uma normalização dessa distância através do cálculo das 200 *frames* capturadas posteriormente ao fim da seleção da projeção, permitindo obter um resultado mais correto e preciso dessa mesma distância. Este reconhecimento permitiu, também, a definição de uma zona de toque onde posteriormente se procedeu à deteção de toques na superfície.

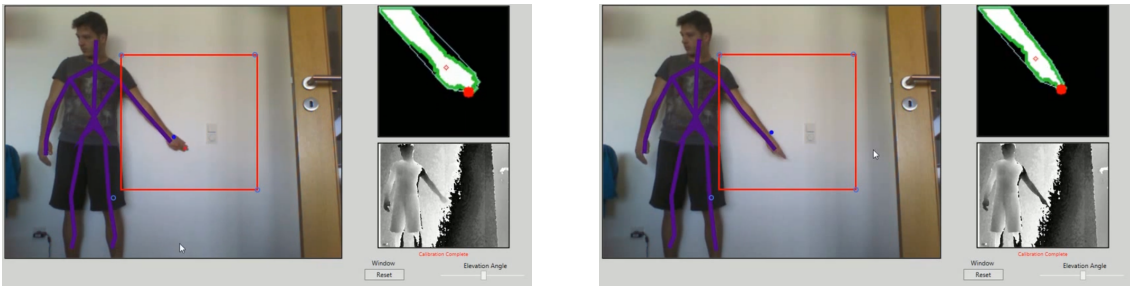
Para a realização da deteção, foram usadas, como referido, as características obtidas através do processo de rastreio e deteção das mãos, definindo-se um ponto de toque principal na superfície.

Inicialmente, a possibilidade de obter a localização da ponta de cada dedo foi a ideia principal no desenvolvimento do sistema. Contudo, a pouca precisão do Kinect e a variação existente no local onde se encontram as extremidades dos dedos, apesar de serem de poucos pixels, torna

este meio ineficiente. Essa ineficiência deve-se sobretudo à variação existente, que apesar de reduzida, torna muito difícil conseguir que o ponto principal de toque se centre em todos os dedos do utilizador, independentemente da orientação da sua mão. Ou seja, quando a imagem está a ser captada e segmentada, um dedo traduz-se num número de pixels tão reduzido, principalmente a grandes distâncias, que obter uma precisão tão elevada tornou-se num objetivo não concretizável.A enorme variação existente levava a que o ponto principal de toque fosse definido não como o centro do dedo mas como o ponto que ficava um pequeno número de pixels ao lado, o que significava o fundo da imagem, isto é, a superfície de toque. Assim, resultava este método num funcionamento incorreto e pouco preciso do sistema.

Passou-se, então, para um processo diferente de deteção do ponto principal de toque, ao ser usada toda a mão, em extensão. Para isso foi estudado o normal toque na superfície, e concluíu-se que o ponto mais distante referente à mão, na sua orientação, está normalmente a ponta dos dedos e seria um bom método de definir esse mesmo ponto, uma vez que a área para deteção desse ponto seria bastante maior. Desta forma tornou-se a interação bastante fluída e o mais natural possível. Foi ainda definido um pequeno gesto de abertura do polegar, para uma interação secundária, de forma a não se perder a naturalidade na interação.

Aplicando os métodos descritos anteriormente para a obtenção do maior número de pontos referentes à mão do utilizador, ou seja, analisando os defeitos importantes existentes na área convexa da mão, obteve-se com sucesso esse ponto, bem como a sua localização com relativa precisão, e ainda a abertura ou não do polegar para a realização de outro tipo de ação, como observado na figura 5.4.



(a) Toque na superfície, com o ponto vermelho a representar o local do toque.

(b) Sem toque na superfície.

Figura 5.4: Simulação de um toque na superfície.

