



IPCB



Instituto Politécnico de Castelo Branco  
Escola Superior de Tecnologia

Mestrado em Desenvolvimento  
de Software e Sistemas Interactivos

NATURAL INTERACTIONS:  
An application for Gestural Hands Recognition

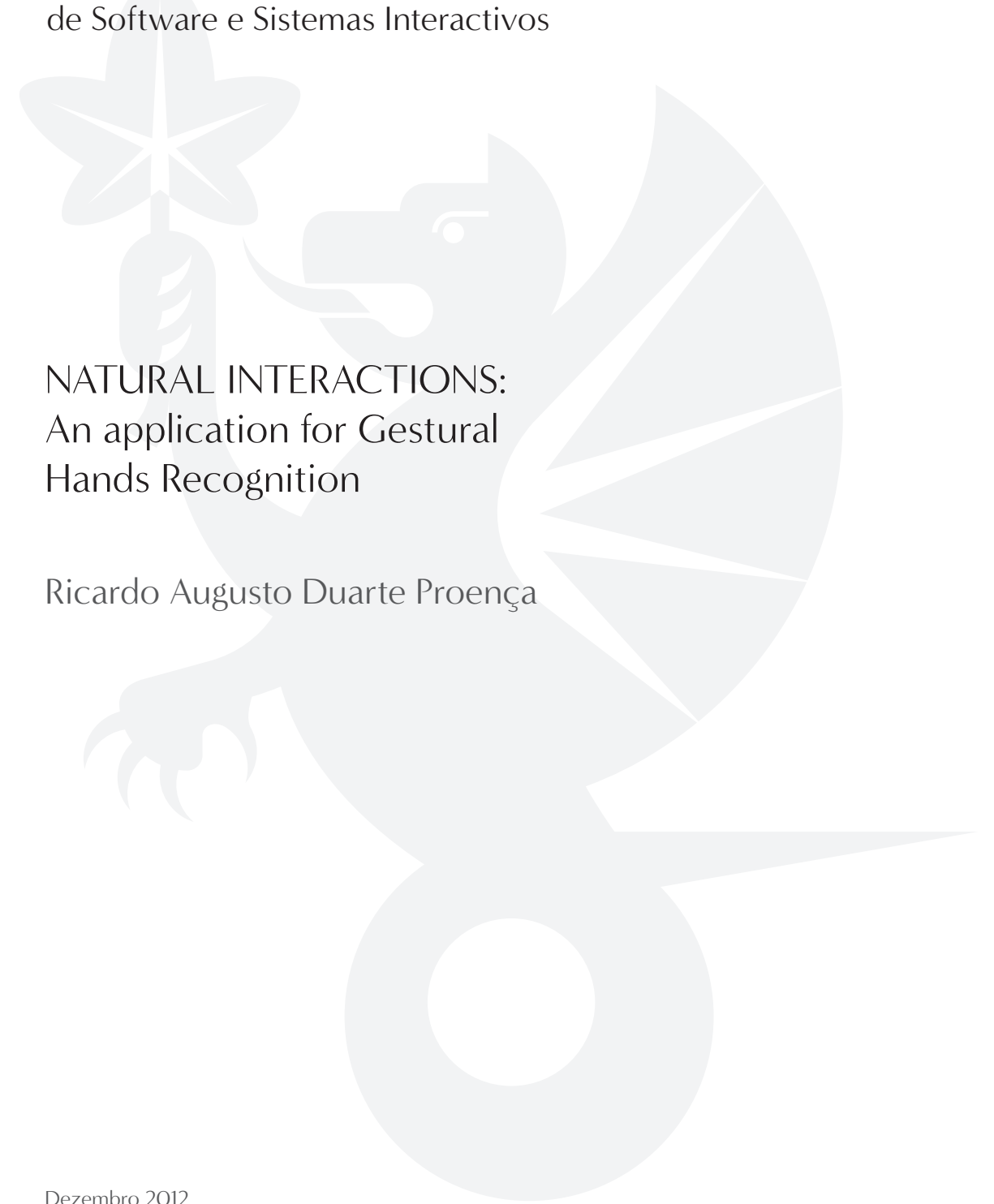
Ricardo Augusto Duarte Proença

Dezembro 2012

NATURAL INTERACTIONS:  
An application for Gestural  
Hands Recognition

Ricardo Augusto Duarte Proença

Dezembro 2012





Instituto Politécnico de Castelo Branco  
Escola Superior de Tecnologia



# NATURAL INTERACTIONS: An application for Gestural Hands Recognition

Ricardo Augusto Duarte Proença

Trabalho de Mestrado  
Desenvolvimento de Software e Sistemas Interactivos

Trabalho Efectuado sob a orientação da  
Professora Doutora Arminda Guerra Lopes

Dezembro 2012

Instituto Politécnico de Castelo Branco  
Escola Superior de Tecnologia

# NATURAL INTERACTIONS: An application for Gestural Hands Recognition

Ricardo Augusto Duarte Proença

Dissertação apresentada ao Instituto Politécnico de Castelo Branco para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Desenvolvimento de Software e Sistemas Interactivos, realizada sob a orientação científica da Doutora Arminda da Conceição dos Santos Guerra Lopes, Professora da Unidade Técnico-Científica de Informática do Departamento de Informática de Escola Superior de Tecnologia do Instituto Politécnico de Castelo Branco.



Dedico este trabalho ao meu filho Pedro, para que tenha um futuro brilhante.



### **Agradecimentos**

À minha orientadora, por todas as suas críticas, sugestões e pelo seu entusiasmo contagiante transmitindo-me sempre a motivação, confiança e vontade necessárias para a realização deste trabalho.

À minha família, por toda a compreensão e tolerância que depositou em mim desde o primeiro momento.

Aos meus pais, por me inculcaram o amor pelo estudo, à realização profissional entre outros valores que regem a minha vida.

À minha esposa, pela sua paciência, espírito de sacrifício e pela coragem que sempre me transmitiu, sem a qual eu não teria conseguido chegar ao fim.

Ao meu filhote por me conseguir fazer sorrir mesmo nos momentos mais difíceis.



### **Palavras-chave**

Interação Homem-máquina, reconhecimento de gestos, visão computacional, tecnologias assistivas

### **Resumo**

Este trabalho apresenta um sistema para o desenvolvimento de novas interfaces homem-máquina com foco no reconhecimento de gestos estáticos de mãos humanas. A proposta é auxiliar o acesso a certos objetos para o ocupante de uma cadeira de rodas inteligente, a fim de facilitar a sua vida diária. A metodologia proposta baseia-se no uso de processos computacionais simples e de *hardware* de baixo custo. O seu desenvolvimento envolve uma abordagem completa aos problemas de visão computacional, com base nas etapas da captura de imagem de vídeo, segmentação de imagens, extração de características, reconhecimento e classificação de padrões. A importância deste trabalho relaciona-se com a necessidade de construir novos modelos de interação que permitam, de uma forma natural e intuitiva, a simplificação da vida quotidiana de uma pessoa com dificuldades motoras.



**Keywords**

Human-Computer Interaction, gesture recognition, computer vision, disable person, assistive technology

**Abstract**

This thesis presents a system for the development of new human-machine interfaces focused on static gestures recognition of human hands. The proposal is to give access to certain objects to the occupant of an intelligent wheelchair in order to facilitate their daily life. The proposed methodology relies on the use of simple computational processes and low-cost hardware. Its development involves a comprehensive approach to the problems of computer vision, based on the steps of the video image capture, image segmentation, feature extraction, pattern recognition and classification. The importance of this work relates to the need to build new models of interaction that allow, in a natural and intuitive way, to simplify the daily life of a disable person.



# Índice geral

1.	Introdução .....	1
1.1.	Enquadramento.....	1
1.2.	Objetivos .....	2
1.3.	Estrutura do relatório.....	2
2.	O contexto: Tecnologias Assistivas .....	4
2.1.	Veículos Autónomos Assistidos.....	5
3.	Estado da Arte .....	7
4.	Solução Proposta.....	11
5.	Interações Naturais .....	14
5.1.	Tipos de NUI .....	16
5.1.1.	Interfaces multimodais.....	16
5.1.2.	Interfaces tácteis .....	16
5.1.3.	Interfaces tangíveis .....	17
5.1.4.	Interfaces baseadas no reconhecimento da voz.....	17
5.1.5.	Interfaces baseadas no rastreamento dos olhos .....	17
5.1.6.	Interfaces gestuais.....	18
6.	Desenho da interface: descrição do processo .....	23
6.1.	Enquadramento conceptual.....	23
6.1.1.	Conceitos básicos .....	23
6.1.2.	Processamento Digital de Imagens .....	28
6.1.3.	Aquisição de imagem .....	29
6.1.4.	Pré-processamento .....	29
6.1.5.	Segmentação .....	30
6.1.6.	Extração de características.....	34
6.1.7.	Descritores .....	37
6.1.8.	Classificação e reconhecimento de padrões.....	38
6.2.	Hardware Utilizado.....	39
6.3.	Software utilizado .....	40
6.4.	Base de Dados de Gestos .....	41
6.5.	Visão Geral do sistema .....	42
6.6.	Arquitetura do Sistema .....	42
6.6.1.	Captura da imagem .....	42
6.6.2.	Pré-processamento .....	43
6.6.3.	Segmentação .....	44

6.6.4.	Pós-processamento .....	47
6.6.5.	Deteção de feições .....	48
6.6.6.	Extração das características .....	50
6.6.7.	Reconhecimento de padrões.....	51
6.6.8.	Considerações .....	52
7.	Testes.....	54
7.1.	Base de Dados .....	54
7.2.	Segmentação da mão .....	54
7.3.	Extração de características.....	56
7.4.	Treino e reconhecimento SVM.....	59
7.5.	Resultados da classificação .....	59
7.6.	Tempos totais de processamento .....	61
8.	Conclusão .....	63
8.1.	Conclusões.....	63
8.2.	Trabalho Futuro.....	64
9.	Bibliografia.....	66
10.	Anexos - Base de Dados de Imagens .....	72
10.1.	Anexo A - Gestos da BD .....	73
10.2.	Anexo B - Gestos utilizados para teste .....	74

# Índice de figuras

FIG. 1 – SITUAÇÃO-PROBLEMA .....	11
FIG. 2 – SISTEMA PROPOSTO .....	12
FIG. 3 – GESTOS RECONHECIDOS PELO SISTEMA.....	12
FIG. 4 – CLASSIFICAÇÃO DOS SISTEMAS DE IHM E ÁREAS DE APLICAÇÃO .....	21
FIG. 5 - ESPAÇO DE COR RGB (HTTP://MSDN.MICROSOFT.COM/EN-US/LIBRARY/WINDOWS/DESKTOP/DD372185(V=VS.85).ASPX) .....	24
FIG. 6 – CANAIS DE CORES: A) IMAGEM RGB B) CANAL RED B) CANAL GREEN D) CANAL BLUE (HTTP://EN.WIKIPEDIA.ORG/WIKI/LENNA) .....	25
FIG. 7 - ESPAÇO DE CORES HSV (HTTP://MSDN.MICROSOFT.COM/EN-US/LIBRARY/WINDOWS/DESKTOP/AA511283.ASPX) ....	25
FIG. 8 - IMAGEM CONVERTIDA PARA ESCALA DE CINZA (HTTP://EN.WIKIPEDIA.ORG/WIKI/LENNA).....	26
FIG. 9 – HISTOGRAMA: A) IMAGEM EM TONS DE CINZA B) HISTOGRAMA DA IMAGEM Na FIG. 9A) PODEMOS VERIFICAR A EXISTÊNCIA DE DOIS TIPOS PRINCIPAIS DE OBJETOS QUE SÃO O CÉU E A AREIA DO DESERTO E NA FIG. 9B) CONFIRMAMOS A REPRESENTAÇÃO DESTAS REGIÕES ATRAVÉS DOS DOIS PICOS DO HISTOGRAMA, INDICANDO QUE ESTES NÍVEIS DE CINZA OCORREM COM MAIS FREQUÊNCIA NA IMAGEM. ....	27
FIG. 10 - IMAGEM BINÁRIA (HTTP://EN.WIKIPEDIA.ORG/WIKI/LENNA) .....	28
FIG. 11 – SISTEMA BÁSICO DE PROCESSAMENTO DE IMAGENS DIGITAIS .....	29
FIG. 12 – DETECÇÃO DE BORDAS (HTTP://EN.WIKIPEDIA.ORG/WIKI/LENNA) .....	31
FIG. 13 – A) LIMIAR BAIXO B) LIMIAR ÓTIMO C) LIMIAR ALTO (HTTP://EN.WIKIPEDIA.ORG/WIKI/LENNA).....	32
FIG. 14 – SUBTRAÇÃO DE FUNDO: A) QUADRO CAPTADA B) IMAGEM COM FUNDO SUBTRAÍDO (HTTP://WWW.BOX10.COM/SKILL-JAM.HTML) .....	34
FIG. 15 – TIPOS DE CARACTERÍSTICAS .....	35
FIG. 16 - ÉTAPAS FUNDAMENTAIS DO TRABALHO PROPOSTO .....	42
FIG. 17 – QUADRO CAPTURADO COM A MÃO BEM POSICIONADA.....	43
FIG. 18 – REGIÃO DE INTERESSE (ROI).....	43
FIG. 19 – ROI SUAVIZADA .....	44
FIG. 20 – PROCESSO DE SEGMENTAÇÃO DE MÃO HUMANA .....	44
FIG. 21 – A) IMAGEM RGB NORMALIZADO .....	45
FIG. 22 – CANAL VERMELHO EXTRAÍDO .....	46
FIG. 23 – A) CANAL VERMELHO EXTRAÍDO COM SUBTRAÇÃO DE FUNDO.....	46
FIG. 24 – HISTOGRAMA DA IMAGEM COM O VALOR ÓTIMO DE LIMIAR IDENTIFICADO.....	47
FIG. 25 – IMAGEM BINARIZADA .....	47
FIG. 26 – IMAGEM APÓS A APLICAÇÃO DOS FILTROS MORFOLÓGICOS.....	48
FIG. 27 – CONTORNOS DA MÃO .....	48
FIG. 28 – INVÓLUCRO CONVEXO DA MÃO.....	49
FIG. 29 – SALIÊNCIAS CONVEXAS DO CONTORNO .....	49
FIG. 30 – RETÂNGULO ENVOLVENTE .....	50
FIG. 31 – RETÂNGULO ENVOLVENTE NORMALIZADO.....	50
FIG. 32 – SEPARAÇÃO DE CLASSES UTILIZANDO O ALGORITMO SVM.....	52

FIG. 33 – REPRESENTAÇÃO DAS ETAPAS DA SEGMENTAÇÃO COM VÁRIOS TIPOS DE ILUMINAÇÃO.....	55
FIG. 34 - SEGMENTAÇÃO EM FUNDO NÃO UNIFORME .....	56
FIG. 35 – CÓDIGO PARA CALCULAR AS CARACTERÍSTICAS.....	57
FIG. 36 – NORMALIZAÇÃO DAS CARACTERÍSTICAS EXTRAÍDAS.....	58
FIG. 37 – CÓDIGO PARA CÁLCULO DO TEMPO DE PROCESSAMENTO.....	61

# Índice de tabelas

TAB. 1 – PRINCIPAIS PROJETOS SOBRE CADEIRAS DE RODAS.....	10
TAB. 2 – CARACTERÍSTICAS DO COMPUTADOR UTILIZADO .....	39
TAB. 3 – CARACTERÍSTICAS DA CÂMARA WEB UTILIZADA .....	40
TAB. 4 – MATLAB VS. OPENCV .....	41
TAB. 5 – CARACTERÍSTICAS DO SOFTWARE UTILIZADO .....	41
TAB. 6 – LIMIAR ÓTIMO PARA SEGMENTAÇÃO.....	55
TAB. 7 – VALORES NORMALIZADOS DO VETOR DE CARACTERÍSTICAS .....	57
TAB. 8 – VALORES DOS PARÂMETROS DO VETOR DE CARACTERÍSTICAS .....	59
TAB. 9 – CLASSIFICAÇÃO UTILIZANDO DIFERENTES FUNÇÕES DE KERNEL PARA AS AMOSTRAS DE TREINO E TESTE .....	59
TAB. 10 – TABELA COM VALORES TOTAIS DOS TESTES REALIZADOS. ....	60
TAB. 11 – RESULTADOS DO RECONHECIMENTO DOS GESTOS PREDEFINIDOS. ....	60
TAB. 12 – BALANÇO FINAL DO RECONHECIMENTO DE GESTOS. ....	60
TAB. 13 – RESULTADOS OBTIDOS EM TERMOS DE TEMPOS DE PROCESSAMENTO .....	62

## Lista de abreviaturas

BD	Base de Dados
CLI	Command Line Interface (Interface por linha de comandos)
GUI	Graphical User Interface (Interface gráfica)
ICA	Análise de Componentes Independentes
IG	Interface Gestual
IHM	Interação Homem-Máquina
KNN	K Nearest Neighbours (K-vizinhos mais próximos)
LDA	Análise Discriminante Linear
PCA	Análise de Componentes Principais
NUI	Natural User Interfaces (Interfaces naturais)
T	Limiar Ótimo
TA	Tecnologias Assistivas
TUI	Tangible User Interfaces (Interfaces tangíveis)
VNS	Very Nervous System
VC	Visão Computacional

# 1. Introdução

## 1.1. Enquadramento

Testemunhamos passivamente o crescimento dos acidentes rodoviários e o crescente número de pessoas que sofrem sequelas variadas, algumas das quais afetam totalmente a mobilidade do indivíduo no seio da sociedade. Tal facto, condiciona inevitavelmente o seu bem-estar físico, psicológico e social. Porém, são vários os efeitos colaterais, alguns dos quais passam mesmo pela necessidade de recorrer a meios de mobilidade alternativos como a cadeira de rodas.

O aumento no número de pessoas com dificuldades motoras reclama, desde há muito tempo, tecnologias assistivas que proporcionem um aumento da qualidade de vida a todos os indivíduos, sem exceção, o que nem sempre acontece, porque muitos são os obstáculos com que a tecnologia se depara, sendo necessários meios muito dispendiosos ou estudos mais aprofundados para melhorar o acesso a qualquer cidadão.

Para que me serve uma cadeira de rodas inteligente, quando a mobilidade não é totalmente completa? Posso andar livremente nos espaços preparados para o efeito, mas como concretizar ações autónomas, independentes, tão simples como o apanhar um pequeno objeto de que necessito, sem a presença de terceiros?

Estas questões, aparentemente inócuas, são sintomáticas de que nem tudo está devidamente pensado numa sociedade que pugna pela igualdade de oportunidades. Basta, por

isso, parar uns minutos e recolher factos que perpassam pela retina dos nossos olhos, com um olhar analítico que permita agir.

Por conseguinte, existe a necessidade de construir novos modelos de Interação Homem-Máquina (IHM) que permitam, de uma forma natural e intuitiva, a manipulação de máquinas que facilitem e proporcionem, incondicionalmente, uma interação com o mundo de forma transparente, simples e sem recursos económicos demasiadamente dispendiosos e que proporcionem uma vida mais autónoma perante o quotidiano.

## 1.2. Objetivos

O trabalho proposto tem como objetivo principal o desenvolvimento de um sistema de reconhecimento de gestos estáticos de mãos humanas de diferentes pessoas efetuados apenas com uma mão, situadas em frente de uma câmara de vídeo digital. O cenário de fundo deverá ser simples e uniforme. A mão humana é apresentada para a câmara sem nenhum dispositivo assistivo que ajude a sua deteção. Esta interface pode ser aplicada a uma cadeira de rodas inteligente, a fim de aumentar a qualidade da vida das pessoas com incapacidade e/ou envelhecidas.