A precisão obtida com a realização destes métodos de extração resultou de um processo de suavização dos valores obtidos, uma vez que os valores retornados apresentavam uma variação bastante elevada e era necessário proceder à sua suavização. A criação de um *array* que permitisse o cálculo da média dos últimos valores conhecidos dos pontos de informação extraídos, como a localização e a profundidade da mão e dedos, tornou-se um excelente método de suavização. Foram definidos os últimos oito valores obtidos para a realização da suavização, uma vez que com o aumento do número de valores para o cálculo da média a suavização introduzida tornava o sistema pouco fluído e natural.

O reconhecimento da zona de toque na superfície, apresenta, no entanto, também, a limitação já referida de o esqueleto obtido através do *Skeleton Stream* sofrer demasiadas perturbações quando se encontra demasiado perto de uma superfície. Com o rastreio do braço e da mão a ser mal realizado, sendo este deslocado muitas vezes como estando integrado na própria superfície e no interior dos limites da janela de calibração, leva a que existam falsas deteções de toque resultantes desse mau rastreio e deteção do Kinect. Uma vez mais, este é um problema intrínseco do Kinect, não podendo serem eliminadas na totalidade estas falsas deteções. Tentou-se, mesmo assim, que as interferências existentes não fossem interpretadas como uma interação direta com o sistema, através da limitação espacial da localização da mão em *frames* consecutivas.

5.3 Controlo do Computador

Para a realização do controlo do computador, foi necessário inicialmente fazer a conversão e mapeamento das coordenadas de toque para as coordenadas do monitor ou projetor que estava a ser utilizado. Para isso, usou-se uma expressão matemática que permite realizar esse mapeamento da forma mais correta possível, obtendo-se uma precisão bastante elevada em relação ao local de toque na projeção e o local onde se está na realidade a interagir com o computador. As movimentações e ações do cursor foram pensadas inicialmente apenas como interações de toque, realizando-se uma ligação entre o sistema e a API de toque do sistema operativo utilizado. Contudo, não existiam essas mesmas ligações na linguagem de programação utilizada, tendo que ser criada uma classe que permitisse a ligação à API e que simulasse a realização desses mesmos toques. A simulação e ligação à API de toque permitiram, ao serem realizadas as ações definidas de interação, obter um resposta do computador, quando o toque é realizado, com o aparecimento de um circunferência em redor desse mesmo local e ainda obtendo-se uma ligação direta, através de um botão na barra de aplicações, ao teclado da interface de toque do sistema operativo.

Este método, tinha a condicionante de que quando utilizada, por exemplo, uma aplicação de desenho não realizava uma ligação continua enquanto era realizada uma ação de deslocamento sem retirar a mão da superfície, sendo apenas representados pontos por toda a zona de deslocamento. Procedeu-se assim à implementação de uma ligação ao controlo do cursor do sistema operativo em simultâneo: realizando-se a ligação ao seu movimento e às ações efetuadas, conseguiu-se suprimir esse problema. Adicionou-se, ainda, um elemento que permite uma resposta visual ao utilizador de onde está a realizar a interação e uma ligação ao movimento de *scroll*.

Conseguiu-se, através destas ligações, proceder ao controlo do computador através das principais ações de interação do rato e de toque, como são o caso do clique esquerdo (ou *Tap*), o clique direito (ou *Holding*) e *scroll* (ou *slide*).

5.4 Sumário

Neste capítulo foram analisadas as principais decisões tomadas e os principais resultados obtidos durante o desenvolvimento do sistema, analisando de uma forma mais profunda a deteção e

rastreio das mãos do utilizador, o reconhecimento da zona de toque e o controlo do computador.

Durante o processo de deteção e rastreio das mãos, percebeu-se que as principais falhas existentes durante o reconhecimento das mesmas por parte do sistema, se devem ao sensor Kinect perder momentaneamente, o rastreio de determinadas partes do esqueleto, retornando um posicionamento incorreto da mão do utilizador e consequentemente uma errada segmentação da mesma. Foram analisados os diferentes espaços de cor para deteção de pele e justificado o uso do HSV, uma vez que apresenta melhores resultados em diversos ambientes luminosos, além de permitir a separação das componentes da luminância e da crominância. Analisou-se, ainda, o processo de extração de características, concluindo-se que apenas os pontos referentes à ponta dos dedos, não seriam suficientes para a posterior definição da zona de toque ou interação principal da mão com o sistema, sendo necessário um maior número de pontos. Foram ainda definidas as limitações do sistema em termos luminosos, e de cores presentes nas proximidades da mão.