Pretendemos propor a possibilidade de construção de um mecanismo de IHM considerando o mínimo de restrições de ambiente ao nível da iluminação, para que a interação seja feita da maneira mais natural possível. No entanto, poderá ser necessário efetuar alguns ajustes iniciais no ambiente de forma a não comprometer a sua usabilidade.

Foram definidos 6 gestos que devem ser executados com a mão direita, sendo o reconhecimento efetuado através de uma base de dados (BD) com os gestos predefinidos.

Ao nível do reconhecimento de gestos, a forma geométrica da mão assume um papel primordial. O reconhecimento será baseado apenas ao nível das bordas da mão, sendo, portanto, suficientemente rápido para poder ser aplicado em tempo real.

O desenvolvimento deste projeto comporta uma abordagem completa aos problemas de VC, baseado nas etapas de captura da imagem de vídeo, segmentação da imagem, extração de características, reconhecimento de padrões treinados e interpretação de resultados.

Este trabalho procura que o sistema desenvolvido opere em tempo real, obrigando a um esforço no sentido de minimizar os tempos de processamento das tarefas pretendidas. O ideal seria possuir capacidades de processamento maiores que 30 quadros por segundo mas, neste trabalho, propomos uma capacidade de processamento por volta dos 30 quadros por segundo devido ao esforço computacional exercido pelos algoritmos utilizados.

## 1.3. Estrutura do relatório

No *Capítulo 1 - Introdução* -Faz-se um enquadramento do tema a desenvolver e apresentam-se os seus objetivos principais.

No *Capítulo 2 - O contexto: Tecnologias Assistivas* - exploramos a temática da mobilidade para pessoas com dificuldades e apresentamos uma resenha sobre veículos autónomos assistidos.

No *Capítulo 3 - Revisão do Estado da Arte*, pretendemos apresentar o estado atual do desenvolvimento de mecanismos de interação com cadeiras de rodas inteligentes.

No *Capítulo 4 - Solução Proposta* - propomos expor o contexto, a motivação e a justificação para o desenvolvimento deste trabalho.

No *Capítulo 5 - Interações Naturais* - desenvolvemos a noção de interface natural, tipificando-a e apresentando exemplos da sua utilização.

No *Capítulo 6 - Desenho da Interface: descrição do processo* - discutimos conceitos importantes relativos a todas as etapas do desenvolvimento do sistema proposto sem os quais nos parece impossível descrevê-lo para depois traçarmos o desenvolvimento do sistema de reconhecimento de gestos, incluindo as abordagens adotadas para resolver as diferentes etapas do problema e sua implementação.

No *Capítulo 7 - Testes* - mostramos uma análise aos testes efetuados ao sistema e apresentamos os resultados obtidos bem como as suas aplicações e limitações.

No *Capítulo 8 - Conclusões* - expomos as conclusões do estudo e sugerimos as suas aplicações futuras.

## 2. O contexto: Tecnologias Assistivas

“Para as pessoas sem deficiência, a tecnologia torna as coisas mais fáceis. Para as pessoas com deficiência, a tecnologia torna as coisas possíveis.” (Radabaugh, 2012)

Desde sempre, a evolução tecnológica caminhou no sentido da facilitação da vida. Utilizamos diariamente recursos particularmente desenvolvidos para auxiliar e simplificar os dinamismos do quotidiano, como as canetas, automóveis, computadores, telefones, um infinito catálogo de ferramentas, que já fazem parte integrante da nossa rotina.

O termo Tecnologia Assistiva (TA) é utilizado, no presente trabalho, para identificar todo o conjunto de meios que possa contribuir para facultar ou alargar as capacidades operantes de pessoas com deficiência, de molde a promover uma vida independente e a efetiva inclusão social.

A TA deve ser pensada como um apoio para fomentar o aumento de uma prática deficitária ou proporcionar a execução de uma função pretendida que se encontra impedida por motivo de deficiência ou envelhecimento. O seu objetivo primordial é facultar uma maior independência, qualidade de vida e inclusão social à pessoa com deficiência, através da ampliação da sua comunicação, mobilidade, controlo de ambiente circundante, habilidades e trabalho (Bersch, 2009).

Neste contexto, a mobilidade pode ser auxiliada por bengalas, muletas, andarilhos, carrinhos, e cadeiras de rodas manuais ou elétricas e qualquer outro veículo, equipamento ou estratégia. As cadeiras de rodas são um artefacto preponderante e bastante utilizado para as pessoas portadoras de incapacidades motoras, usualmente recomendadas a utilizadores que não

possuam força nem coordenação motora para poderem andar. A escolha por este tipo de veículo toma em consideração as habilidades sensoriais e motoras do utilizador, como o seu desejo de se locomover de modo independente (Bersch, 2009).

## 2.1. Veículos Autónomos Assistidos

Desde os anos 80 que existem estudos com a finalidade de ajudar a locomoção de pessoas dependentes de cadeiras de rodas motorizadas, através do desenvolvimento de técnicas nas áreas de robótica móvel. De início, as propostas abordavam a cadeira como veículos autogeridos, desenvolvidos para operar em parques fabris. Colocavam-se marcas de referência no próprio ambiente e a localização do veículo era encontrada através do reconhecimento dessas marcas para gerar automaticamente a trajetória ideal para um certo local estabelecido (Madarasz, Heiny, Crompt, & Mazur, The design of an autonomous vehicle for the disabled, 1986), ou seguia linhas magnéticas colocadas sobre o piso do ambiente, sendo que todas as trajetórias possíveis já se encontravam pré-estabelecidas (Wakaumi, Nakamura, & Matsumura, 1992).

As pesquisas eram então centradas em veículos capazes de seguir trajetórias e de desviar-se de obstáculos que eram providos de sensores, como câmaras de vídeo e ultrassom (Mazo, et al., 1995), (Katevas, et al., 1997), (Wang, Kang, Ishimatsu, & Ochiai, 1997) e debruçavam-se em volta dos problemas de controlo, navegação e desvio de obstáculos sem que fosse dada muita importância ao utilizador.

O veículo era tratado como um robô móvel e o utilizador como uma carga a ser transportada de um ponto para outro num ambiente adaptado ao "robô".

Só em 1994 é que este paradigma se alterou, com a introdução da ideia de controlo assistido para desvio de obstáculos (Bell, Borenstein, Levine, Koren, & Jaros, 1994). Em 1996, pela primeira vez, foi apresentado um veículo autónomo para pessoas deficientes, conhecido como VAHM (Bourhis & Pino, 1996), com inovações ao nível do controlo, ou seja, o veículo dispunha de três modos de navegação: manual, assistido e autónomo, permitindo assim ao utilizador escolher o nível de controlo desejado.

No modo manual, o utilizador dirige o veículo através de uma interface, como por exemplo, o *joystick*. No modo assistido, o veículo passava a seguir comandos dados pelo utilizador como: " *siga o corredor*", " *contorne a parede*", ou " *passe pela porta*", e evitava assim colisões com obstáculos. No modo autónomo, o utilizador escolhia o destino pretendido e o veículo gerava as trajetórias automaticamente, sem a intervenção do utilizador. Portanto, o utilizador escolhia o nível de controlo mais adequado a uma determinada situação.

Esta nova abordagem trouxe novas questões para as áreas de investigação, indo mais além de que um típico projeto mecatrónico. Incluindo várias áreas, tornou-se necessário conciliar profissionais da área das ciências exatas, humanas e biológicas. A partir daí, os trabalhos desenvolvidos focalizaram-se essencialmente em técnicas de controlo assistido e adaptativo, fáceis de interagir com o utilizador (Tahboub & Asada, 1999), (Luo, Chen, & Lin, 1999), (Nakanishi, Kuno, Shimada, & Shirai, 1999), chegando-se então ao termo que ainda hoje é aplicado: cadeira de rodas inteligente. Hoje, estas cadeiras denotam um desenvolvimento apontando para a maximização da funcionalidade da cadeira, do conforto, independência e

autoestima do utilizador. São incorporadas tecnologias de biomecânica, ergonomia, interface e controlo que passaram a permitir ao utilizador uma navegação e desvio de obstáculos mais segura e confortável.

### 3. Estado da Arte

Na revisão de literatura que efetuámos, deparámo-nos com várias propostas de cadeiras de rodas inteligentes para satisfazer a necessidade de navegação e de desvio de obstáculos.

No entanto, verificamos que a grande maioria dos estudos apenas se relaciona com a deslocação da cadeira única e exclusivamente, sem haver preocupações de mobilidade mais completa.

Na verdade, esta relação não é totalmente eficaz para a resolução dos conflitos que um indivíduo sem mobilidade tem de ultrapassar.

Uma cadeira de rodas inteligente robotizada que observa o utilizador e o meio ambiente em que este está inserido é proposta por (Kobayashi, Kinpara, Shibusawa, & Kuno, 2009). É pretendido compreender as intenções do utilizador a partir dos seus comportamentos e das informações ambientais, mesmo quando este se encontra fora da cadeira de rodas e reconhece os comandos indicados por gestos. Embora o sistema seja baseado no reconhecimento facial, ele pode ser modificado para usar os movimentos da boca, olhos, ou quaisquer outras partes do corpo para se poder adaptar a qualquer dificuldade do utilizador.

Também (Kumar & Dinesh, 2012) expõem uma proposta cuja ideia é controlar a cadeira de rodas usando reconhecimento gestual para pessoas com deficiência física. Os autores centraram-se mais nos tipos de gestos para estudar como estes influenciam o reconhecimento através do computador e descobriram que existem gestos típicos para cada função, que capturam e guardam para a formulação de um padrão específico, usado no processo de obtenção dos

resultados. É utilizada uma câmara de vídeo para capturar as imagens dos gestos de mão comparando-o com os dados definidos e pré-armazenados para, de seguida, enviar o sinal ao microcontrolador que controla o motor específico para mover a cadeira de rodas.

(Chhabria & Dharaskar, 2012) apresentam uma metodologia para aplicar técnicas de reconhecimento de gestos, reconhecimento de voz e rastreamento dos olhos ao nível do controlo do movimento da cadeira de rodas. Ao nível da interface gestual, é apresentado um esquema que reduz o tamanho da BD de gestos, usado para armazenar diferentes posturas de seres humanos. Para interagir através dos olhos, um novo algoritmo é proposto para reconhecer o movimento do olho para a esquerda ou direita. O recurso ao piscar os olhos também não foi esquecido e pode ser utilizado pelo algoritmo para controlar o arranque e paragem da cadeira de rodas.

Para pessoas com maiores dificuldades, o trabalho de (Jia, Hu, Lu, & Yuan, 2007) apresenta um sistema de controlo de mãos livres para cadeiras de rodas inteligentes que assenta numa interface baseada no reconhecimento visual de gestos da cabeça do utilizador da "RoboChair". Os gestos reconhecidos são utilizados para gerar os comandos de controlo de movimento para o controlador de sinais digitais de movimento de baixo nível para que possa controlar o movimento do robô de acordo com a intenção do utilizador. São utilizados os algoritmos Adaboost para a deteção de rosto e o Camshift para rastreamento de objetos que, combinados no sistema, alcançam uma deteção de rosto precisa e um rastreamento e reconhecimento de gestos em tempo real. Este é um sistema extremamente útil para os utilizadores que têm restrições dos movimentos dos membros, causadas por doenças como a doença de Parkinson e a tetraplegia.

(Kobayashi, Kinpara, Shibusawa, & Kuno, 2009), propõem uma cadeira de rodas robotizada que observa o utilizador e o meio ambiente em que este está inserido. Ele pretende compreender as intenções do utilizador a partir dos seus comportamentos e das informações ambientais mesmo quando este se encontra fora da cadeira de rodas e reconhece os comandos indicados por gestos.

Embora o sistema seja baseado no reconhecimento facial, ele pode ser modificado para usar os movimentos da boca, olhos, ou quaisquer outras partes do corpo, a fim de se poder adaptar a qualquer dificuldade do utilizador.

(Shan, Tan, & Wei, 2007) utilizam o filtro de partículas e o deslocamento médio, do inglês "*Particle Filtering*" e "*Meanshift*", que são duas abordagens bem sucedidas no reconhecimento de objetos por integrar as suas vantagens num algoritmo que permita o desenvolvimento de uma interface de controlo para uma cadeira de rodas de mão robotizada através do reconhecimento de gestos dinâmicos.

Em (Tsui, Kimb, Behal, Kontak, & Yanco, 2011) é apresentado um sistema completo que, através de um braço robótico, permite ao utente recolher autonomamente um objeto desejado de uma prateleira. A posição do objeto é calculada utilizando visão estereoscópica, a partir de uma câmara colocada no ombro do utilizador, e o braço de robô move-se para a posição mais adequada e ajusta a sua orientação, caso seja necessário, de modo a que a sua garra esteja perpendicular ao objeto pretendido. O braço do robô compara a imagem da câmara com as imagens da BD e escolhe a melhor combinação. Depois, o braço do robô move-se para alinhar os

pontos característicos escolhidos, desloca-se para frente e fecha a sua garra para agarrar o objeto.

A interface é bastante intuitiva e tem dois modos de operação: através de um ecrã tátil ou de um *joystick*. Esta interface mostra a visão da câmara, semelhante à visão do utilizador, que apenas precisa de apontar (com a mão no ecrã tátil ou através do *joystick*) o objeto no ecrã para recolher o objeto

A Tab. 1 apresenta uma visão geral dos principais projetos sobre cadeiras de rodas inteligentes que utilizam a visão computacional como recurso, contendo uma breve descrição sobre os objetivos propostos e as tecnologias utilizadas por cada grupo de investigação.

Trabalho	Descrição
Autonomous Wheelchair (Madarasz, Heiny, Crompt, & Mazur, The design of an autonomous vehicle for the disabled, 1986)	Navegação autónoma através de um mapa em ambientes desordenados.
Smart Wheelchair (Nisbet, Craig, Odor, & Aitken, 1995)	Utilizada para ajuda no treino de mobilidade. Segue linhas e move-se para trás quando colide com objetos.
CPWNS (Yoder, Baumgartner, & Skaar, 1996)	Reproduz rotas que são ensinadas ao sistema através da direção manual da cadeira de rodas entre o ponto inicial e o ponto final.
The Intelligent Wheelchair (Gribble, Browning, Hewett, Remolina, & Kuipers, 1998)	Navegação autónoma através do reconhecimento de marcas artificiais no chão.
Tetra Nauta (Balcells & Gonzalez, 1998)	Navegação autónoma através de marcas artificiais no solo.
Wheelesley (Yanco, 1998)	Navegação assistida.
Rolland	Monitoriza e aprende as características do ambiente

(Roefler & Lankenau, 2000)	enquanto navega para planejar trajetórias. Através de treino prévio, aprende igualmente o desvio de obstáculos.
LOUSON III (Tse & Luo, 2000)	Controlo para seguir um alvo.
Intelligent Wheelchair System (Murakami, Kuno, Shimada, & Shirai, 2001)	O utilizador comanda a cadeira de rodas através de reconhecimento de expressões faciais. A resposta da cadeira adapta-se ao ambiente em que está inserida.

Tab. 1 - Principais projetos sobre cadeiras de rodas

## 4. Solução Proposta

Considerando o caso de um indivíduo que necessita de ter acesso a um objeto que se encontra colocado numa superfície com grande área, facto inibidor da sua ação normal e autónoma, urge pensar uma nova alternativa que ajude a minimizar essa inibição

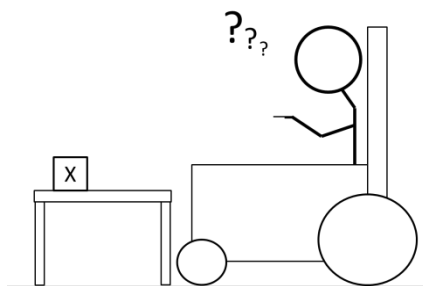


Fig. 1 - Situação-problema

É nosso entendimento, no seio da investigação que temos vindo a realizar, que face à evolução das tecnologias de aquisição e processamento de imagens, é possível construir um sistema computacional que faça com que as máquinas reconheçam gestos humanos. Todavia, a construção deste sistema pressupõe a utilização de vários algoritmos de programação bastante complexos. Elencamos genericamente os procedimentos que tal acarreta: (1) captura da imagem; (2) segmentação da mão humana; (3) extração de características; (4) classificação dos gestos.

Diferentes tipos de algoritmos para o reconhecimento de gestos têm sido estudados, no entanto, o grande desafio prende-se com o facto de tornar esta interação independente de

dispositivos que facilitem a identificação da mão humana, como luvas ou sensores de captura de movimento.

Consideramos que o desenvolvimento de algoritmos de identificação de gestos, com base em câmaras de vídeo de baixo custo, é um grande estímulo para a área da IHM.

Face aos resultados que temos obtido no decurso da presente investigação, cremos ser de propor um sistema que permita ajudar o ocupante de uma cadeira de rodas a interagir com o mundo, nomeadamente a ter acesso a determinados objetos, com vista a facilitar o seu quotidiano.

Ao nível de *hardware*, o protótipo é composto por uma cadeira de rodas inteligente, um computador portátil com uma webcam incorporada e um braço mecânico. Quanto ao *software*, este é constituído por sistema de reconhecimento de gestos estáticos, efetuado em tempo real, através de uma mão situada em frente à webcam.

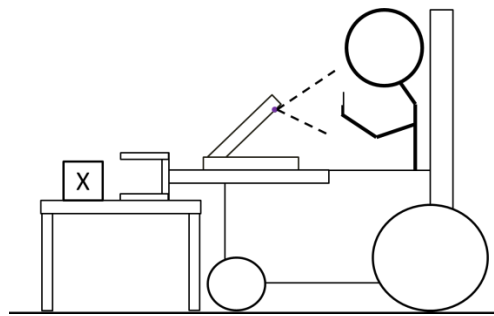


Fig. 2 - Sistema proposto

Face a este enquadramento, podemos já afirmar que, para darmos cumprimento a este conceito de mobilidade, temos por base o princípio de que cada gesto executado será transposto para uma ação do braço mecânico. Essa transposição será definida a partir dos 6 gestos que a seguir apresentamos, onde cada um cumpre uma função específica no contexto de facilitação da mobilidade do indivíduo, a saber: (a) pausa; (b) esticar braço mecânico com velocidade 1; (c) esticar braço mecânico com velocidade 2; (d) abrir garra; (e) fechar garra; (f) recolher braço até à posição inicial.

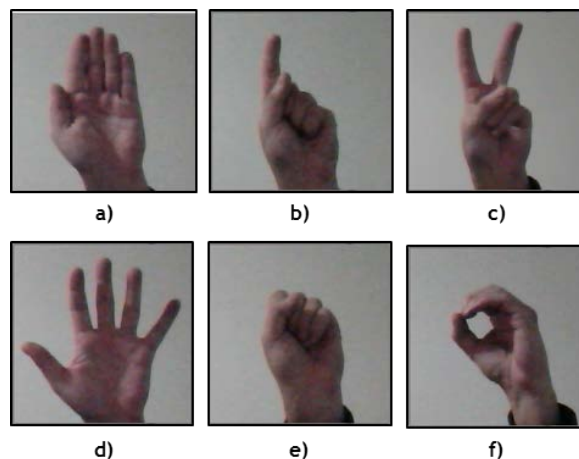


Fig. 3 - Gestos reconhecidos pelo sistema

Pretendemos demonstrar com este método que é possível construir um mecanismo de IHM, facilitando a vida do indivíduo portador de inibição, a partir de um número que consideramos aceitável ao nível das restrições de ambiente, já que muitas vezes, este é considerado fator condicionante deste tipo de operações, *a priori*.