Para o reconhecimento da zona de toque na superfície foi analisada a precisão do sensor Kinect, levando a que a janela de calibração fosse efetuada com uma média de valores de profundidade capturados num curto espaço de tempo. Foi analisado o processo de definição do posicionamento da mão para realizar toques na superfície, definindo-se a mão em extensão como um gesto natural e fluído para a realização dessa interação. Justificou-se ainda a necessidade de suavização da localização dos pontos dos dedos e da mão, devido à variação retornada do processo de extração de informação. Referiu-se, ainda, a possibilidade de deteções erradas de toques na superfície, devido a dificuldade de rastreio do esqueleto por parte do Kinect junto a superfícies.

Quanto ao controlo do computador, analisou-se o porquê de ser necessário a simulação de toques na superfície e a necessidade de se realizar uma ligação à API de toque do sistema operativo e ao cursor, de forma a permitir uma correta interação com o sistema e oferecendo uma resposta visual ao utilizador das ações realizadas.

Capítulo 6

Conclusões e Trabalho Futuro

Neste capítulo pretende-se destacar os principais resultados e conclusões obtidas durante o desenvolvimento do sistema implementado e ainda propor trabalhos futuros e melhoramentos. Inicialmente, aborda-se a realização dos objetivos e posteriormente são descritos alguns processos que podem ser acrescentados ao sistema e um exemplo de como pode ser possível a sua realização.

6.1 Resultados

Esta dissertação tinha como objetivo principal a criação de uma interface tátil com recurso ao sensor Kinect, transformando superfícies como paredes em enormes ecrãs sensíveis ao toque, através de uma projeção. O objetivo proposto foi alcançado, permitindo o sistema desenvolvido controlar o ambiente de um computador através de toque direto na superfície onde está a ser realizada a projeção, isto apesar de algumas limitações, principalmente intrínsecas ao sensor.

Como demonstrado nos capítulos anteriores, foi possível desenvolver um sistema de deteção e rastreio das mãos e extração de características das mesmas, algo que o Kinect não realiza, juntamente com a criação de um sistema de deteção de toque em superfície, permitindo esta combinação a criação e desenvolvimento do sistema pretendido. É possível, com o sistema, detetar diferentes tipos de gestos, que correspondem a diferentes tipos de interação, e traduzir esses mesmos gestos em ações no ambiente de trabalho projetado, tornando essa mesma projeção num ecrã interativo de fácil utilização.

Era essencial, para o desenvolvimento deste sistema, a deteção e rastreio das mãos do utilizador, de forma a serem retiradas características das mesmas, para permitirem ao sistema saber a localização dos dedos do utilizador e a sua profundidade. Todo este processo iniciou-se com um pré-processamento e uma segmentação das mãos em termos de cor de pele e profundidade, e posteriormente com a extração de informação com o *wrapper* Emgu CV, e resultou num processo bastante eficiente, obtendo-se resultados bastante corretos em termos de precisão.

Relativamente ao processamento das características da superfície de projeção, com a realização de uma janela de calibração todo o método de desenvolvimento de uma zona de deteção de toque e controlo do computador, em conjunto com a análise das características da mão, resultaram

numa deteção de toques com bastante precisão, com o local tocado a ser devidamente mapeado para o local correto do computador e consequentemente da projeção. A deteção de toque, ou não, na superfície foi também realizada com um bom grau de certeza, desde que se respeitem as regras de funcionamento descritas no processo de desenvolvimento, relativamente ao posicionamento da mão e do corpo do utilizador. As ações definidas para as interações, são também detetadas com um bom grau de certeza.