Mais uma vez, podemos comprovar que o simples facto de o sujeito ter de utilizar uma luva, ou qualquer outro dispositivo que facilite a identificação da mão humana é, desde logo, inibidora da ação livre, sistemática e autónoma do indivíduo.

O escopo deste trabalho é limitado ao reconhecimento dos gestos, isto é, deseja-se identificar o gesto estático mostrado pelo ocupante da cadeira, entre um conjunto de gestos pré-definidos pelo sistema, não tendo sido construído, até ao momento, o protótipo que esboçamos.

Em última instância, lançamos o desafio no sentido de despertar atenções e possíveis interesses para a implementação deste tipo de equipamento que será, no nosso entendimento, mais desinibidor da ação do indivíduo, porque não o condiciona a elementos que nem sempre estão ao dispor de qualquer um de nós, mesmo em condições de mobilidade normais.

## 5. Interações Naturais

Interações naturais realizadas por Interfaces Naturais (NUI - Natural User Interfaces) podem ser consideradas como toda e qualquer tecnologia que permita ao utilizador interagir de forma mais intuitiva e direta com as máquinas e com a informação que desta provem.

Esta nova geração tecnológica não só utiliza novas modalidades de entrada de dados, como o multitoque, a voz, a caneta, os dedos, o rastreamento de movimento ou através de marcadores fiduciais mas também nos traz uma nova filosofia de interação com os dispositivos (Blake, Natural User Interfaces in .NET, 2010), isto é, passamos a poder pensar em utilizar uma NUI para interagir com o ambiente através de dispositivos que possam estar embutidos nos nossos eletrodomésticos, na nossa casa, ou mesmo nas nossas roupas.

Como o próprio nome indica, as interfaces naturais têm como objetivo fundamental abrir possibilidades de interações naturais com o utilizador, rejeitando os dispositivos mecânicos de acesso à informação.

A primeira abordagem relacionada com esta temática data de 1991, quando Mark Weiser elaborou uma visão em torno de dois conceitos fundamentais que ainda hoje orientam a pesquisa na área científica de IHM. O primeiro conceito é computação ubíqua (Weiser, The Computer for the 21st Century, 1991) que, cobrindo principalmente o *hardware*, tem como objetivo tornar a IHM tão natural que possa integrar a informática com as ações e comportamentos naturais das pessoas. O segundo conceito prende-se com a “tecnologia calma” (Weiser & Brown, Designing Calm Technology, 1995), onde Weiser meditou sobre uma tecnologia de design de interação

baseada na calma e no conforto, dando primazia ao relacionamento do utilizador com os conteúdos. As interfaces naturais serão, porventura, o passo lógico para atingir esta forma de interação, onde o utilizador interage (de uma forma primitiva) com o computador, através da utilização da fala, gestos, voz, ou até mesmo a movimentação dos olhos.

Para (Monson-Haefel, 2010), o termo NUI é definido como uma Interface Homem-Máquina que modela os aspetos das interações diretas entre as pessoas e o seu ambiente natural.

O mesmo termo é definido em (Blake, Natural User Interfaces in .NET, 2010), como uma interface para o utilizador, concebida para uso de comportamentos humanos naturais, para interagir diretamente com o conteúdo.

Contudo, a maneira mais fácil de entender a NUI é confrontá-la com os outros tipos de interfaces mais tradicionais, tais como as interfaces gráficas (GUI) e as interfaces por linha de comandos (CLI), originado, assim, um novo sentido para o termo: *“Uma Interface Natural é uma interface projetada para reutilizar as habilidades existentes que o utilizador tem para interagir diretamente com o conteúdo”* (Blake, Natural User Interfaces in .NET, 2010).

Segundo Blake (Blake, Multi-touch Development with WPF and Silverlight, 2010), tanto a CLI como a GUI utilizam elementos artificiais nas suas interfaces, ou seja, a CLI utiliza o texto para entrada e saída de informação e a GUI utiliza, por exemplo, o rato para entrada e janelas, menus, ícones como saída de informação.

Todavia, se CLI e GUI se definem ao nível dos dispositivos de entrada, a NUI define-se como o modo como é realizada a interação. Refere-se a um estilo de interface que para o utilizador é tão fácil de controlar que parece invisível.

Na nossa opinião, Blake apresenta uma interface desenhada para que o utilizador possa utilizar comportamentos naturais a um ser humano tais como o toque, os gestos ou a fala, para interagir com os conteúdos, focando-se nas próprias interações, em vez de se focar na interface.

Por conseguinte, numa interface natural, qualquer tecnologia pode ser utilizada, desde que a interação se baseie na reutilização de aptidões não computacionais já existentes no utilizador, inatas ou adquiridas através da prática, vivência e interação com o ambiente que o rodeia (Buxton B. , 2010).

A aplicação destas capacidades simples e intrínsecas ao ser humano, e que podem ser adaptadas a diferentes tarefas, mostra-se como uma grande vantagem das interfaces naturais. A curva de aprendizagem é bastante acentuada, dado que o modo de interagir pode ser alcançado através da observação de outra pessoa apresentando a habilidade uma vez ou duas (Blake, Natural User Interfaces in .NET, 2010). Estas capacidades possuem ainda características e sensações de fluência e naturalidade iguais às das pessoas quando interagem com o mundo real. Estas características baseiam-se em diferentes tecnologias de interação e podemos tipificá-las em interfaces gestuais, tácteis, multitoque, tangíveis, baseadas no reconhecimento de voz, baseadas no rastreamento dos olhos e multimodais.

## 5.1. Tipos de NUI

### 5.1.1. Interfaces multimodais

As interfaces multimodais podem ser pensadas como aquelas de uso intuitivo, não tradicionais e capazes de estabelecer uma comunicação mais real com os seus utilizadores.

(Oviatt, 2003) define os sistemas multimodais de entrada como aqueles que têm a habilidade necessária para processar uma combinação de modalidades - tais como a fala, o toque, a escrita, os gestos, o olhar, os movimentos da cabeça e do corpo - coordenadas com os sistemas de saída, com o objetivo de reconhecer as formas naturais de linguagem e comportamento humano e, desta forma, suportar a IHM de uma forma transparente, flexível, eficiente e expressiva.

(Sarter, 2006), cuja investigação se centraliza nas interfaces multimodais de saída, debate o uso de feedback visual, audível e também táctil para a criação de sistemas cujos proveitos são a sinergia, a redundância e uma maior capacidade de transferência de informação.

Cada uma das modalidades não tradicionais de entrada apresenta um conjunto próprio de impedimentos, como por exemplo o reconhecimento pleno de escrita com a mão livre ou de gestos e expressões faciais e corporais. A incorporação destas modalidades de comunicação em interfaces de entrada é um desafio tecnológico imenso. Os formatos de entrada podem ser tão simples como dois dispositivos apontadores, ou podem incluir reconhecimento de voz e de imagem. A integração perfeita destas múltiplas habilidades permite alcançar uma complementaridade de expressão ao utilizador e, ao mesmo tempo, fornece ao sistema a capacidade de escolher quais as formas de entrada e saída mais apropriadas para cada situação, reconhecendo a informação como uma construção de várias expressões.

Para (Wigdor & Wixon, 2011) e (Clayton, 2011), as NUI são as interfaces capazes de fazer o utilizador sentir e agir tal como uma extensão de seu próprio corpo, diminuindo ou eliminando as dificuldades das IHM.

### 5.1.2. Interfaces tácteis

As interfaces tácteis interagem com os utilizadores através do toque, o qual pode ser construído a partir de diferentes tipos de recetores e permite várias abordagens para sintetizar sensações (toque leve, toque pesado, pressão, dor, vibração, calor e frio). Ambientes virtuais, telemedicina e robótica são algumas das áreas que irão beneficiar do desenvolvimento desta tecnologia. Imaginemos, por exemplo, o apelo que será para os consumidores a capacidade de sentir amostras de tecido antes de comprar *online* uma peça de roupa. Por outro lado, o projeto "The Haptic Radar" (Cassinelli, Zerroug, & Ishikawa, 2009), em desenvolvimento na Universidade de Tóquio, pretende aumentar a consciência espacial, sentindo o ambiente em redor do utilizador, traduzindo-o em vibrações e outros estímulos sensoriais. Como resultado, um utilizador invisual será capaz de evitar a colisão com objetos a poucos metros deles. O potencial deste projeto é tremendo uma vez que poderá servir para evitar colisões em ambientes perigosos para pessoas portadoras de deficiências visuais e para sensibilização de condutores.

### 5.1.3. Interfaces tangíveis

Estas interfaces podem ser definidas como aquelas que compreendem interações realizadas com artefactos físicos, como estímulos para interferir no contexto e representações de informação digital. A capacidade das interfaces tangíveis (TUI) em reconhecer interações aplicadas sobre objetos e empregá-las no contexto de um sistema computacional abre novas possibilidades de tornar as interações com computadores mais próximas dos processos do mundo real. No campo de gestão de emergências, por exemplo, as informações em tempo real da situação num local de uma ocorrência, com a possibilidade de interagir de forma mais natural com o ambiente retratado num mapa, pode representar um melhor atendimento, reduzindo danos e prejuízos. Esta característica é bastante importante em ambientes com processos críticos, onde a rápida compreensão da situação atual pode proporcionar uma melhoria na tomada de decisão (Radicchi, Nunes, & Botega, 2010). Uma aplicação diferente desta tecnologia surge com os Siftables (Merrill, Kalanithi, & Maes, 2007). Estes são blocos que conseguem pensar e que, com base no seu arranjo em relação aos outros, podem tornar-se num qualquer número de aplicações interativas, tais como jogos de matemática, sequenciadores de música ou misturadores de tinta.

### 5.1.4. Interfaces baseadas no reconhecimento da voz

A habilidade de comunicar através da voz com os nossos dispositivos é uma grande promessa da ciência. Conduzir um carro, cozinhar uma refeição, reparar um motor ou realizar uma cirurgia são atividades que poderiam beneficiar de interações através do reconhecimento por voz. Atualmente já existem diversas aplicações nesta área, nomeadamente o fala-para-texto ou a marcação por voz nos telemóveis.

Por outro lado, para facilitar a interação de computadores para pessoas com deficiência (Chathuranga, Samarawickrama, Chandima, Chathuranga, & Abeykoon, 2010) presenteiam-nos uma solução abrangente para o problema onde, como alternativa para o rato, são utilizadas técnicas de VC e tecnologias de reconhecimento de voz. Através dos movimentos do nariz, o utilizador consegue mover o ponteiro do rato e através de comandos por voz pode fazer as outras ações relativas ao manuseamento do rato. O *hardware* é composto por uma câmara de vídeo digital e um microfone.

O reconhecimento da voz está em expansão com promessas, a curto prazo, de aplicações, como por exemplo, um tradutor em tempo real universal.

### 5.1.5. Interfaces baseadas no rastreamento dos olhos

O rastreamento dos olhos (em inglês Eye Tracking) é um sistema que acompanha o movimento dos olhos face ao monitor e regista todo o seu percurso tornando possível analisar padrões de leitura e consulta visual de utilizadores face, por exemplo, a uma página da internet. As áreas da publicidade e do marketing estão bastante atentas à sua evolução, pois baseia-se na agregação de informações sobre onde o olhar do consumidor persiste. Já a medicina de

reabilitação (Lina, Huana, Chana, Yehb, & Chiua, 2004) desenvolveu um sistema de rastreamento ocular para a reabilitação de deficiências no movimento dos olhos.

(Corcoran, Nanu, Petrescu, & Bigioi, 2012) mostram como a detecção da face em tempo real combinado com rastreamento dos olhos podem ser utilizados em ambientes de jogos. As informações recolhidas logram ser utilizadas para melhorar o design das interfaces dos jogos com o intuito de proporcionar uma forma mais inteligente, isto é, adaptável a determinados comportamentos do utilizador, como a alteração do seu humor.

### 5.1.6. Interfaces gestuais

As interfaces baseadas no reconhecimento gestual já se encontram disponíveis no mercado, ao nível dos jogos. O sucesso da Nintendo Wii e da Microsoft Kinect já fazem parte do quotidiano de muitas famílias.

O número de investigações sobre Interface Gestual (IG) tem sofrido uma intensificação, graças à consolidação de técnicas de reconhecimento de imagens, de objetos e de movimentos pertencentes na área da Visão Computacional (VC) (Buxton, Billinghurst, Guiard, Sellen, & Zhai, 1994). Neste tipo de interface, a interação do utilizador com um determinado sistema ocorre por meio do reconhecimento de gestos que podem ser feitos com a cabeça, com os braços, com as mãos e outros membros do corpo humano.

Um bom exemplo desta complementaridade remonta ao ano de 1986, onde podemos encontrar o sistema Very Nervous System (VNS) de David Rokeby (Rokeby, 2010) que oferece um nível sofisticado de interação gestual através de uma câmara de vídeo. O computador deteta a localização da totalidade dos gestos de todos os membros do corpo humano e interpreta-os através do mapeamento destes a sons e outros parâmetros musicais.

No entanto, o intento do nosso trabalho limita-se às interfaces gestuais através do reconhecimento de gestos de mão. Por isso, iremos apresentar com mais detalhe este tipo de interfaces principiando com a definição de gestos de mãos. Convém alertar que, a partir deste ponto, quando nos referimos a gestos, referimo-nos a gestos realizados por mãos.

#### Gestos de mãos

Ao longo do tempo, toda a pesquisa efetuada sobre reconhecimento de gestos tem tido como objetivo primordial a criação de um sistema que possa identificar gestos humanos específicos para poder usá-los na transmissão de informações ou no controlo de dispositivos.

As pessoas comunicam frequentemente através de gestos. São usados para tudo. Existem evidências que apontam que os gestos beneficiam a linguagem falada, sendo parte do processo de uma linguagem (McNeill & Levy, 1982).

Podemos definir gesto por *“movimento do corpo, principalmente da cabeça e dos braços, para exprimir ideias ou sentimentos”* (Porto Editora, 2012). No entanto, como é que os investigadores o definem?

Existem inúmeras definições e discussões sobre a definição de gesto.

Do ponto de vista biológico e sociológico, a sua definição é muito vaga, abrindo oportunidade aos investigadores de classificarem os gestos de forma aberta. *“A noção de gesto é*

*o abraçar de todos os tipos de situações em que um indivíduo emprega movimentos cuja intenção comunicativa é fundamental” (Nespoulous, Perron, & Roch, 1986).*

Os gestos associados à fala são referenciados como gesticulação. Os que operam de forma independente da comunicação são referidos como autônomos. Estes podem ser organizados numa língua de comunicação, como a Língua Gestual Portuguesa, mas também podem representar comandos de movimento.

Segundo Mulder, podemos caracterizar o gesto humano segundo três papéis funcionais: semióticos, ergóticos e epistémicos (Mulder, 1996).

A função semiótica do gesto pretende comunicar informações significativas. A sua estrutura é convencional e resulta de uma experiência cultural partilhada. O gesto “adeus” ou a própria língua gestual ilustram esta função.

A função ergótica do gesto é associada à noção de trabalho, que corresponde à capacidade dos seres humanos em manipular o mundo real para criar artefactos ou para alterar o estado do meio ambiente por “manipulação direta”. O resultado destes gestos poderá ser o molde de barro de uma cerâmica.

A função epistémica do gesto permite que os seres humanos aprendam com o meio ambiente através de experiências tácteis. Quando colocamos a mão num objeto e descobrimos de que material é feito é um exemplo demonstrativo desta função.

Outro método de classificação categoriza os gestos em quatro dicotomias: símbolo-ato, transparência-opacidade, autonomia semiótica-multissemiótica e centrípeta-centrífugo (Nespoulous, Perron, & Roch, 1986), sendo que os últimos dois não revelam pertinência para o estudo em questão porque não se direccionam para o controlo de um objeto nem para a comunicação interpessoal.

A dicotomia símbolo-ato indica a noção de que alguns gestos são puras ações, enquanto outros servem como símbolos. Um gesto de ação ocorre quando uma pessoa conta o seu dinheiro, enquanto um gesto simbólico ocorre quando uma pessoa coloca o polegar para pedir boleia. Esta ideia aponta para a possibilidade de usar gestos representativos de movimentos reais em dispositivos de IHM.

A dicotomia transparência-opacidade refere-se à facilidade com que os outros podem interpretar gestos. A transparência é associada à universalidade, que afirma que alguns gestos padrão têm significados transculturais. Na realidade, os gestos têm significados padrão, mas não têm o mesmo significado em todas as sociedades (Birdwhistell, 1970). Afortunadamente, isto significa que os gestos utilizados para o controlo de dispositivos de IHM podem ser escolhidos livremente.

Outra possibilidade de classificação de gestos padrão utiliza três categorias: mimética, dística e arbitrária (Nespoulous, Perron, & Roch, 1986).

Nos gestos miméticos, os movimentos formam uma representação de um objeto ou recurso. Por exemplo, colocando a mão no queixo e deslocando-a para baixo pode ser usado para representar a barba. Estes gestos são abundantemente utilizados nas representações de língua gestual.

Os díticos são usados para apontar para objetos, são transparentes no seu contexto e representam intenções reais. Apontando com a mão para uma cadeira para pedir permissão para se sentar é um exemplo prático deste tipo de gestos. Tal como os anteriores, são muito úteis na língua gestual.

Gestos arbitrários são aqueles cuja interpretação tem de ser estudada devido à sua opacidade. Embora não sejam vulgares do ponto de vista cultural, uma vez que eles têm de ser estudados e compreendidos, podem ser usados sem qualquer informação verbal complementar. Por conseguinte, estes podem ser criados para utilização específica de um dispositivo de IHM.

Em contraste com a taxonomia gestual apresentada, a interação com os computadores atuais é feita de forma totalmente livre e o paradigma dominante é a manipulação direta. No entanto, podemos conjecturar acerca desta manipulação direta, pois os gestos oferecidos parecem restritos na forma em como se inserem no nosso quotidiano. Essa deficiência torna-se óbvia quando consideramos a proficiência dos seres humanos no uso de gestos no seu dia a dia e então considerar esses gestos para serem utilizados na IHM, eliminando a curva de aprendizagem destes em relação aos gestos de entrada que os computadores podem entender.

#### **Gestos estáticos e dinâmicos**

Os estudos relacionados com sistemas de IHM através de gestos concentram-se, geralmente, nas interfaces de entrada e saída onde os gestos pressupõem a utilização das mãos para instruir uma máquina, e o seu significado depende do sistema que os integra e avalia. Nestes sistemas, os gestos são criados por uma mão em pose ou por um movimento físico em duas ou três dimensões, traduzidos para comandos computacionais ou comandos de movimento. Os gestos simbólicos de parar, iniciar, agarrar e largar são perfeitamente comuns na nossa sociedade e podem ser usados em ambos os casos.

Pavlovic (Pavlovic, Sharma, & Huang, 1997) apresenta uma definição de gesto manual que combina a evolução da postura e da posição no tempo, com a intenção de comunicar ou manipular, diferenciando os gestos em dois tipos: dinâmicos e estáticos.