A combinação destes processos permitiu atingir o objetivo proposto no início desta dissertação. Contudo, limitações como a luminosidade do espaço onde está a ser realizada a projeção podem levar a uma não correta segmentação da mão pela cor da pele, uma vez que podem acontecer alterações na pigmentação da pele. A realização de projeções em superfícies brilhantes que possam refletir a luz da projeção pode interferir também na correta utilização do sistema. Há, ainda, os problemas intrínsecos ao sensor, como a profundidade variável e pouco precisa de um mesmo ponto da imagem ou o ineficiente rastreio das articulações junto a superfícies, podem interferir de forma a tornar o sistema suscetível a erros e a interações erradas.

6.2 Trabalho Futuro

Pretende-se agora apresentar propostas de a possíveis melhoramentos e de trabalho futuro que pode ser implementado no sistema desenvolvido. Relativamente aos aperfeiçoamentos a realizar no sistema, é importante salientar:

- Deteção automática da superfície, por exemplo, através de um detetor de cantos como é o caso do detetor de Harris, juntamente com a deteção em termos de cor da projeção. Pode-se com este processo lançar uma janela em modo *Full Screen*, com uma cor definida, e através da deteção dessa mesma cor realizar uma binarização da imagem captada e posteriormente obter a localização dos cantos da projeção automaticamente, com recurso a esse mesmo detetor;
- Criação de superfícies e janelas de calibração que não sejam paralelogramas retângulos, possibilitando a deteção de superfícies com inclinação. Para isso seria necessário realizar a deteção da projeção independentemente da forma e criar uma operação matemática que permiti-se mapear a projeção para um ecrã virtual que fosse retangular, e tratar todos os dados obtidos nesse mesmo ecrã virtual;
- Desenvolvimento de mais ações de interação, como são o caso do *zoom*, do *pinch* e do *rotate*, com a utilização das duas mãos. Contudo, era importante ultrapassar as limitações existentes no rastreio do esqueleto quando se tenta realizar estes movimentos, sendo necessário o posicionamento lateral e resolver oclusões em certas partes do corpo do utilizador.
- Implementação de apenas um gesto específico para cada tipo de ação com o sistema, podendo esta implementação recorrer a técnicas de reconhecimento de padrões, como descritores de Fourier, uma vez que possuem invariância à rotação e escala, de forma a melhor

detetar esse mesmo gesto. Realizar um sistema de treino que permita detetar interação caso o gesto aconteça e a mão esteja em contacto com a superfície.

- Desenvolvimento do sistema, utilizando o novo sensor Kinect. Sendo que este apresenta uma maior sensibilidade e precisão em relação ao esqueleto obtido pelo *Skeleton stream*, foi ainda melhorado o sistema de deteção de profundidade, sendo agora muito mais preciso.

Referências

[AD86] R. Adler e P.J. Desmares. Saw touch systems on spherically curved panels. In *IEEE 1986 Ultrasonics Symposium*, pages 289–292. IEEE, 1986.

[ATD01] Alberto Albiol, Luis Torres e Edward J Delp. Optimum color spaces for skin detection. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 1, pages 122–124. IEEE, 2001.

[Bay76] B.E. Bayer. Color imaging array, July 20 1976. US Patent 3,971,065.

[Biz12] Microsoft BizSpark. The microsoft accelerator for kinect, 2012. Ultimo acesso em 2013/02/05. URL: http://www.microsoft.com/bizspark/kinectaccelerator/.

[Bux12] Bill Buxton. Multi-touch systems that i have known and loved, Agosto 2012. Ultimo acesso em 2013/01/27. URL: http://www.billbuxton.com/multitouchOverview.html.

[CH92] T.T. Chairman-Hewett. *ACM SIGCHI curricula for human-computer interaction*. ACM, 1992.

[CLV12] L. Cruz, D. Lucio e L. Velho. Kinect and rgbd images: Challenges and applications. *SIBGRAPI Tutorial*, 2012.