Os gestos dinâmicos implicam necessariamente evolução temporal (Birk, Moeslund, & Madsen, 1997) e necessitam de movimento para transmitir a mensagem pretendida. O gesto de dizer “adeus” é um exemplo de um gesto dinâmico com a mão aberta que apenas tem significado se for efetuado com movimento.

Por outro lado, os gestos estáticos não necessitam de movimento para transmitir a mensagem. A mesma mão aberta estática, se for efetuada sem movimento, transmite a informação de “parar”.

Em termos de VC, podemos afirmar que um gesto estático é uma configuração da mão em pose, que pode ser representado por uma única imagem, e um gesto dinâmico é um gesto com movimento, representado por uma sequência de imagens.

Assim, com base no uso de gestos humanos, verificamos a necessidade de reconhecer alguns requisitos e tarefas para conceber uma arquitetura ideal para os sistemas de reconhecimento de gestos:

- Escolher gestos que se encaixem de forma útil num ambiente;
- Criar um sistema que possa reconhecer gestos humanos que não sejam perfeitos;

- Criar um sistema que possa usar em simultâneo gestos estáticos e dinâmicos como componentes da transmissão de informação;
- Conceber o gesto o mais rápido possível, mesmo sem o gesto estar concluído;
- Empregar um método de reconhecimento que utilize uma pequena quantidade de tempo de processamento e de memória;
- Produzir um sistema expansível que possa reconhecer outros tipos de gestos;
- Criar um ambiente que permita o uso de gestos para controlo remoto de dispositivos.

### Sistemas de reconhecimento de gestos

A recente importância dada a aplicações baseadas em gestos simbólicos prende-se com a sua potencialidade na utilização em interfaces para IHM. Os gestos podem naturalmente aumentar a interação entre o utilizador e o computador, substituindo dispositivos como o rato, o teclado, o *joystick* ou mesmo botões no controlo de máquinas. No entanto, o principal impulso para o desenvolvimento de interfaces gestuais advém do crescimento de aplicações em ambientes virtuais (Krueger, 1991), como por exemplo a simulação de cirurgias.

Na maioria destas aplicações, os gestos são vistos como manobreadores de objetos virtuais que podem ser gerados por computador, simulando objetos 2D ou 3D ou abstrações de controlo de objetos reais, como por exemplo braços robóticos.

Porém, nem todas as aplicações de gestos são concebidas com o objetivo de manipular objetos. Podem também ser empregues na transmissão de mensagens. A utilização na comunicação através da língua gestual ou em teleconferências facultam este tipo de enjões (Pavlovic, Sharma, & Huang, 1997).

A partir da literatura atual, notamos a existência de numerosos exemplos que utilizam técnicas de VC para desenvolver sistemas de reconhecimento de gestos, aplicados em diferentes áreas.

Notamos que a maioria dos estudos realizados sobre esta temática se centra em algumas áreas de interesse específicas. A figura seguinte resume essas áreas.

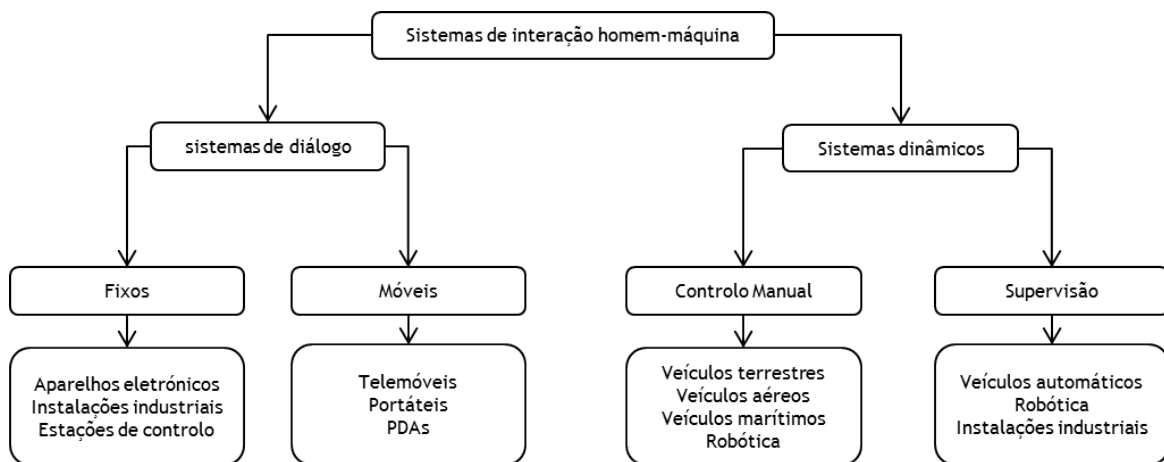


Fig. 4 - Classificação dos sistemas de IHM e áreas de aplicação

Podemos enunciar várias aplicações de interfaces gestuais como o desenvolvimento de técnicas de identificação forense, interação de crianças muito jovens com os computadores,

monitorização de estados emocionais de pacientes ou controlo remoto de televisão (Mitra & Acharya, 2007), mas destacamos os estudos que apresentamos a seguir.

Fourney apresenta-nos *Maestro*, um sistema de reconhecimento de gestos para interação direta com apresentações eletrónicas que permite aos utilizadores utilizarem gestos de navegação para controlar as apresentações (Fourney, Terry, & Mann, 2010). O processo de segmentação é feito através da limiarização da imagem através da cor, mas os utilizadores necessitam de um par de luvas, uma vermelha e outra azul. O reconhecimento do movimento é baseado na correspondência de modelos e utiliza várias características da trajetória das mãos, como por exemplo, a localização do início e fim do gesto, comprimento do caminho percorrido, momento de inércia, entre outros.

Segundo Rocchetti, os sistemas baseados no reconhecimento de gestos podem suportar interação com jogos de computador sem a ajuda de qualquer controlador. As principais novidades na sua abordagem residem na utilização da informação contextual que pode ser inferida a partir da configuração de jogo e da criação de algoritmos capazes de reconhecer as interações de um jogador (Rocchetti, Marfia, & Semeraro, 2011). O algoritmo de reconhecimento das mãos recebe o fluxo de vídeo capturado e analisa, quadro a quadro, a identificação de um conjunto de pontos-chave onde as mãos e antebraços de um jogador foram detetados. O algoritmo de rastreamento de mãos determina a posição exata de ambas as mãos. Finalmente, o algoritmo de reconhecimento de ações verifica, em termos de tempo e trajetória, a ação tomada pelo utilizador.

Rautaray em (Rautaray & Agrawal, 2010) apresenta um trabalho que implementa técnicas de reconhecimento de gestos para desenvolver uma interface para controlar o *software* VLC Player através de gestos. A aplicação desenvolvida é composta por um módulo central que utiliza a Análise de Componentes Principais (PCA) para encontrar os vetores de características dos gestos. O reconhecimento do gesto é feito pelo algoritmo dos K-vizinhos mais próximos (KNN). A análise teórica da sua abordagem apresenta como fazer o reconhecimento num fundo estático. As imagens de treino são feitas cortando o gesto do fundo estático e a deteção do movimento da mão é efetuada através do algoritmo Lucas Kanade.

Em (Zang, et al., 2009), os investigadores propõem uma nova forma de entrada para os jogos virtuais e outras aplicações que utilizam o movimento da mão. Enquanto as formas anteriores de entrada que utilizam as mãos, recorrem apenas ao acelerómetro ou aos sensores eletromagnéticos, este projeto tem duas formas de entrada e combina-as para uma entrada mais precisa e reproduzível para jogos. O acelerómetro é utilizado para determinar a direção e a rotação dos gestos, assim como os seus movimentos. O sensor eletromagnético deteta as correntes elétricas que os músculos fazem enquanto se movem com o corpo, dando assim conhecimento ao sistema da mudança de localização do utilizador. Para testar esta inovação, foi construído um jogo virtual do cubo de Kubrik que permitia manipular o cubo em tempo real no espaço 3D utilizando o sistema de entrada de gestos.

## 6. Desenho da interface: descrição do processo

### 6.1. Enquadramento conceptual

Apresentamos os conceitos fundamentais relativos ao processamento de imagens digitais. Parece-nos que as informações versadas serão necessárias para uma melhor compreensão da abordagem proposta. Iniciaremos com alguns conceitos básicos inerentes a um trabalho desta dimensão, para depois apresentarmos os métodos mais atuais de processamento de imagens digitais, extração de características, reconhecimento de objetos e respetiva classificação.

#### 6.1.1. Conceitos básicos

##### Espaços de cor

A cor é uma percepção visual que é reconhecida pelo homem. Ela deriva do espectro da luz que interage no olho transmitindo informação através do nervo ótico para o sistema nervoso. Este fenómeno ainda não é completamente compreendido, mas sabemos que as cores são determinadas pela natureza da luz refletida nos objetos.

Como tentativa de solucionar esta insuficiência, surgiu a necessidade de especificar as cores de uma forma padrão. Surgem então os modelos ou espaços de cor. Podemos definir espaço de cor como um modelo matemático usado para descrever cada cor a partir de fórmulas representativas das componentes da cor. A representação de qualquer cor pode então ser feita à custa da combinação desses componentes.

Existem vários modelos de cor, com diferentes características e propriedades, adaptadas a problemas distintos como a construção de equipamentos para IHM. Destacamos dois modelos comumente utilizados: o RGB e o HSV.

O modelo RGB é baseado na observação de que pela mistura de vermelho (R), verde (G) e azul (B), em várias proporções, é possível obter uma vasta variedade de cores. Portanto, podemos construir uma imagem a cores usando os componentes vermelho, verde e azul da luz detetada para cada píxel. A cor de cada um destes é, em seguida, determinada pelo peso de cada cor primária. Na televisão a cores, por exemplo, são utilizados três sensores de imagem, um atrás de um filtro ótico vermelho, outro atrás de um filtro ótico verde e outro ainda por trás de um filtro ótico azul. No final, as imagens vermelhas, verdes e azuis são sobrepostas para se visualizar a imagem a cores.

Numa imagem em conformidade com o modelo RGB, o valor de cada píxel pode ser pensado como um vetor de três componentes com os valores de vermelho, verde e azul. Assim, o espaço de cor pode ser exposto de forma que R, G e B sejam considerados como os eixos ortogonais num espaço de cor tridimensional que pode ser representado em termos de um cubo de cores.

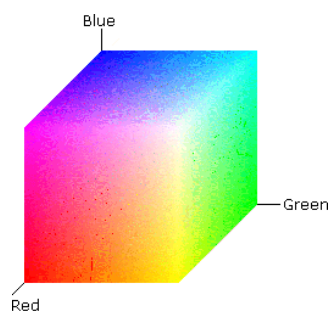


Fig. 5 - Espaço de cor RGB ([http://msdn.microsoft.com/en-us/library/windows/desktop/dd372185\(v=vs.85\).aspx](http://msdn.microsoft.com/en-us/library/windows/desktop/dd372185(v=vs.85).aspx))

Na Fig. 5, podemos encontrar, nos seus vértices, as cores primárias (vermelho, verde e azul), as cores secundárias (magenta, amarelo e ciano) que se obtêm através das cores primárias. O vértice que representa a origem dos eixos não expõe o brilho de qualquer uma das cores primárias e, portanto, corresponde à cor preta. Por sua vez, o vértice oposto será obviamente a cor que representa o brilho máximo, ou seja, o branco. A diagonal que une estes vértices representa a escala de cinza.

Numa imagem digital, os píxeis da imagem terão três valores, cada um para os componentes vermelho, verde e azul. Portanto, uma imagem de cor pode ser representada através da composição de três canais de cores, correspondentes às três cores primárias ( $r$ ,  $g$ ,  $b$ ). O canal vermelho terá os valores do vermelho de todos os píxeis, o canal verde os valores verdes de todos os píxeis e o canal azul os valores azuis de todos os píxeis.

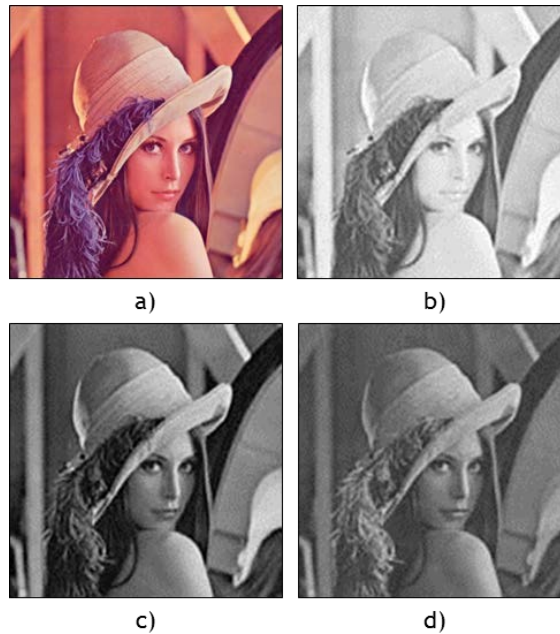


Fig. 6 - Canais de cores: a) Imagem RGB b) Canal RED c) Canal GREEN d) Canal BLUE  
(<http://en.wikipedia.org/wiki/Lenna>)

Genericamente, cada um dos três componentes de cor é representado por 8 bits. Como exemplo, para codificar a cor verde teríamos de colocar um valor de 0 para as componentes vermelho e azul, enquanto a componente verde teria o valor de 255 (0,255,0).

O modelo HSV é bastante utilizado pois a informação de cor contida no matiz (H) e saturação (S) pode ser dissociada do brilho (V). O matiz especifica a cor dominante, a cor pura e pode ser qualquer cor do espectro de cores. A saturação mede a quantidade de branco contido na cor pura variando a clareza da cor. Assim, o matiz e a saturação especificam as informações de cor. O brilho é utilizado para representar a luz do matiz e saturação, isto é, um objeto na luz solar ou na sombra terá valores de brilho diferentes.

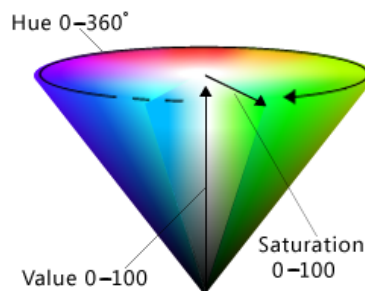


Fig. 7 - Espaço de cores HSV  
(<http://msdn.microsoft.com/en-us/library/windows/desktop/aa511283.aspx>)

A Fig. 7 representa o modelo HSV. O valor que codifica o matiz de cores avança ciclicamente a partir de vermelho a 0°, através do verde e azul, e volta para o vermelho a 360°. Os valores intermédios codificam as cores compostas, tais como o amarelo ou o roxo. O valor da saturação vai desde nenhuma cor (0), com um tom cinzento, até um tom da cor viva (100), sem cinzento. O valor (V) codifica o brilho, preto em 0 e luminosidade total em 100.

Este modelo é bastante útil porque possibilita distinguir uma determinada cor com elevada resolução. Através do processamento de cada componente da cor (separando as componentes da saturação e do brilho), consegue-se diferenciar diferentes objetos, mesmo com uma iluminação não constante.

### Imagem em escala de cinza

Uma imagem em escala ou tons de cinza não tem informação sobre a cor, tem apenas a informação sobre a intensidade de luz. Estas imagens são resultado de um cálculo sobre a intensidade da luz em cada píxel para cada cor. O valor de cada píxel é uma única amostra de um espaço de cores que varia entre o preto como a menor intensidade e o branco como maior intensidade.



Fig. 8 - Imagem convertida para escala de cinza (<http://en.wikipedia.org/wiki/Lenna>)

Para efetuar a conversão de uma imagem no espaço de cor RGB para escala de cinza, será necessário calcular a média dos três canais (r, g, b) e armazenar essa informação num único canal de cor.

### Histogramas

Um histograma de uma imagem é um gráfico da frequência relativa de ocorrência de cada um dos valores de cor permitidos para cada píxel (Solomon & Breckon, 2011). Normalmente representado através de um gráfico de barras, cada um destes valores fornece, para cada nível de cor, o número ou a percentagem de píxéis correspondentes à imagem.

Para uma imagem de escala cinza simples, o histograma pode ser construído através da simples contagem do número de vezes que cada valor de escala cinza (0-255) ocorre dentro da imagem. O eixo horizontal representa os valores de nível cinzento. Cada barra vertical representa o número de vezes que o nível de cinzento correspondente ocorreu na imagem.

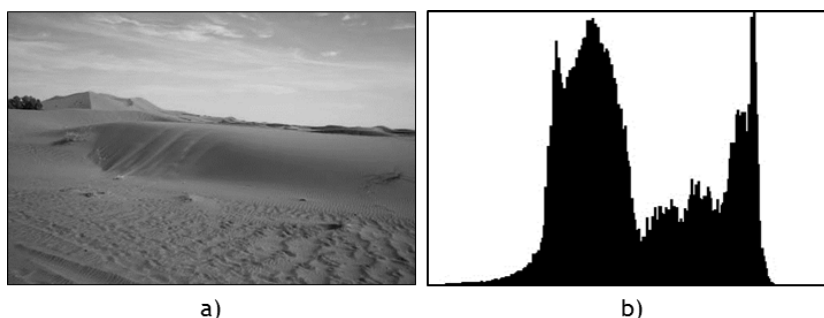


Fig. 9 - Histograma: a) imagem em tons de cinza b) histograma da imagem Na Fig. 9a) podemos verificar a existência de dois tipos principais de objetos que são o céu e a areia do deserto e na Fig. 9b) confirmamos a representação destas regiões através dos dois picos do histograma, indicando que estes níveis de cinza ocorrem com mais frequência na imagem.

Este tipo de histograma pode ajudar na detecção de objetos numa imagem, pois estes tendem a ter níveis de cinza análogos. Assim, através da localização dos picos do histograma e respetivos vales, podemos detetar diferentes objetos.

Convém referir que a identificação de objetos numa imagem não é simplesmente uma questão de localização de picos e vales de um histograma, mas este pode prover outras informações qualitativas e quantitativas, como por exemplo, o nível de cinza mínimo, máximo, médio ou a predominância de pixéis claros ou escuros.

### Filtros

Um filtro digital é um termo genérico para se referir a uma operação matemática aplicada a uma imagem. São utilizados como ferramentas para incrementar a qualidade da imagem ou para fazer sobressair determinados objetos ou características da imagem.

Existem diferentes abordagens para ao nível de aplicação de filtros para PDI, mas podemos dividi-los em duas grandes categorias: filtros passa-baixo e filtros passa-alto. Na primeira categoria são aplicados filtros com valores próximos de zero, para altas frequências; são igualmente utilizados para a redução do ruído, suavizando a imagem. Na segunda categoria são aplicados filtros com valores próximos de zero para as baixas frequências. Este filtro faz com que as transições entre diferentes regiões da imagem se tornem mais nítidas mas, em compensação, enfatizando o ruído que possa existir na imagem.

### Imagem binária

Uma imagem binária é uma imagem digital na qual há apenas dois valores possíveis para cada píxel. As cores mais utilizadas para a representação desses valores são o preto e o branco.



Fig. 10 - Imagem binária (<http://en.wikipedia.org/wiki/Lenna>)

Estas imagens apresentam enormes benefícios em termos de simplicidade e velocidade de tratamento. São muito utilizadas no PDI, onde uma cor representa o objeto e a outra representa o fundo.

### 6.1.2. Processamento Digital de Imagens

O PDI refere-se a todo o tipo de processamento de uma imagem realizado por meio de um computador, e visa a extração de informações a partir de imagens.

Podemos decompor o PDI em três níveis distintos de processamento: baixo, médio e alto.

Os processos de baixo nível envolvem operações primitivas de pré-processamento como, por exemplo, filtragem de ruídos ou melhoria de contraste. Este nível recebe uma imagem digital na sua entrada e devolve outra imagem digital na saída.

Os processos de nível médio circunscrevem-se às tarefas de segmentação de objetos, ou seja, a divisão da imagem em regiões que contenham objetos de interesse, ou a descrição destes em função dos seus atributos, como por exemplo contornos, áreas ou perímetros. Estes processos caracterizam-se por possuir na sua entrada imagens e na sua saída atributos relativos a essas imagens.

Denomina-se por processos de alto nível aos processos que pretendem identificar e atribuir um significado aos objetos reconhecidos, executando assim funções cognitivas normalmente associadas à visão humana.

A figura seguinte ilustra um sistema de PDI reunindo os três níveis descritos.

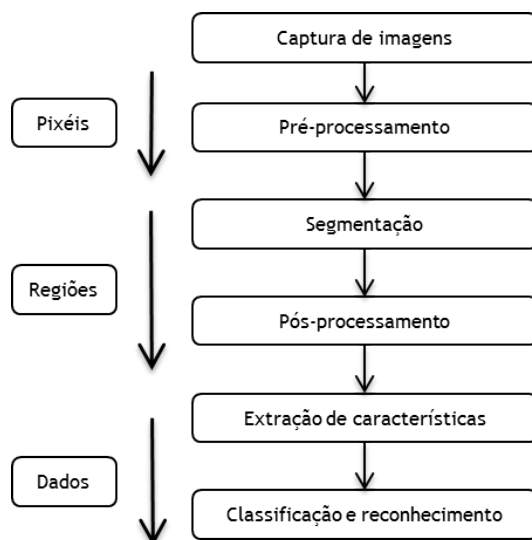


Fig. 11 - Sistema básico de processamento de imagens digitais

### 6.1.3. Aquisição de imagem

Uma imagem pode ser definida como uma função bidimensional  $f(x,y)$ , onde  $x$  e  $y$  são coordenadas espaciais e a amplitude de  $f$  em cada par de coordenadas  $(x,y)$  é chamada de intensidade da imagem no ponto. Caso o par de coordenadas  $(x,y)$  e a amplitude  $f$  sejam valores finitos e discretos, considera-se que a imagem é digital. (Gonzalez & Woods, 2002)

Na sua primeira etapa, aquisição ou captura de imagem, as imagens são convertidas em matrizes bidimensionais de números inteiros, onde o par composto pelos índices das linhas e colunas identificam um determinado ponto na imagem, vulgarmente denominado por píxel. O correspondente valor de cada píxel identifica a sua cor que pode ser apresentada através de um número específico de bits. Este número é chamado de profundidade de cor. A partir desta matriz, podemos implementar diversos algoritmos computacionais.

### 6.1.4. Pré-processamento

As técnicas de pré-processamento têm como objetivo melhorar a qualidade da imagem de forma a realçar características consideradas pertinentes e delimitam-se em duas categorias: domínio espacial e domínio da frequência.

O primeiro baseia-se no processamento das coordenadas de posição utilizando filtros que manipulam a imagem no plano, enquanto o segundo método se apoia em filtros que utilizam o espectro da imagem, após a aplicação da transformada de Fourier na imagem.

Um filtro digital é um termo genérico empregue para nos referimos a uma operação matemática aplicada a uma imagem. É aplicado como ferramentas para incrementar a qualidade da imagem ou para fazer sobressair determinados objetos ou características da imagem.

Existem diferentes abordagens ao nível de aplicação de filtros, mas podemos dividi-los em dois grandes grupos: filtros passa-baixo e filtros passa-alto. Os primeiros são aplicados com valores próximos de zero para altas frequências e são utilizados para a redução do ruído

suavizando a imagem. Os segundos são usados com valores próximos de zero para as baixas frequências. Este filtro faz com que as transições entre diferentes regiões da imagem se tornem mais nítidas mas, em compensação, enfatiza o ruído que possa existir na imagem.

### 6.1.5. Segmentação

A definição para a segmentação de imagens está diretamente relacionada com a área na qual será aplicada. Na área da visão computacional, a segmentação refere-se ao processo de decomposição de uma imagem digital em várias regiões homogêneas, de tal modo que os pixels em cada uma dessas regiões distintas possuam invariavelmente duas características: uniformidade em relação a uma determinada propriedade ou atributo; contraste em relação à região vizinha (Kain, 1989). Estes conjuntos de propriedades da imagem podem incluir níveis de cinza, contraste, valores espectrais ou outras propriedades.

A segmentação de imagens tem basicamente dois objetivos: decompor a imagem em partes para posterior análise; realizar uma mudança de representação (Shapiro & Stockman, 2001). O resultado final de um processo de segmentação é um número finito de regiões homogêneas, que individualizam as diferentes regiões contidas numa imagem. Sendo assim, a imagem seria formada por um conjunto de regiões, separadas pelas suas bordas, de tal forma que cada pixel da imagem pertenceria unicamente a uma determinada região (Acharya & Ray, 2005).

O desenvolvimento de algoritmos de segmentação assenta numa de duas propriedades básicas: descontinuidade e similaridade entre pixels (Gonzalez & Woods, 2002). No primeiro grupo de algoritmos, a divisão da imagem pode ser fundamentada através da distinção de vários tons de cinza ou em mudanças bruscas de níveis de cinza entre regiões. As principais áreas de interesse dentro desta categoria são a detecção de pontos isolados, linhas e bordas numa imagem.

As principais abordagens da segunda categoria são baseadas na observação do interior dos objetos e não nas fronteiras que os delimitam, presumindo que os pixels que compõe um objeto têm propriedades similares e os pixels de objetos distintos têm propriedades distintas. Assim, a separação das regiões é efetuada através da detecção de características similares relativamente a um determinado critério que, por exemplo, poderá ser um nível de cinza.

Importa reconhecer que nenhuma técnica de segmentação é universalmente aplicável para todas as imagens e que nenhuma técnica de segmentação é perfeita. Esta conclusão advém do modo como uma imagem digital é adquirida, uma vez que um pixel pode pertencer parcialmente a duas (ou mais) regiões. A maioria dos métodos existentes pretende atribuir cada pixel a única região, que é uma abordagem suficiente para a maioria das aplicações.

#### **Método baseado na detecção de bordas**

Pontos, linhas e bordas contêm um grande número de informações acerca das diferentes regiões de uma imagem. Apesar de bordas e linhas serem ambas detetadas a partir de uma mudança abrupta do nível de cinza (Gonzalez & Woods, 2002), há uma diferença considerável entre as duas. Uma linha pode ser incorporada dentro de uma única região, ou seja, poderá ser encontrada dentro desta, mesmo quando o seu valor de cinzento seja diferente do valor médio

de cinza da região em que se encontra. Por outro lado, uma borda demarca fundamentalmente duas regiões distintas, implicando que uma aresta seja a fronteira entre as duas regiões (Acharya & Ray, 2005).



Fig. 12 - Detecção de bordas (<http://en.wikipedia.org/wiki/Lenna>)

A detecção de bordas será, essencialmente, uma operação que deteta alterações consideráveis do nível de intensidade numa imagem.

Com base na teoria, existem dois principais métodos que abordam este tipo de segmentação: a técnica do histograma em tons de cinza e a técnica do gradiente de base (Kang, Yang, & Liang, 2009).

A técnica do histograma em tons de cinza depende principalmente da seleção de um limiar  $T$  que permita distinguir o objeto de interesse do fundo de uma imagem. Como é extremamente difícil procurar os níveis de cinza com intensidades máxima e mínima por causa do impacto do ruído, é necessário substituir as curvas do objeto e do fundo com duas curvas Gaussianas (Kang, Yang, & Liang, 2009), cujo cruzamento é o vale do histograma. O limiar  $T$  será o valor do ponto de interseção entre este vale.

O método baseado no gradiente envolve operadores de gradiente que procuram diferenças de cores numa mesma gama de frequência. Quando é encontrado um valor alto de magnitude do gradiente é possível localizar uma transição entre duas regiões diferentes. Os operadores de convolução mais usados são o Sobel, o Canny, o Laplace e o Laplaciano Gaussiano (LOG) (Gonzalez & Woods, 2002).

Os dois os métodos requerem um equilíbrio entre a precisão e a imunidade ao ruído, porque este pode trazer falsas bordas. Assim, estes algoritmos são adequados para imagens simples e sem ruído.

#### Método da limiarização

O intento da limiarização reside na separação de regiões de uma imagem quando esta se pode dividir em duas classes (o fundo e o objeto). Devido ao facto da limiarização produzir uma imagem binária à saída, o processo também é designado por binarização. A forma mais simples de limiarização consiste em determinar ou predefinir um limiar ótimo ( $T$ ), que permita distinguir os objetos numa imagem. Isto é feito através da conversão dos pixels cujos níveis de cinza são maiores ou iguais ao  $T$  em brancos e os outros em pretos, ou vice-versa.

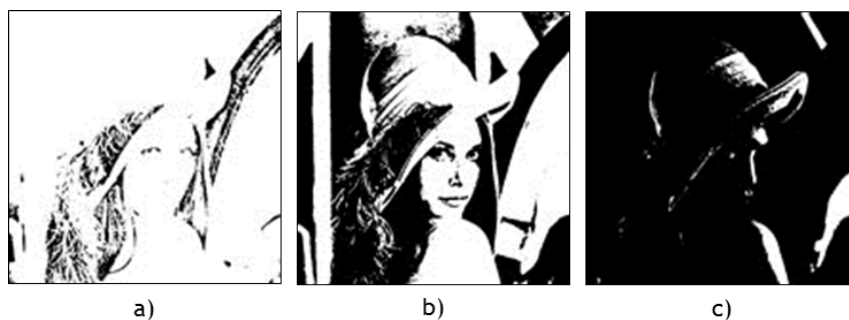


Fig. 13 - a) Limiar baixo b) limiar ótimo c) limiar alto (<http://en.wikipedia.org/wiki/Lenna>)

Como se pode verificar na Fig. 13, a principal dificuldade nesta operação reside na obtenção do valor ótimo do limiar. Este pode ser obtido experimentalmente ou por tentativa e erro. No entanto, cada vez mais se tentam aplicar métodos de detecção automática. Uma técnica bastante usada na detecção automática do limiar recorre ao estudo do histograma de cor da imagem. Através da procura de picos e vales num histograma, consegue-se obter a separação de objetos claros em fundos escuros e objetos escuros em fundos claros. Quando  $T$  é constante, a abordagem é chamada de limiarização global, senão é chamada de limiarização local (Zhang, Qu, & Wang, 2006).

A limiarização global pode falhar sempre que a iluminação de fundo é irregular. Na limiarização local, os limiares múltiplos são usados para compensar a iluminação irregular (Varshney, Rajpal, & Purwar, 2009). A limitação da limiarização advém da desconsideração pelas características espaciais de uma imagem e, devido a isso, é sensível ao ruído (Kang, Yang, & Liang, 2009). A principal vantagem da utilização desta técnica deriva da sua simplicidade matemática e do seu custo computacional, mas necessita de imagens com características específicas como um bom contraste e uma boa nitidez.

#### Método baseado na detecção de regiões

Comparado com os métodos anteriores, este tipo de algoritmos revela-se bastante simples e mais imune ao ruído (Kang, Yang, & Liang, 2009).

Ao contrário dos métodos anteriores que se baseiam em mudanças abruptas da intensidade dos pixels, este método particiona as imagens de acordo com determinados critérios pré-definidos (Gonzalez & Woods, 2002). Apresentamos as técnicas mais comuns para este tipo de algoritmos: crescimento de regiões e divisão de regiões.

A primeira é um procedimento que agrupa os pixels de toda a imagem em sub-regiões ou regiões maiores baseado num critério (por exemplo, tamanho, semelhança) e é processado em 4 passos (Chang & Li, 1994): (1) selecionar um grupo de pixels na imagem original; (2) selecionar um conjunto de critérios de similaridade (níveis de intensidade de cor ou de tons de cinza); (3) fazer crescer regiões anexando os pixels vizinhos que têm propriedades semelhantes; (4) parar com o crescimento, quando não existam mais pixels que preencham o critério para inclusão na região.

A segunda propõe uma divisão da imagem em conjuntos de regiões arbitrárias desconexas para depois fundir essas regiões, numa tentativa de satisfazer as condições de segmentação (Kang, Yang, & Liang, 2009).

### **Métodos baseados em EDP (equação diferencial parcial)**

A segmentação de imagens baseada neste método é realizada maioritariamente pelo modelo de contorno ativo ou cobras. Em 1987, Kass et al. (Jiang, Zhang, & Nie, 2009) desenvolveu este método para encontrar objetos familiares na presença de ruído. A ideia central da cobra é transformar um problema de segmentação num quadro PDE.

Cobras ou contornos ativos são curvas geradas por computador, que se movem dentro da imagem para encontrar os limites dos objetos sob a influência de forças internas e externas.

### **Métodos baseados em Redes Neurais Artificiais**

O método proposto é totalmente diferente de algoritmos de segmentação convencionais. Neste caso, uma imagem é inicialmente mapeada numa rede neuronal, onde cada neurónio representa um píxel (Kang, Yang, & Liang, 2009), convertendo este problema para um processo de minimização de energia. A rede neuronal é treinada com o conjunto de amostras, para depois extrair as bordas da imagem segmentada usando equações dinâmicas que dirigem o estado de cada neurónio para a energia mínima definida pela rede neuronal.

As principais vantagens deste método são: rápido processamento; boa segmentação, mesmo quando os dados são desviantes dos parâmetros normais; insensível ao ruído. No entanto, existem algumas desvantagens que derivam da informação sobre a segmentação a qual deverá ser previamente conhecida, a inicialização poderá influenciar o resultado da segmentação e o tempo de treino é geralmente muito longo (Kang, Yang, & Liang, 2009).

### **Métodos baseados em agrupamentos**

Este método implementa técnicas de aprendizagem automática não supervisionada, onde é preciso identificar um conjunto finito de categorias conhecidas como classes para classificar os píxeis (Dehariya, Shrivastava, & Jain, 2010). O critério de similaridade entre os píxeis é definido, e depois os semelhantes são agrupados. A aglomeração de píxeis em grupos baseia-se no princípio de maximizar a similaridade intraclasse e maximizar a similaridade entre classe. A qualidade de um resultado depende tanto da medida de similaridade usada pelo método como da sua implementação. Estes algoritmos são classificados como agrupamento rígido, agrupamento k-médias, agrupamentos difusos, entre outros.

### **Métodos baseados na subtração de fundo**

A subtração de fundo é um método para detetar objetos em movimento, calculando as diferenças, para cada píxel, entre um quadro que está a ser apresentado e uma imagem de fundo. Com este recurso, é depois aplicado um limiar para separar o objeto pretendido do fundo da imagem, como mostra a Fig. 14. No entanto, no caso de não existir informação prévia da imagem de fundo, esta tarefa torna-se extremamente difícil de concretizar.



Fig. 14 - Subtração de fundo: a) quadro captada b) imagem com fundo subtraído  
(<http://www.box10.com/skill-jam.html>)

Considera-se que as imagens de fundo podem ser de dois tipos: estáticas e dinâmicas. As imagens estáticas representam logicamente imagens que não apresentam movimento, geralmente associadas a interiores e captadas através de câmaras fixas. Por consequência, as imagens dinâmicas, associadas aos exteriores, podem alterar-se devido a diferentes fatores, como por exemplo, alteração na iluminação, movimento da câmara ou do movimento de objetos de fundo, como por exemplo uma carro em movimento.

#### 6.1.6.Extração de características

Definir característica é uma tarefa que não se apresenta claramente circunscrita, pois o que compõe uma característica varia consoante o problema que se pretende resolver. No entanto, podemos arriscar que uma característica será uma parte com interesse numa determinada imagem. Pontos, manchas de cor, cantos, bordas e arestas são regularmente consideradas características de uma imagem e, portanto, no contexto de processamento de imagem, tentamos extraí-las e analisá-las para posterior tratamento.

As características podem ser extraídas com base na sua aparência ou com base na forma.

A extração de características baseadas na aparência pretende descrever a imagem através do seu aspeto. Utiliza a cor, a rugosidade e a textura dos objetos presentes na imagem, recorrendo as intensidades dos pixéis para caracterizar os objetos pretendidos.

Quando nos baseamos na forma, a extração das características focaliza-se nos atributos intrínsecos dos objetos que compõem a imagem, tais como momentos, retângulos envolventes, áreas, contornos, vértices, entre outros.

A Fig. 15 apresenta o resumo dos tipos de características.

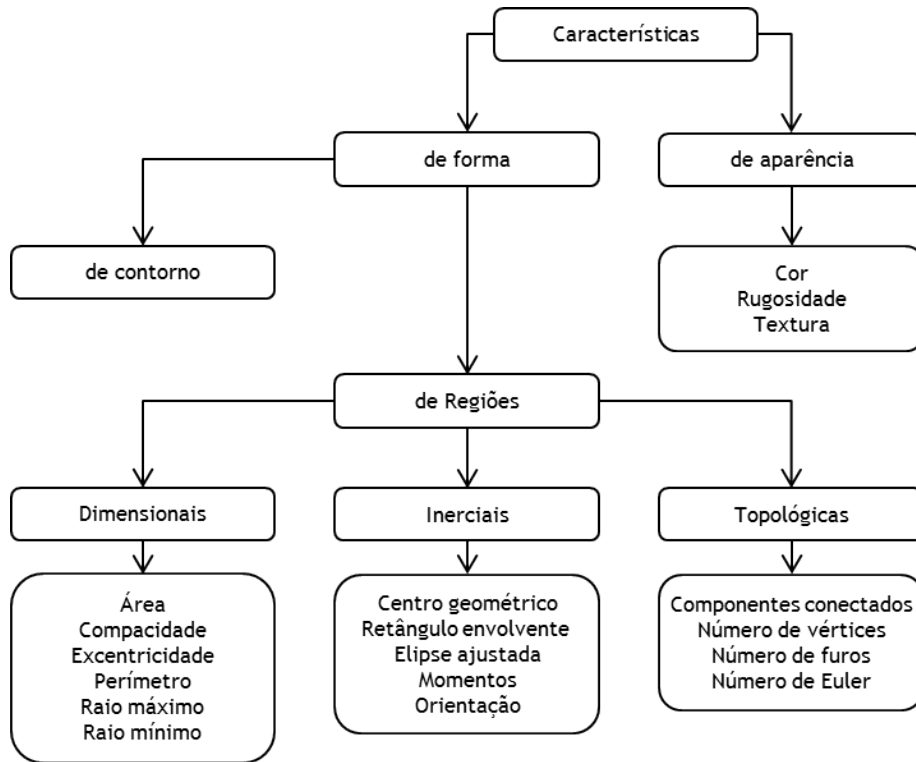


Fig. 15 - Tipos de características

#### Método baseado na aparência

Métodos baseados em aparência, tais como a PCA (Turk & Pentland, 1991), a Análise Discriminante Linear (LDA) (Belhumeur, Hespanha, & Kriegman, 1996) ou Análise de Componentes Independentes (ICA) (Bartlett, Movellan, & Sejnowski, 2002), projetam as imagens num subespaço dimensional menor para simplificar a classificação das imagens com base numa certa distância (ou semelhança) com as imagens pré-conhecidas.