[CMN86] S.K. Card, T.P. Moran e A. Newell. *The psychology of human-computer interaction*. CRC, 1986.

[Con] Aura Conci. Rgb para hsv. Ultimo acesso em 2013/06/17. URL: http://www2.ic.uff.br/~aconci/RGBparaHSV.html.

[DFAB04] A. Dix, J. Finlay, G. Abowd e R. Beale. *Human-computer interaction*. Prentice hall, 2004.

[Ele] Tyco Electronics. Elo touchsystems. Ultimo acesso em 2013/01/27. URL: http://www.elotouch.com/.

[FDF90] James D Foley, Van Dam e S.K. Feiner. *Computer graphics: principles and practice*. Addison-Wesley, 1990.

[FSMA08] B. Freedman, A. Shpunt, M. Machline e Y. Arieli. Depth mapping using projected patterns, April 2 2008. US Patent App. 12/522,171.

[fWT12] Kinect for Windows Team. Near mode: What it is (and isn't), Janeiro 2012. Ultimo acesso em 2013/01/28. URL: http://

blogs.msdn.com/b/kinectforwindows/archive/2012/01/20/near-mode-what-it-is-and-isn-t.aspx.

[Gen12] Freak'n Genius. Freak'n genius, 2012. Ultimo acesso em 2013/02/05. URL: http://www.freakngenius.com/.

[Han05] J.Y. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 115–118. ACM, 2005.

[HW90] Y. Han e R.A. Wagner. An efficient and fast parallel-connected component algorithm. *Journal of the ACM (JACM)*, 37(3):626–642, 1990.

[IKK12] IKKOS. Ikkos training, 2012. Ultimo acesso em 2013/02/05. URL: http://www.ikkostraining.com/.

[Int12] Ubi Interactive. Ubi, 2012. Ultimo acesso em 2013/02/05. URL: http://www.styku.com/business/.

[Jin12] Jintronix. Jintronix, 2012. Ultimo acesso em 2013/02/05. URL: http://www.jintronix.com/.

[Joh72] R.G. Johnson. Touch actuable data input panel assembly, June 27 1972. US Patent 3,673,327.

[Kas84] L.R. Kasday. Touch position sensitive surface, November 20 1984. US Patent 4,484,179.

[KE12] Kourosh Khoshelham e Sander Oude Elberink. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437–1454, 2012.

[KH90] G. Kurtenbach e E.A. Hulteen. Gestures in human-computer communication. *The art of human-computer interface design*, pages 309–317, 1990.

[Mal82] J.B. Mallos. Touch position sensitive surface, August 24 1982. US Patent 4,346,376.

[Man12] ManCTL. Skanect, 2012. Ultimo acesso em 2013/02/05. URL: http://skanect.manctl.com/.

[Met82] N. Metha. A flexible machine interface. *MA Sc. Thesis, Department of Electrical Engineering, University of Toronto*, 1982.

[Mic12a] Microsoft. Kinect for windows sensor components and specifications, 2012. Ultimo acesso em 2013/02/05. URL: http://msdn.microsoft.com/en-us/library/jj131033.aspx.

[Mic12b] Microsoft. Microsoft - pixelsense, 2012. Ultimo acesso em 2013/02/01. URL: http://www.microsoft.com/en-us/pixelsense/default.aspx.

[Mic12c] Microsoft. Pointer flags, Novembro 2012. Ultimo acesso em 2013/06/13. URL: http://msdn.microsoft.com/en-us/library/windows/desktop/hh969211(v=vs.85).aspx/css.

[Mon] Leonardo Monteiro. Formação das cores. Ultimo acesso em 2013/06/17. URL: http://www.ufrgs.br/engcart/PDASR/formcor.html.

[MR97] N. Matsushita e J. Rekimoto. Holowall: designing a finger, hand, body, and object sensitive wall. In *Proceedings of the 10th annual ACM symposium on User interface software and technology*, pages 209–210. ACM, 1997.