Todavia, apesar dos vários estudos comparativos, os investigadores relatam frequentemente resultados contraditórios em relação ao desempenho dos métodos baseados em aparência. (Liu & Wechsler, 1999) e (Bartlett, Movellan, & Sejnowski, 2002) reclamam que o ICA supera o PCA, enquanto (Baek, Draper, Beveridge, & She, 2002) apresenta um relatório onde alcança melhores resultados com o PCA. Já (Delac, Grgic, & Grgic, 2006) indicam que o desempenho dos métodos baseados em aparência é fortemente dependente da medida de distância empregada e que, com a combinação adequada ao método utilizado, não é possível indicar a superioridade de qualquer uma das três técnicas.

#### Método baseado na forma

Este método consiste, basicamente, na medição de semelhanças entre as diversas formas representadas pelas suas características. Habitualmente, podemos descrever formas utilizando geometria simples desde que estas apresentem muitas dissemelhanças mas, por si só, não devem ser empregues para serem descritores de forma. Geralmente são utilizadas como filtros para eliminar falsos positivos ou para combinar com outros descritores de forma.

Existem inúmeros parâmetros geométricos de forma, que são calculados para explorar a forma da mão, pelo que apenas vamos descrever alguns. Expomos, assim, as características geométricas que foram calculadas no âmbito deste projeto:

**Perímetro:** contorno da mão;

$$perimetro = \sum \text{pixeis do contorno}$$

**Área:** número total de píxeis do objeto mão;

$$area = m_{00}$$

**Proporção:** relação entre largura e altura.

$$proporcao = \frac{comprimento}{altura}$$

**Extensão:** relação da área do contorno e a área do retângulo delimitador do contorno.

$$extensao = \frac{area}{\text{área retângulo delimitador}}$$

**Solidez:** relação da área de contorno para a sua área convexa.

$$solidez = \frac{area}{area\ convexa}$$

**Diâmetro equivalente:** diâmetro do círculo cuja área é igual à área de contorno.

$$diametroEq = \sqrt{4 \times area / \pi}$$

**Análise de Componentes Independentes (ICA):** ângulo de todos os pontos do contorno.

$$orientacao = \frac{1}{2} \text{arc tan} \left( \frac{2 \times m_{11}}{m_{20} - m_{02}} \right)$$

**Eixo principal de inércia:** ângulo de rotação no qual o momento de inércia em relação a um eixo é máximo e, em relação ao outro, é mínimo.

$$eixosInercia_{1,2} = \sqrt{\frac{1}{2} \times (m_{20} + m_{02}) \pm \sqrt{4 \times m_{11}^2 - (m_{20} - m_{02})^2}}$$

**Compacidade:** representa o grau em que uma forma é compacta.

$$compacidade = \frac{1}{4} \pi \left( \frac{perimetro^2}{area} \right)$$

**Excentricidade:** define a relação entre o seu raio máximo e o raio mínimo de um contorno.

$$excentricidade = \frac{(m_{20} + m_{02})^2 + 4m_{11}}{area}$$

**Circularidade:** mede a semelhança da forma de um objeto em relação à forma de um círculo;

$$circularidade = \frac{perímetro^2}{4\pi \times área}$$

**Retangularidade:** mede a semelhança da forma de um objeto em relação à forma de um retângulo;

$$retangularidade = \frac{área}{altura \times comprimento}$$

No que diz respeito às características estatísticas, utilizamos os 7 Momentos Hu. Os momentos de uma imagem possibilitam calcular propriedades geométricas de objetos presentes numa imagem como a posição, o tamanho ou a orientação. Estas propriedades são relevantes para a implementação de sistemas de reconhecimento de gestos e objetos e para extração de características para posterior reconhecimento de padrões, onde há a necessidade de encontrar descritores de objetos que sejam invariantes em relação à translação, rotação e escala (Hu, 1962) apresenta 7 momentos invariantes.

$$\begin{aligned}\phi_1 &= m_{20} + m_{02} \\ \phi_2 &= (m_{20} - m_{02})^2 + 4 \times m_{11}^2 \\ \phi_3 &= (m_{30} - m_{12})^2 + (3 \times m_{21} - m_{03})^2 \\ \phi_4 &= (m_{30} - m_{12})^2 + (m_{21} + m_{03})^2 \\ \phi_5 &= (m_{30} - 3m_{12})(m_{30} + m_{12})[(m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2] \\ &\quad + (3m_{21} - m_{03})(m_{21} + m_{03})[3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2] \\ \phi_6 &= (m_{20} - m_{02})[(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2] + 4m_{11}(m_{30} + m_{21})(m_{21} + m_{03}) \\ \phi_7 &= (3m_{21} - 3m_{03})(m_{30} + m_{12})[(m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2] \\ &\quad - (m_{30} - 3m_{21})(m_{21} + m_{03})[3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2]\end{aligned}$$

Nos últimos anos, inúmeras técnicas têm sido propostas como modo de descrição da forma e medição de semelhanças. No entanto, existem três abordagens diferentes no que toca à classificação da forma (Mingqiang, Idiyo, & Joseph, 2008):

O método sustentado nas bordas ou regiões é proposto pela MPEG-7. É o mais comum e baseia-se na utilização dos pontos das bordas das formas.

Os métodos baseados no domínio do espaço fazem corresponder as formas com recurso aos pontos, enquanto as técnicas baseadas no domínio das características utilizam a correspondência das formas com recurso a vetores.

Por último, os métodos de preservação de informação e não-preservação de informação permitem a reconstrução clara de uma forma através do seu descritor.

### 6.1.7.Descritores

A extração dos atributos relatados na alínea anterior, quaisquer que sejam os escolhidos, permite obter um conjunto de valores que passaremos a denominar de vetor de características,

também chamados de descritores que permitem determinar com precisão cada objeto. Estes descritores devem ser representados por uma estrutura de dados adequada ao algoritmo de reconhecimento, visando eliminar a redundância e aumentar a exclusividade dos parâmetros escolhidos.

Suponhamos que os descritores usados para descrever uma moeda sejam as coordenadas normalizadas  $x$  e  $y$  do centro e a medida do raio (distância do centro a um ponto qualquer da circunferência). Neste caso, um vetor de três elementos (coordenada  $x$ , coordenada  $y$  e raio) é uma estrutura de dados adequada para armazenar estas informações sobre cada uma das diferentes moedas em circulação.

### 6.1.8. Classificação e reconhecimento de padrões

Hoje em dia, a classificação de imagens é um domínio com bastante maturidade. Subsistem distintas abordagens para descobrir relações entre imagens ou segmentos de imagem.

As aplicações desenvolvidas para o reconhecimento de padrões, também denominadas classificadores, pertencem à subárea da Inteligência Artificial que se dedica a desenvolver técnicas automatizadas capazes de classificar informações baseadas num conhecimento *a priori* ou em informações estatísticas extraídas dos padrões.

Iremos designar por padrão uma entidade à qual se pode atribuir um nome. Estes podem ser representados como vetores multidimensionais, onde cada dimensão é um único atributo. Assim, estes atributos (ou características) dos padrões são categorizados em quantitativas que representam valores contínuos, discretos ou intervalos de valores e qualitativas que assumem valores nominais (como a cor) e ordinais (como o frio e o quente). Posto isto, podemos definir classificação como um ato de atribuição de classes a objetos (Costa & Jr., 2001). Para tal, a criação de classificadores envolve sempre a solicitação de um modelo e o uso de padrões de treino para aprender ou estimar os parâmetros desconhecidos desse modelo.

Podemos diferenciar dois tipos de aprendizagem principais: a classificação supervisionada e a classificação não supervisionada.

A classificação supervisionada tem como objetivo encontrar conceitos a partir de amostras pré-fornecidas ao sistema e que estão pré-classificadas, isto é, exemplos que já estão catalogados nas suas respetivas classes. O classificador é treinado para replicar a decisão correta para novas amostras. Assim, torna-se possível classificar novos conjuntos de dados não rotulados com base na classificação já existente. Se essas classes possuírem valores discretos, o problema é categorizado como classificação. Pelo contrário, se as classes possuírem valores contínuos, o problema é categorizado como regressão. O maior obstáculo deste método prende-se com a construção de um conjunto de amostras satisfatórias por cada classe que permita a classificação de todos os dados. Neste tipo de aprendizagem destacamos as Máquinas de Suporte Vetorial (SVM), as Redes Neurais Artificiais (ANN), a Aprendizagem Bayesiana (NB), o K Vizinhos Mais Próximos (KNN), as Florestas Aleatórias (RF) e as Árvores de Decisão (DT).

A classificação não-supervisionada pressupõe que o conjunto de amostras não está catalogado nem os padrões de treino se encontram classificados. O sistema tenta classificar os conjuntos de dados agrupando os semelhantes nas classes com maior correspondência. Neste

caso, a informação disponível é menor, pelo que a classificação se torna menos precisa. Todavia, esta poderá ser a única solução para problemas sem informação precisa acerca das classes que reúnem os dados. Aqui destacamos o PCA, o ICA, o K-Médias, e o Modelo Oculto de Markov (HMM).

## 6.2. Hardware Utilizado

Podemos caracterizar o *hardware* utilizado para o desenvolvimento deste projeto como *hardware* atual a partir do ano de 2010, não sendo necessário nenhum equipamento específico especial para que o *software* funcione corretamente.

Apontamos, porém, uma situação - a câmara digital. A escolha da câmara digital deverá fornecer um sinal estável, sem grandes variações de luminosidade.

Este facto tem uma enorme importância pois pode representar a eliminação de vários problemas.

A câmara de captura de vídeo utilizada não fornecia um sinal estável ao nível da luminosidade e, em consequência, fazia com que o *software* exibisse, por vezes, inconformidades no reconhecimento e segmentação de cores.

O computador utilizado neste sistema possui as características apresentadas na Tab. 2.

Processador	Intel Core 2 Quad 9400 @2.6GHz
Motherboard	ASUSTek P5Q-PRO
Memória	DDR2 4 GB
Disco rígido	Western Digital 1TB 7200rpm

Tab. 2 - Características do computador utilizado

Utilizámos uma câmara de captura de vídeo da marca Logitech para a aquisição de imagens com as características expostas na Tab. 3.

Marca	Logitech
Modelo	Logitech Webcam C100
Tipo de sensor	2.3
Captura de vídeo	640 X 480 pixéis
Resolução de vídeo	30 quadros por segundo
Interface	USB 2.0

Versão do driver	13.31.1044.0
------------------	--------------

Tab. 3 - Características da câmara web utilizada

### 6.3. Software utilizado

O desenvolvimento do sistema proposto presume a utilização de plataformas de desenvolvimento. No que diz respeito à área da Visão Computacional, podemos encontrar algumas bibliotecas dedicadas, destacando-se a aplicação Matlab e a biblioteca OpenCV. O Matlab é um *software* vocacionado sobretudo para o cálculo numérico, cálculo de matrizes e processamento de sinais e o OpenCV é uma biblioteca multiplataforma de código fonte livre e gratuito, constituída por um conjunto de funções nas áreas do Processamento de Imagem e Visão Computacional.

Em relação ao MatLab, podemos aludir ao seu ambiente gráfico e à exposição dos problemas do ponto de vista matemático. A sua linguagem de script utiliza comandos muito parecidos à forma como escrevemos expressões algébricas e não requerem a preocupação, por parte do programador, em incluir bibliotecas, declarar variáveis ou gerir a memória. No entanto, esta facilidade no seu uso deriva do facto do Matlab ser construído na linguagem Java, sendo esta construída sobre a linguagem C, o que o torna lento ao nível da execução de código, dificilmente utilizado para processamento de imagens em tempo real.

A biblioteca OpenCV é rápida e eficiente em todas as suas versões (Linux, Windows, OS X e Android). Apresenta estruturas de dados básicos ideais para a realização de operações com matrizes. Utiliza código altamente otimizado para processamento de imagens, mas o seu tratamento de erros deixa muito a desejar. A documentação não é aconselhável a iniciantes mas facilmente encontramos exemplos da sua utilização. Em baixo, apresentamos uma tabela com o resumo dos seus principais atributos:

ATRIBUTOS	MATLAB	OPENCV
Facilidade de utilização	X	
Rapidez		X
Recursos necessários		X
Custo financeiro		X
Ambiente de desenvolvimento	X	
Gestão de memória	X	
Portabilidade		X

Desenvolvimento de competências de programação		X
Depuração de erros	X	
Ajuda e exemplos de código	X	X

Tab. 4 - MatLab vs. OpenCV

Analisando a Tab. 4, realçamos a necessidade de refletir sobre o tipo de atributos em que estes ambientes de desenvolvimento apresentam resultados satisfatórios. Embora o Matlab exiba bons resultados nos parâmetros relacionados com a usabilidade, apresentando uma boa curva de aprendizagem baseada no seu ambiente de desenvolvimento, numa gestão da memória eficaz e numa excelente secção de ajuda, a execução do código é lenta e o seu preço também não constitui uma vantagem.

Por outro lado, a biblioteca OpenCV apresenta melhores resultados no que diz respeito aos atributos relacionadas com a eficiência do produto final. Na realidade, o OpenCV exige um esforço suplementar por parte do programador no que se refere à gestão de memória e depuração de erros, mas apresenta enormes vantagens relativamente à racionalização de recursos e à sua rapidez. Aliás, a sua enorme velocidade relativamente ao MatLab apresenta-se como um fator preponderante.

O Matlab é um *software* genérico, projetado para diversas utilizações, enquanto o OpenCV foi construído unicamente para o processamento de imagens, onde cada estrutura de dados foi concebida para o programador tendo em conta a área da Visão Computacional.

Assim, podemos constatar que o Matlab apresenta uma melhor curva de aprendizagem, enquanto que o OpenCV se apresenta mais difícil mas mais eficiente.

Concluindo, o OpenCV será o mais adequado para a elaboração de um sistema eficaz utilizando o mínimo de recursos possível bem como para o desenvolvimento de competências de programação.

Resumimos na Tab. 5 o *software* utilizado.

Biblioteca De Visão Computacional	OpenCV (Open Source Computer Vision Library) 2.4.2
Ambiente de desenvolvimento	Microsoft Visual Studio C++ 2010 Express Edition
Sistema Operativo	Windows 7 Professional 32 bits

Tab. 5 - Características do *software* utilizado

#### 6.4. Base de Dados de Gestos

As escolhas das imagens representativas dos gestos foram selecionadas aleatoriamente. Pretendemos que várias produções do mesmo gesto tenham variações intra-gesto média, mas que variem significativamente entre as diferentes classes de gestos (variância inter-gesto alta). Não é

necessário que os gestos da mesma classe sejam perfeitamente iguais em cada elemento do vetor de características, mas deve haver pelo menos um elemento no vetor que seja díspar para podermos simular a naturalidade dos gestos efetuados pelas diferentes pessoas que possam utilizar o nosso sistema.

A criação da BD de gestos seguiu os seguintes procedimentos:

1. Foram gravados, com uma câmara de captura de vídeo, diferentes imagens com várias posições (poses) de mão e escolhemos 6 poses como gestos padrão, ou seja, gestos a serem reconhecidos;
2. Executámos a aplicação e exportámos cada quadro;
3. As imagens exportadas foram selecionadas aleatoriamente;
4. Para cada quadro, foram também captados os respetivos vetores de características;
5. O vetor de características foi armazenado numa BD.

## 6.5. Visão Geral do sistema

O método proposto obedece a uma sequência lógica de tarefas distintas e complementares ao nível do processamento digital de imagens (Gonzalez & Woods, 2002). A Fig. 16 mostra o diagrama de blocos do sistema.

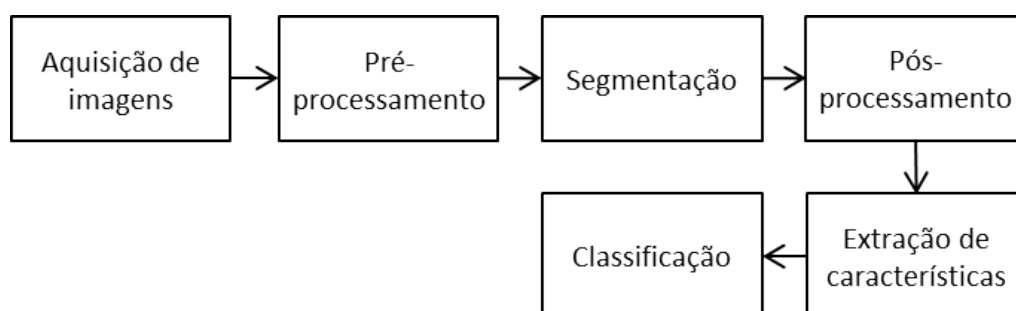


Fig. 16 - Etapas fundamentais do trabalho proposto

O diagrama apresentado representa uma abordagem completa aos problemas de visão computacional, baseado nas etapas de captura da imagem de vídeo, segmentação da imagem, extração de características, reconhecimento de padrões treinados e interpretação de resultados.

## 6.6. Arquitetura do Sistema

### 6.6.1. Captura da imagem

A câmara de captura de vídeo está ligada ao computador através da interface USB e posicionada através de um suporte em frente da mão humana que executará os gestos. A distância que separa a mão humana da câmara deverá ser de modo a que a mão humana ocupe aproximadamente a totalidade do quadro marcador de gestos representado na imagem capturada.



Fig. 17 - Quadro capturado com a mão bem posicionada

Cada quadro capturado pela câmara é tratado como uma imagem estática e, para tal, precisa de ser capturada do vídeo original. Ela tem uma resolução de 640x480 pixéis reproduzida no espaço de cor RGB, e cada canal de cor é representado por 8 bits.

### 6.6.2. Pré-processamento

Nesta fase, pretendemos seleccionar apenas a informação necessária para que a tarefa da segmentação possa produzir bons resultados.

Primeiramente, pretendemos delimitar a área que nos apresenta a mão humana e que contém toda a informação que pretendemos trabalhar. A parte restante da imagem é descartada pois não oferece dados relevantes. Com isto, reduzimos significativamente o número de pixéis que iremos tratar.



Fig. 18 - Região de Interesse (ROI)

Depois, é aplicado um filtro de suavização na imagem capturada. Esta é uma operação simples de processamento de imagens e frequentemente utilizada com o intuito de eliminar alguns ruídos que possam estar presentes na imagem.

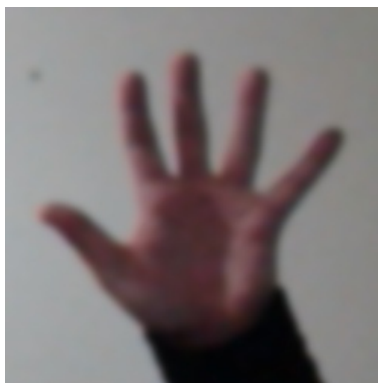


Fig. 19 - ROI suavizada

O filtro utilizado é do tipo passa-baixo, Gaussiano com máscara de 5x5. Assim conseguimos suavizar a imagem de uma forma ponderada e simétrica.

### 6.6.3. Segmentação

A segmentação é a parte principal de todo o processo e pressupõe uma sequência de tarefas lógicas. O seu objetivo é a separação da mão humana em relação ao fundo da imagem. Todos os quadros recolhidos pela câmara de captura de vídeo passam por este processo. Esquematicamente, as suas etapas podem ser observadas na Fig. 20.