[MSL01] J Birgitta Martinkauppi, Maricor N Soriano e Mika V Laaksonen. Behavior of skin color under varying illumination seen by different cameras at different color spaces. In *Photonics West 2001-Electronic Imaging*, pages 102–112. International Society for Optics and Photonics, 2001.

[NAC+95] A.F. Newell, J.L. Arnott, A.Y. Cairns, I.W. Ricketts e P. Gregor. Intelligent system for speech and language impaired people: a portfolio of research. In *Extra-ordinary human-computer interaction*. Cambridge University Press, 1995.

[Nor02] D. Norman. *The design of everyday things*. Basic books, 2002.

[Pli12] Matt Plichta. Multi-touch systems that i have known and loved, Fevereiro 2012. Ultimo acesso em 2013/01/28. URL: http://lab.agent-x.com/2012/02/02/kinect-openni-and-processing/.

[Ram10] Margarida Ramos. Modelos de cor - aditivo e subtrativo, Fevereiro 2010. Ultimo acesso em 2013/06/17. URL: http://api-margaridaramos.blogspot.pt/2010/02/modelos-de-cor-aditivo-e-subtractivo.html.

[Rek02] J. Rekimoto. Smartskin: an infrastructure for freehand manipulation on interactive surfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, pages 113–120. ACM, 2002.

[Rot05] Tim Roth. Dsi - diffused surface illumination, Junho 2005. Ultimo acesso em 2013/01/24. URL: http://iad.projects.zhdk.ch/multitouch/?p=90.

[Saf08] D. Saffer. *Designing Gestural Interfaces: Touchscreens and Interactive Devices*. O'Reilly Media, Incorporated, 2008.

[Spa] RGB Color Space. Color spaces.

[Sty12] Styku. Styku smart fitting room, 2012. Ultimo acesso em 2013/02/05. URL: http://www.styku.com/business/.

[Tec12] GestSure Technologies. Gestsure, 2012. Ultimo acesso em 2013/02/05. URL: http://www.gestsure.com/.

[Uni11] International Telecommunication Union. Bt.601 : Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios, Março 2011. Ultimo acesso em 2013/06/17. URL: https://www.itu.int/rec/R-REC-BT.601/_page.print.

[Vox12] Voxon. Voxiebox, 2012. Ultimo acesso em 2013/02/05. URL: http://get.voxiebox.com/.

[WA12] J. Webb e J. Ashley. *Beginning Kinect Programming with the Microsoft Kinect SDK*. Apress, 2012.

[Zha12] Z. Zhang. Microsoft kinect sensor and its effect. *Multimedia, IEEE*, 19(2):4–10, 2012.

[ZSQ99] Benjamin D Zarit, Boaz J Super e Francis KH Quek. Comparison of five color models in skin pixel classification. In *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999. Proceedings. International Workshop on*, pages 58–63. IEEE, 1999.

Anexo A

Anexos

A.1 Classe desenvolvida para a Segmentação das Mãos

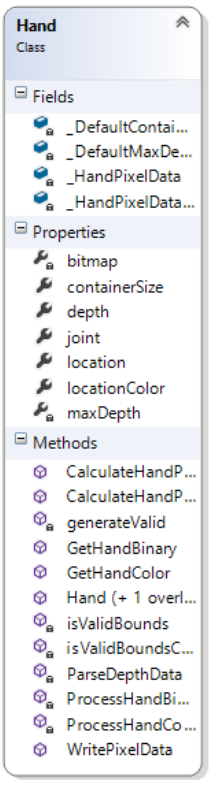


Figura A.1: Classe de processamento que realiza a segmentação das mãos.