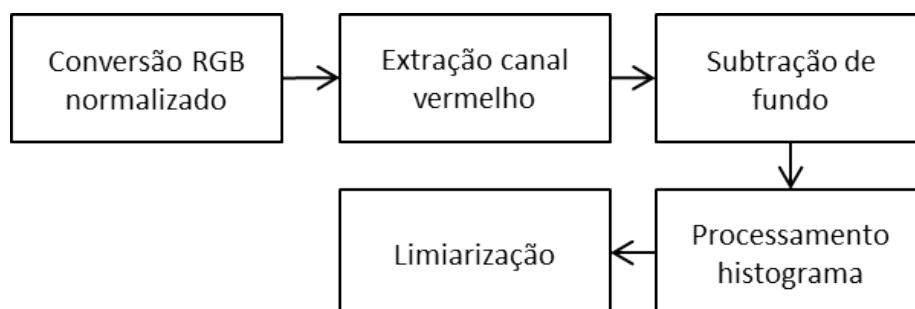


Fig. 20 - Processo de segmentação de mão humana

O primeiro passo do processo passa por classificar cada píxel como pertencendo ou não a um determinado objeto. O espaço de cores RGB apresenta problemas quando ocorrem mudanças na iluminação, pois as suas cores variam substancialmente sempre que tal ocorre. Esta dificuldade afeta gravemente a classificação pretendida. Assim, para superar este obstáculo, podemos usar espaços de cor diferentes sendo que alguns deles permitem separar a informação da luminosidade dos píxeis.

Os espaços de cores que apresentam as características pretendidas são: RGB normalizado, HSV, HSL e YCbCr. Cada um destes espaços de cor contém três componentes, sendo um deles o parâmetro luminosidade, excepto para o RGB Normalizado em que um dos componentes é redundante. Sendo assim, a luminosidade e o parâmetro redundante não se apresentam necessários para a filtragem de cor, sendo então ignorados. A cor é obtida utilizando os outros dois componentes.

Podemos verificar que, no RGB Normalizado, os valores de uma cor permanecem os mesmos, inclusivamente quando existem alterações na iluminação.

As cores em RGB podem ser transformadas para RGB normalizado usando as seguintes fórmulas:

$$r = \frac{R}{R + G + B}$$
$$g = \frac{G}{R + G + B}$$
$$b = \frac{B}{R + G + B}$$

Podemos facilmente reparar que,

$$r + g + b = 1$$

portanto, qualquer um dos três valores pode ser descartado, uma vez que é redundante.

O RGB normalizado é reconhecido por alguns autores como vantajoso para a segmentação de cores (Aryuanto & Limpraptono, 2009) pois elimina o efeito das variações de intensidade de cor nos pixéis.

Depois de esclarecer a nossa escolha relativamente ao espaço de cor, podemos agora expor todo o processo de segmentação.

O procedimento inicia-se com a conversão dos quadros capturados pela câmara de captura de vídeo para o espaço de cor RGB normalizado.



Fig. 21 - a) imagem RGB normalizado

Sabendo que a maioria da cor da pele humana tende a agrupar-se no canal vermelho, extraímos esse canal (Fig. 21).



Fig. 22 - canal vermelho extraído

Depois, subtraímos o fundo à imagem do canal vermelho normalizado. A subtração de fundo permite-nos obter uma imagem em níveis de cinza com alterações de intensidade de cor entre o fundo e a mão, como mostra a Fig. 22.

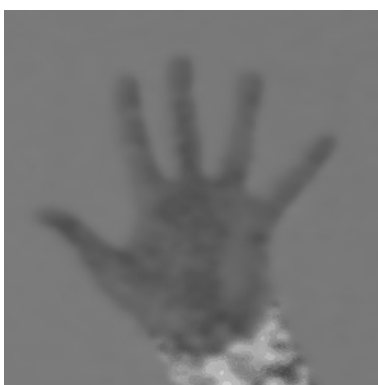


Fig. 23 - a) canal vermelho extraído com subtração de fundo

Seguidamente concebemos o histograma da imagem.

Para construir o histograma de uma imagem, é necessário inicializar com zero todos os elementos de um vetor de  $L$  elementos, onde  $L$  é o número de tons de cinza possíveis. Depois, percorre-se a imagem, píxel a píxel, incrementando a posição do vetor cujo índice corresponde ao tom de cinza do píxel inspecionado. Após toda a imagem ter sido percorrida, cada elemento do vetor conterá o número de píxeis cujo tom de cinza equivale ao índice do elemento.

A segmentação de imagens baseada em histogramas é uma das técnicas mais simples e utilizadas. O histograma é utilizado para selecionar os níveis de cinza em que se dá a separação de píxeis em regiões. Podemos considerar que numa imagem simples apenas existem duas entidades: o fundo e o objeto. O fundo é geralmente um nível de cinzento e ocupa a maior parte da imagem. O seu nível de cinza é um grande pico no histograma. O objeto da imagem é outro nível de cinza, que será outro pico, com algum espaçamento em relação ao primeiro.

Através do histograma, calculamos a valor ótimo da limiarização utilizando o Método do Vale.

Este método propõe que o vale mais profundo entre os dois picos deverá ser definido como limiar. Assim, conseguimos uma separação aceitável entre o fundo e o objeto.

A Fig. 24 mostra o histograma gerado e o limiar ótimo identificado.

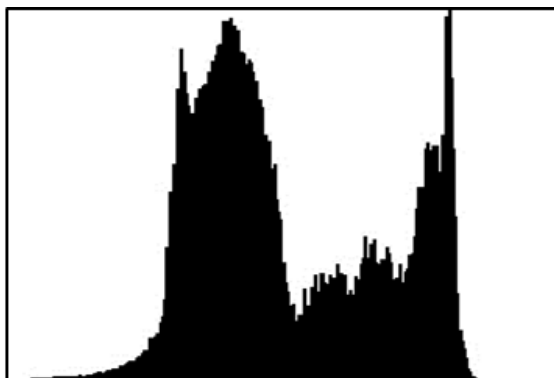


Fig. 24 - Histograma da imagem com o valor ótimo de limiar identificado

O valor encontrado no Método do Vale será então utilizado como critério de seleção para a limiarização. Todos os píxeis da imagem são comparados com esse critério e são alterados. Os píxeis cujo valor seja maior que o critério passarão a ter cor preta, e os restantes serão igualados à cor branca. Desta forma destaca-se uma região da imagem. Esta operação produzirá uma imagem binarizada (Fig. 25).



Fig. 25 - Imagem binarizada

#### 6.6.4. Pós-processamento

Nesta fase, pretendemos refinar a informação atrás recolhida através da aplicação de filtros morfológicos à imagem binarizada. O seu intuito é o de eliminar pequenos buracos presentes nas bordas e no interior da imagem segmentada.

Os filtros utilizados foram: o filtro de dilatação e o filtro de erosão. Os dois filtros operam nas bordas internas e externas dos objetos. A quantidade e a forma como os objetos se dilatam ou encolhem obedecem à eleição de uma máscara. A dilatação faz com que os objetos se dilatam ou aumentem de tamanho, enquanto a erosão faz com que eles encolham.

Neste caso, foram aplicados filtros morfológicos utilizando uma máscara de vizinhança 8. A sequência de aplicação dos filtros foi primeiro a dilatação e depois, na imagem dilatada, a erosão. Com a dilatação, procura-se, em primeiro lugar, expandir o objeto através das suas bordas internas e externas, para preencher possíveis buracos e expandir as bordas uniformemente. Depois, aplicando o filtro morfológico erosão na imagem dilatada, procura-se regressar ao objeto original; apenas as bordas externas serão afetadas, porque os buracos já

totalmente preenchidos na etapa de dilatação serão mantidos. A aplicação destes filtros nessa sequência é conhecida também como fechamento.

A Fig. 26 exibe os resultados obtidos após a aplicação desta sequência de filtros morfológicos.



Fig. 26 - Imagem após a aplicação dos filtros morfológicos

#### 6.6.5. Detecção de feições

A etapa anterior facultava uma imagem onde podemos encontrar os pixels da borda que separam os diferentes segmentos de uma imagem, neste caso a mão do fundo da imagem, mas ainda não temos informação sobre as bordas como entidades em si. O próximo passo é ser capaz de montar os pixels da borda em contornos para obter todas as estruturas a partir das quais vamos calcular as características a extrair.

Podemos definir contorno como uma curva (sequência de pontos) que define a borda de um objeto numa imagem e é representado por um vetor de pontos.

Esta fase irá detetar todos os contornos que circundam a mancha da mão segmentada e tem como principal objetivo localizar todos os pixels que pertencem ao contorno da mão.

A extração do contorno é uma etapa importante, porque é através do vetor de contornos que iremos, na etapa seguinte, extrair as características geométricas e os momentos estatísticos.

No nosso trabalho, apenas o maior contorno exterior será obtido para posterior análise.

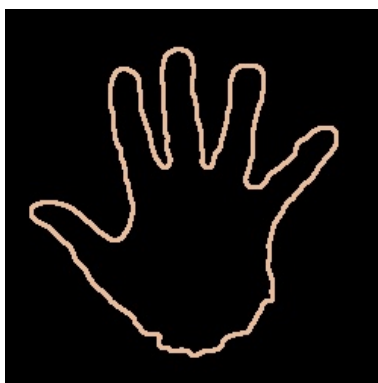


Fig. 27 - Contornos da mão

O resultado é uma imagem em que apenas podemos ver o contorno da mão.

Através do contorno podemos calcular várias estruturas, como o invólucro convexo, as arestas do contorno, retângulos, elipses e círculos envolventes.

Por definição, o invólucro convexo é o menor polígono que, dada uma sequência de pontos (neste caso, o contorno), consegue abranger no seu interior todos os pontos dessa sequência com o menor número de arestas.



Fig. 28 - Invólucro convexo da mão

Depois, para descobrir os dedos da mão, será necessário encontrar as maiores saliências convexas do contorno. Para tal, serão eliminados todos os pontos que, à partida, serão redundantes, ou seja, os pontos que estão rodeados por bastantes pontos que pertençam ao invólucro convexo.

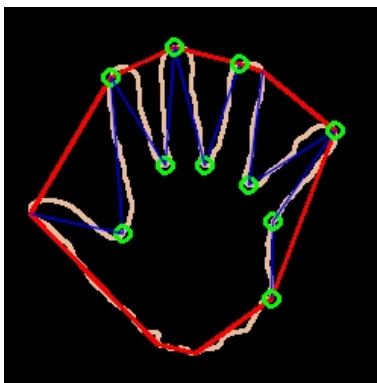


Fig. 29 - Saliências convexas do contorno

Por fim, definimos o retângulo e o retângulo normalizado que incluem todos os pontos que compõem o interior dos limites do contorno da mão.

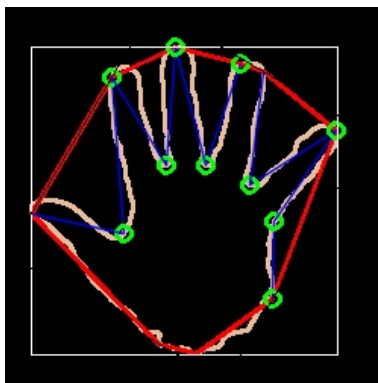


Fig. 30 - Retângulo envolvente

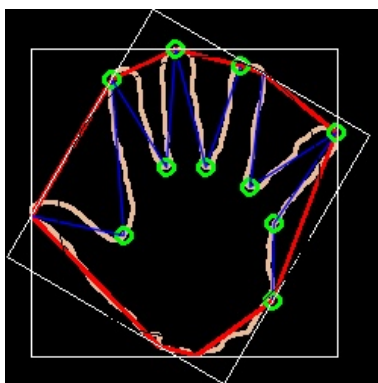


Fig. 31 - Retângulo envolvente normalizado

#### 6.6.6. Extração das características

A extração de características e propriedades representativas dos objetos de interesse é crucial na identificação e diferenciação de gestos estáticos. Qualquer que seja o objeto a ser identificado, é presumível que este possua algumas características que o definam de uma forma singular em relação ao objeto do fundo.

Assim, tendo por base as estruturas construídas na fase anterior, podemos agora proceder à extração de várias características representativas do contorno e do invólucro convexo.

Todos estes recursos serão utilizados para classificar os gestos e reconhecer o gesto do utilizador.

Para seleccionar o conjunto de características mais adequado foi utilizado este conjunto de passos:

Passo 1: Iniciar com uma seleção de características vazia e uma precisão de reconhecimento de zero.

Passo 2: Para cada característica que não esteja na seleção de características, deverá medir-se a sua precisão de reconhecimento. Adicionar a característica que obteve maior precisão de reconhecimento. Se não existir nenhuma característica que aumente a precisão de reconhecimento, parar.

Passo 3: Se o custo computacional ou a complexidade do sistema exceder o limite máximo aceitável, remover a característica adicionada a partir da seleção e parar.

Passo 4: Se a precisão do reconhecimento for suficiente, ou todas as características já se encontrarem incluídas na seleção atual, devemos parar.

Passo 5: Ir para passo 2.

Após a execução deste algoritmo verificámos que as características mais discriminatórias dos seis gestos escolhidos são:

- Compacidade do contorno (C1);
- Proporção do retângulo envolvente (C2);
- Área do retângulo envolvente (C3);
- Momentos Hu 4 do contorno (C4);
- Tamanho do contorno (C5);
- Eixo principal de inércia (C6);
- Momentos Hu 7 do contorno (C7);
- Excentricidade do contorno (C8);
- Ângulo do retângulo envolvente normalizado (C9);
- Tamanho do invólucro convexo (C10);

Alguns problemas durante o reconhecimento de gestos podem ocorrer se houver movimentos elevados de rotação, porque a forma e as proporções das características que representam o gesto podem sofrer alterações.

Convém realçar que estas características são fáceis de extrair, são independentes e são suficientes para distinguir a quantidade de gestos por nós escolhida. Ainda assim, estas não serão as mais apropriadas a todos os tipos de gestos.

### **6.6.7. Reconhecimento de padrões**

Para o reconhecimento de padrões, escolhemos a técnica de aprendizagem automática denominada por Máquinas de Suporte Vetorial (SVM) que vem sendo reconhecida nos últimos anos. Esta técnica tem sido utilizada em distintas tarefas de reconhecimento de padrões, sempre com resultados superiores aos alcançados por técnicas similares em várias aplicações. O seu uso advém de algumas das suas principais características (Smola, 1999):

- Boa capacidade de generalização;
- Robustez em grandes dimensões;
- Base teórica bem definida dentro da Matemática e Estatística.

As SVM são classificadores lineares que maximizam a distância geométrica entre o subespaço vetorial das classes. Esta técnica tem como base a construção de um hiperplano, como superfície de decisão, de modo a maximizar a margem de separação entre as classes.

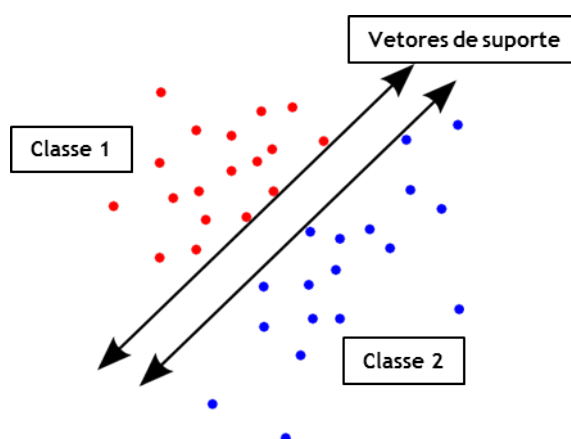


Fig. 32 - Separação de classes utilizando o algoritmo SVM

Note-se, porém, que nem sempre é possível estabelecer diretamente uma separação linear entre classes. Os dados de entrada são mapeados por uma função matemática (*kernel*) para um novo domínio de vetores de características. Esta conversão aumenta as dimensões do espaço o que possibilita a definição de um hiperplano linear que minimiza os erros de classificação.

Para resolver o problema, optamos por selecionar os seguintes parâmetros:

- Problema: suporte para classificação do vetor com N classes que permita a separação imperfeita de classes;
- Tipo de *kernel*: Linear. A discriminação linear é feita no espaço original de características. É a opção mais rápida;
- Critérios de término do processo: definimos como 1000 o máximo de iterações possíveis e uma precisão de 0,000001.

### 6.6.8. Considerações

Ao nível do espaço de cor escolhido para a segmentação, concluímos que tanto o RGB normalizado, o HSV ou o YCbCr poderiam ser utilizados. Isto porque as transformações de espaços de cores não auxiliam na detecção da pele, devido à variedade de cores que esta pode ter. Assim, um esforço computacional extra na conversão do espaço RGB para outro qualquer é desnecessário. Para o nosso trabalho, a importância da segmentação advém da escolha do valor ótimo do limiar. Em casos de imagens com muitas variações de cor deve-se utilizar uma abordagem com vários níveis de limiar adaptativos às várias seções da imagem.

Em relação à aplicação do filtro de suavização, este poderia ter sido aplicado apenas com uma máscara de suavização vertical. A suavização horizontal beneficia, maioritariamente, a eliminação de ruídos provenientes dos sinais de televisão, perdendo definição de imagem em forma de colunas verticais. A suavização vertical propõe a eliminação do granulado da imagem, que aparece em muitas câmaras de captura de vídeo devido à sua forma de quantização e codificação do sinal. A imagem perde definição em forma de linhas horizontais. Ao aplicar os dois tipos de suavização perdemos ainda mais detalhe na imagem.

A aplicação dos dois filtros provém da não valorização da definição da imagem pois, neste caso, os detalhes perdidos são irrelevantes. O que é realmente significativo são as bordas da mão humana.

A escolha da extração das características geométricas foi efetuada porque o processamento de toda a imagem é computacionalmente muito exigente e introduz uma quantidade inaceitável de ruído (imagens com muitos pixels de informação irrelevante no fundo). Com a nossa escolha, pretendemos reduzir as informações redundantes e, com isto, diminuir o esforço computacional necessário mas, ao mesmo tempo, manter a singularidade do objeto específico de interesse.

Relativamente à plataforma escolhida, o OpenCV, realçamos a necessidade da sua complexa instalação e configuração. É necessário integrar as suas bibliotecas na plataforma Microsoft Visual Studio C++ 2010 Express Edition. Esta configuração foi exigente, porque a documentação que acompanha o pacote OpenCV carece de alguns melhoramentos. Só foi possível concluir este processo após efetuar algumas pesquisas pela internet. Ainda assim, após a correta configuração, consideramos o seu desempenho bastante fiável.

O OpenCV suporta várias linguagens de programação, através da utilização de wrappers <sup>1</sup>. Estes existem para as linguagens C#, JAVA e Python.

Sabendo que a programação em si seria provavelmente mais fácil em C# ou Java, o motivo pelo qual decidimos utilizar as funções originais em C/C++ prende-se, uma vez mais, com a minimização do esforço computacional em tarefas não essenciais.

A documentação é apresentada nas linguagens de programação C e C++ e, como já referimos, necessita de aperfeiçoamentos.

Por fim, salientamos o facto de que a biblioteca está em constante evolução. No início do nosso trabalho, a versão existente era a 2.3. No final, a versão era denominada de 2.4.2.

Se, por um lado, mostra o constante avanço, por outro lado expõe os erros que ela ainda contém. Em pleno desenvolvimento deste trabalho, e no caso da utilização do método de classificação KNN, encontramos um erro conhecido por "Memory leak in nearest-neighbor training" que apenas foi resolvido na versão 2.4.2.

Sublinhamos novamente a preocupação relativa ao nível do custo computacional ao longo do desenvolvimento deste trabalho. Os algoritmos foram implementados para terem um desempenho admissível em tempo real com o intuito de realizar uma segmentação em ambientes internos, sujeitos a diferentes condições de iluminação.