A.2 Classe desenvolvida para a Calibração

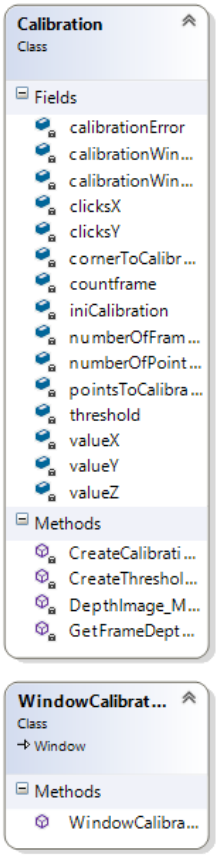
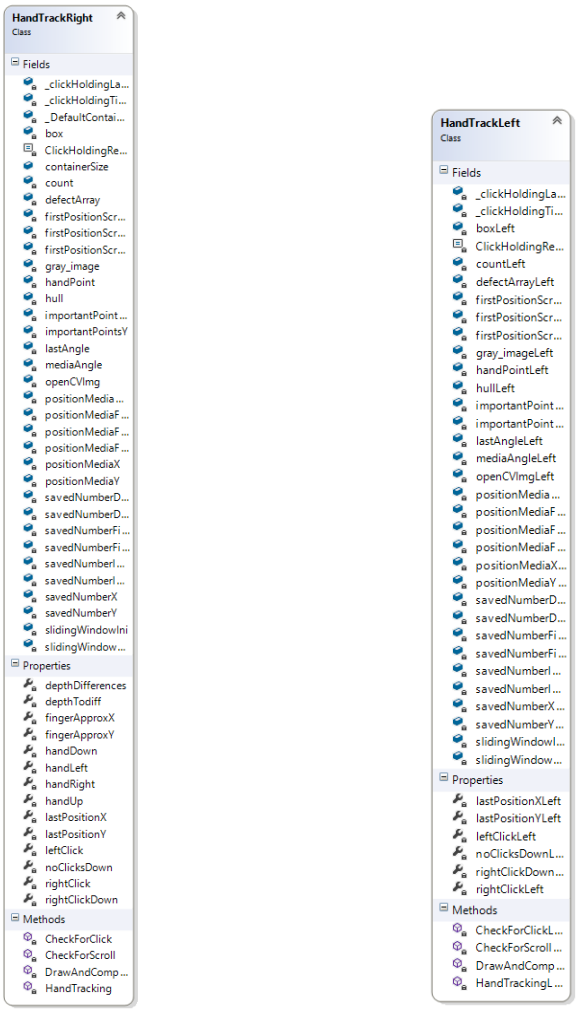


Figura A.2: Classe de processamento que realiza a calibração da superfície.

A.3 Classe desenvolvida para Análise das Características das Mãos e Deteção de Toque



(a) Classe processamento da mão direita

(b) Classe processamento da mão esquerda

Figura A.3: Classes de processamento das mãos do utilizador.

A.4 Classe desenvolvida para o Controlo do Computador

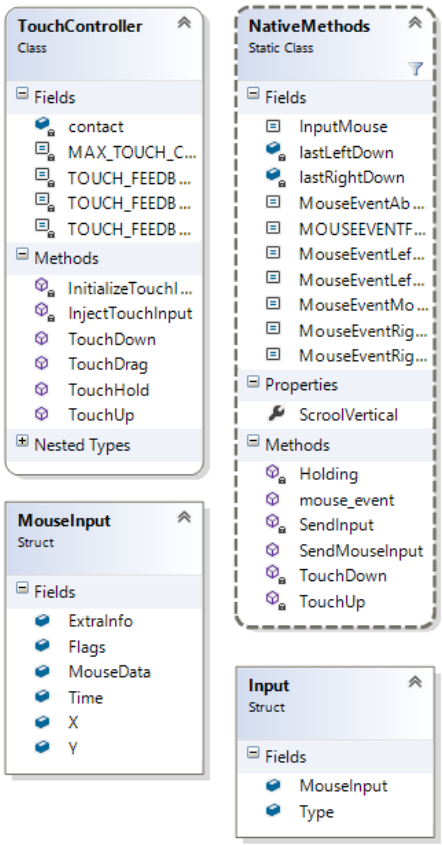


Figura A.4: Classe de processamento que realiza a ligação dos tipos de toque ao computador.