---

<sup>1</sup> Objeto que permite encapsular outro objeto por forma a alterar a interface ou comportamento deste último.

## 7. Testes

### 7.1. Base de Dados

Para a elaboração da BD, foram guardadas 100 imagens representativas de cada gesto. Para testes, foram utilizadas 5 mãos humanas distintas para guardar 100 imagens de cada gesto, para cada uma das mãos.

Nos Anexos deste trabalho expomos algumas destas imagens.

### 7.2. Segmentação da mão

Uma enorme vantagem da abordagem por nós adotada, aponta para a remoção da influência da luminosidade durante o processo de segmentação, fazendo com que esta se torne menos dependente das condições de iluminação, o que tem sido sempre um obstáculo crítico para o reconhecimento de imagem. Os valores do limiar ótimo são obtidos a partir do histograma da imagem utilizando o método dos picos e vales como relatado em 6.6.3.

A figura seguinte procura exemplificar as etapas da segmentação, onde cada coluna representa um passo necessários para a segmentação e as linhas representam os gestos retirados com iluminação reduzida, ótima e alta, respetivamente:

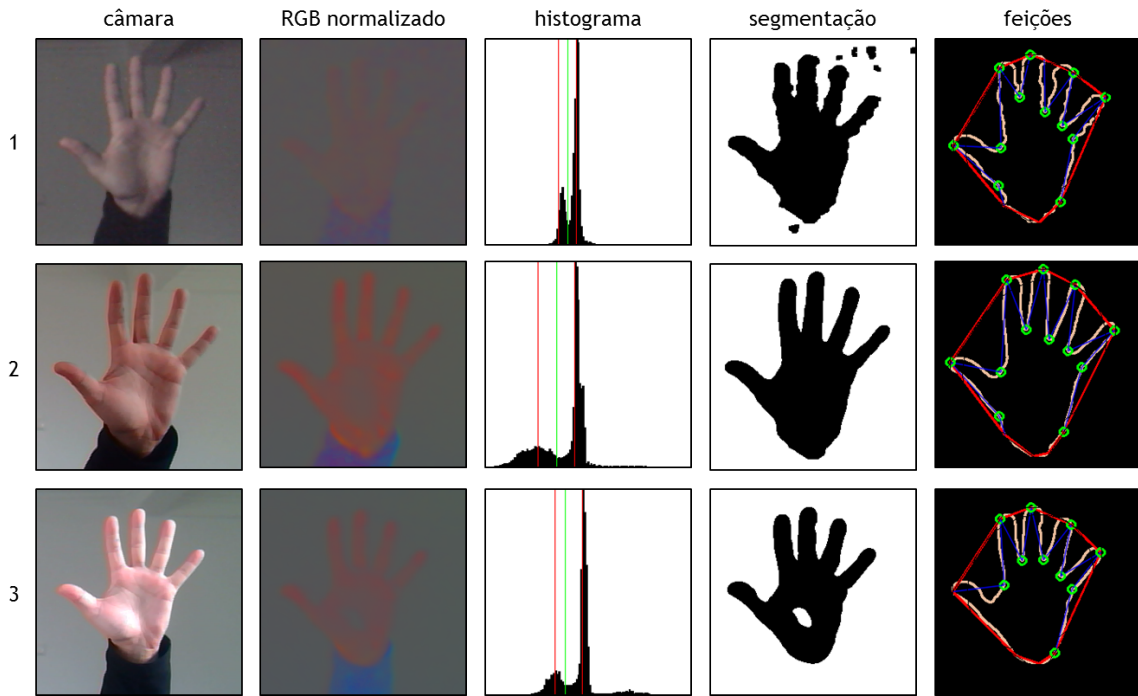


Fig. 33 - Representação das etapas da segmentação com vários tipos de iluminação.

Para os diferentes tipos de iluminação, podemos verificar alterações em todo o processo de segmentação excetuando na etapa final, detecção de feições.

Na coluna das imagens que representam os gestos no espaço de cor RGB normalizado, podemos atestar que a cor da pele é sempre mais notória que as outras cores presentes na imagem.

Certificamos também que o método da detecção dos picos e dos vales para encontrar o limiar ótimo, se revela bastante útil, pois elimina a maioria dos reflexos, como pode ser visto nas imagens da coluna segmentação.

Com base nas figuras anteriores, apresentamos o valor final do limiar ótimo extraído como exemplificado na coluna de imagens histograma.

Imagem	Iluminação	1.º pico	2.º pico	Limiar ótimo
1	Reduzida	109	120	114
2	Ótima	95	118	107
3	Alta	106	123	112

Tab. 6 - Limiar ótimo para segmentação

Depois, as pequenas imperfeições encontradas são eliminadas na etapa de detecção de feições pois o nosso algoritmo toma apenas em conta o maior contorno externo encontrado na imagem. A única restrição prende-se com o tamanho dos objetos de cor de pele que possamos

encontrar na imagem pois se estes tiverem maior tamanho que a mão serão eles o alvo da extração de contornos.

Apresentamos alguns exemplos para ilustrar as situações descritas:

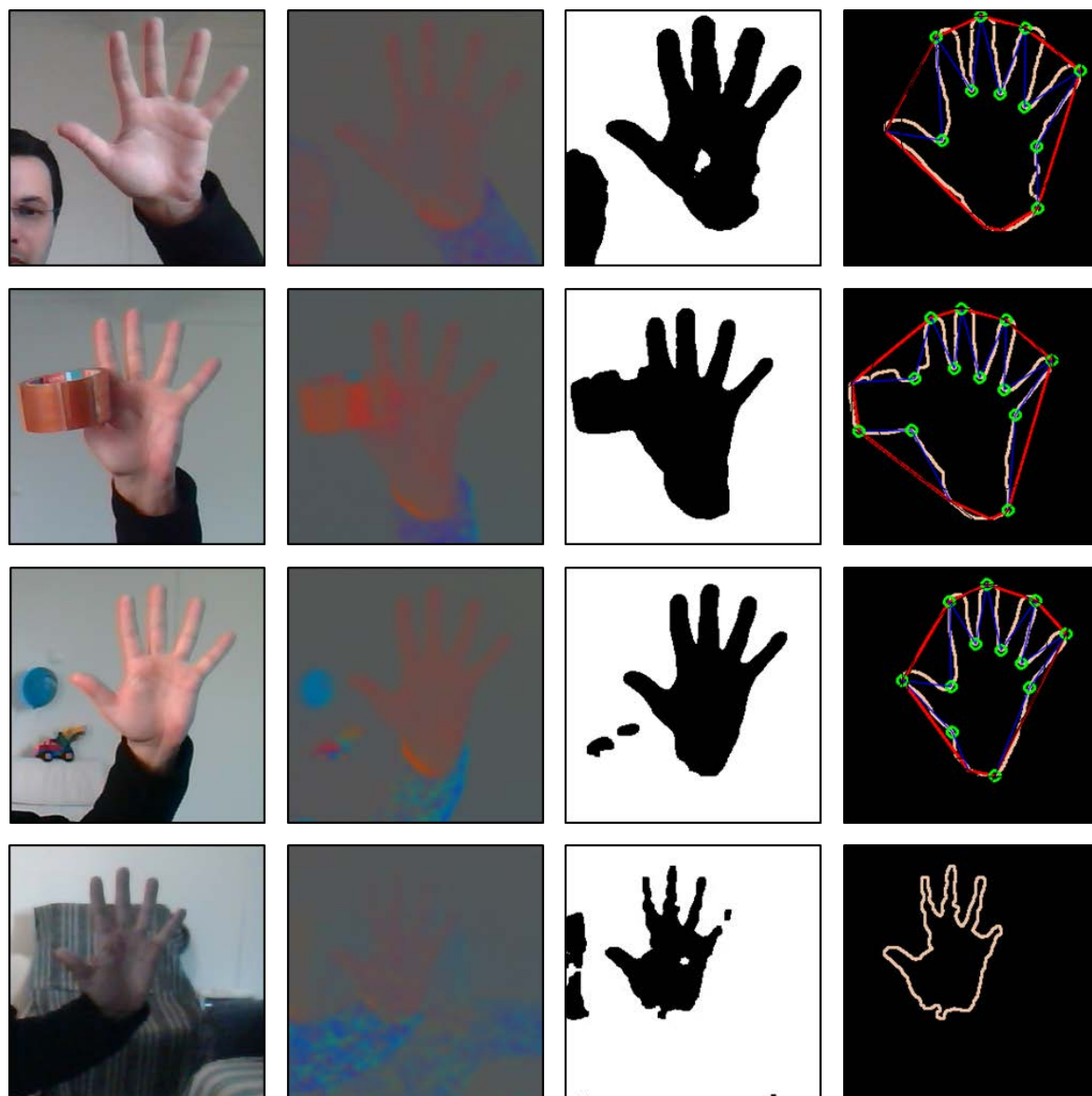


Fig. 34 - Segmentação em fundo não uniforme

Na imagem apresentada podemos observar que o nosso algoritmo apenas apresentou falhas quando notou a presença de um objeto na mão e quando combinado com um fundo não uniforme e uma iluminação reduzida. Isto acontece devido a uma indevida separação de cores aquando da análise do histograma para encontrar o valor ótimo do limiar.

### 7.3. Extração de características

A extração de características é uma fase crítica em todo o processo na medida em que é através dela que será efetuado o reconhecimento dos gestos. O vetor de características deverá captar a “essência” do gesto e deverá gastar pouco tempo de processamento.

A nossa metodologia propõe o uso de características geométricas em conjunto com os momentos invariantes de Hu (Hu, 1962). A utilização dos momentos advém da sua uniformidade de valores mesmo em gestos com rotações e escalas diferentes. Mostramos, seguidamente, o código de cálculo de algumas destas características.

```

. . .
cCompactness = pow(cPerimeter,2) / (cArea-4*CV_PI);
brAspRatio = (double)boundRect.width/boundRect.height;
brArea = boundRect.width*boundRect.height;
. . .

```

Fig. 35 - Código para calcular as características

Depois, é efetuada a normalização destas características para tornarmos os nossos dados consistentes. A normalização tem como objetivo que todos os valores estejam numa distribuição espacial proporcional, impossibilitando que os valores mais altos dominem completamente o processo. Isto poderia ocorrer porque alguns dos valores das características possuem um valor absoluto claramente maior do que os outros, como se pode verificar na tabela seguinte:

Parâmetro	Valor Real	Valor máximo	Valor mínimo	Valor Normalizado
C1	16.751	56.931	15.767	0.023
C2	0.652	1.245	0.364	0.326
C3	12972.0	40796	7062	0.175
C4	0.00002	0.002	0.0000002	0.009
C5	8.0	23	5	0.166
C6	22808.869	34803.625	9082.689	0.533
C7	-0.000000001	0.0000002	-0.0000004	0.610
C8	5.7E+16	1.2228E+17	-2.57937E+13	0.046
C9	-3.850	0	-89.672	0.957
C10	7.0	10	5	0.400

Tab. 7 - Valores normalizados do vetor de características

Assim, a normalização é feita através da fórmula:

$$caract\_norm = \frac{(caract - caract\_min)}{(caract\_max - caract\_min)}$$

*caract\_norm* = valor obtido,

*caract* = valor real,

*caract\_min* = menor valor obtido para a característica na BD,

*caract\_max* = maior valor obtido para a característica na BD,

que foi implementada com o seguinte código:

```

. . .
featVec.at<float>(0) = ((float)cCompactness - min_feat.at<float>(0)) /
(max_feat.at<float>(0) - min_feat.at<float>(0));

featVec.at<float>(1) = ((float)brAspRatio - min_feat.at<float>(1)) /
(max_feat.at<float>(1) - min_feat.at<float>(1));

featVec.at<float>(2) = ((float)brArea - min_feat.at<float>(2)) /
(max_feat.at<float>(2) - min_feat.at<float>(2));
. . .

```

Fig. 36 - Normalização das características extraídas

A partir dos vetores de contornos da Fig. 27, foi criada a seguinte tabela para mostrar os valores dos parâmetros escolhidos para o vetor de características.

Parâmetro	Gesto A	Gesto B	Gesto C	Gesto D	Gesto E	Gesto F
C1	0.014	0.043	0.793	0.182	0.404	0.250
C2	0.258	0.127	0.509	0.089	0.058	0.556
C3	0.535	0.437	0.905	0.381	0.329	0.602
C4	0.018	0.004	0.008	0.452	0.214	0.003
C5	0.277	0.222	0.666	0.333	0.333	0.444
C6	0.688	0.738	0.744	0.444	0.425	0.660
C7	0.609	0.612	0.614	0.385	0.583	0.612
C8	0.024	0.170	0.058	0.173	0.107	0.033

C9	0.987	0.014	0.257	0.961	0.924	0.607
C10	0.800	0.600	0.600	1.0	0.800	0.600

Tab. 8 - Valores dos parâmetros do vetor de características

Podemos verificar que os valores de cada característica são bastantes diferentes em relação a cada gesto, à exceção da característica 10 (tamanho do invólucro convexo). Isto permite-nos uma boa separação linear entre gestos.

#### 7.4. Treino e reconhecimento SVM

Função <i>Kernel</i>	Taxa de reconhecimento
Linear	99.1%
Polinomial	99.1%
RBF	97.0%
Sigmóide	16%

Tab. 9 - Classificação utilizando diferentes funções de *kernel* para as amostras de treino e teste

Da Tab. 9 podemos aferir que o classificador SVM é estável e eficiente, com boa capacidade de generalização para problemas com um número pequeno de amostras. No geral, podemos visualizar que, para diferentes tipos de funções do *kernel* com diferentes métodos de classificação, a precisão da classificação é igual para a função com *kernel* Linear e Polinomial. Isto acontece porque o nosso problema tornou-se linearmente separável quando executamos a normalização das características.

#### 7.5. Resultados da classificação

A tabela seguinte resume os resultados da classificação dos testes de gestos de mãos.

Gestos	A	B	C	D	E	F	TOTAL
A	497	1	0	0	0	2	500
B	4	489	0	0	6	1	500
C	0	0	494	0	4	2	500
D	0	1	0	498	1	0	500

E	0	0	0	0	500	0	500
F	4	0	0	0	0	496	500

Tab. 10 - Tabela com valores totais dos testes realizados.

Podemos encontrar na diagonal da Tab. 10, o número de gestos reconhecidos corretamente e nas células não diagonais, o número de gestos reconhecidos erradamente, isto é, falsos positivos.

De notar que no caso do gesto B e C, os falsos positivos se concentram no gesto E, o que nos permite considerar que o gesto E contém valores para as suas características, próximos dos valores considerados para a margem de separação entre as classes.

Subsequentemente, podemos resumir os resultados em termos de percentagem.

Gestos	Quadros	Corretos	Incorretos	% Corretos	% Incorretos
A	500	497	3	99.4%	0.6%
B	500	489	11	97.8%	2.2%
C	500	494	6	98.8%	1.2%
D	500	498	2	99.6%	0.4%
E	500	500	0	100%	0%
F	500	496	4	99.2%	0.8%

Tab. 11 - Resultados do reconhecimento dos gestos predefinidos.

Analisando a tabela anterior, aferimos que:

- Três gestos A, onze gestos B, seis gestos C, dois gestos D e quatro gestos F foram incorretamente reconhecidos que correspondem a percentagem de 0.6%, 2.2%, 1.2% 0.4% e 0.8% respetivamente;

Gestos	Quadros	Corretos	Incorretos	% Corretos	% Incorretos
A, B, C, D, E, F	3000	2974	26	99.1%	0.9%

Tab. 12 - Balanço final do reconhecimento de gestos.

Em jeito de conclusão, a Tab. 12 permite-nos aferir que o nosso trabalho apresenta uma percentagem de 99.1% de reconhecimento de gestos estáticos.

## 7.6. Tempos totais de processamento

Ao pretendermos que o nosso sistema opere em tempo real, somos obrigados a um esforço adicional para possuir capacidades de processamento por volta dos 30 quadros por segundo. De modo a avaliar esta relação, fez-se a medição de tempos de processamento utilizando as seguintes linhas de código:

```

. . .
// Calculo do numero de quadros por segundo
num_frames++;
elapsed = clock() - last_time;
int fps = 0;
if (elapsed > 1)
{
    fps=floor(num_frames/(float)(1+(float)elapsed/(float)CLOCKS_PER_SEC));
    num_frames = 0;
    last_time = clock() + 1 * CLOCKS_PER_SEC;
    printf("FPS: %d\n", fps);
}
. . .

```

Fig. 37 - Código para cálculo do tempo de processamento

Antes da chamada da rotina de processamento de imagem faz-se a leitura do tempo do sistema em milissegundos. Depois, para cada repetição da rotina é efetuada uma nova leitura do tempo do sistema. O tempo total de processamento é a diferença entre as duas medições.

Operação	Tempo
Captura da imagem	$\cong 1.6 \text{ ms}$
Aplicação de filtro de suavização	$\cong 16.5 \text{ ms}$
Conversão para RGB normalizado	$\cong 4.9 \text{ ms}$
Segmentação através da limiarização	$\cong 1.2 \text{ ms}$
Aplicação de filtros de remoção de ruído	$\cong 2.7 \text{ ms}$
Deteção de feições	$\cong 1.8 \text{ ms}$
Extração de características	$\cong 0.5 \text{ ms}$
Reconhecimento e classificação	$\cong 0.5 \text{ ms}$

TEMPO TOTAL:	$\cong 30 \text{ ms}$
--------------	-----------------------

Tab. 13 - Resultados obtidos em termos de tempos de processamento

Com tempos aproximadamente iguais a trinta milissegundos por cada iteração do nosso algoritmo, aferimos que a metodologia proposta opera em trinta quadros por segundo.

## 8. Conclusão

### 8.1. Conclusões

Ao longo do nosso trabalho foram expostos os conceitos chave que o desenvolvimento de um sistema de reconhecimento de gestos estáticos abrange. Aduzimos várias noções diferentes para cada uma das etapas constituintes do nosso trabalho. Explorámos vários espaços de cor para a representação de imagens, múltiplos tipos de filtros para remover possíveis ruídos, diferentes modos de desagregar regiões de interesse do fundo da imagem e diversos tipos de classificação de objetos. O desafio residiu na seleção dos vários métodos a aplicar e na sua conjugação para produzir o resultado esperado.

Também realizámos pesquisas e análises bibliográficas dos trabalhos relativos à temática em questão, desde algoritmos de reconhecimento de gestos, a segmentações e classificação de imagens e da biblioteca de VC, o OpenCV.

Depois, propusemos um conjunto de técnicas que pudessem fazer com que o trabalho proposto operasse em tempo real para a deteção de gestos da mão em ambientes de trabalho convencionais, com base em algoritmos de limiarização e deteção de contornos.

Foi também criada uma BD com imagens da mão sem qualquer tipo de restrições durante a aquisição, testadas por cinco indivíduos diferentes.

Por fim apresentámos os resultados da metodologia seguida.

Podemos dizer que desenvolvemos um sistema de reconhecimento de gestos estáticos desde a fase de aquisição até à fase de identificação.

Sabendo que a construção do protótipo destacado não seria possível de construir com base nos recursos de que dispúnhamos, restringimos o âmbito do nosso trabalho ao reconhecimento de gestos, ou seja, queríamos identificar o gesto estático mostrado pelo ocupante da cadeira de rodas.

Os objetivos deste trabalho baseavam-se no estudo, implementação e melhoramento das técnicas existentes aplicadas em sistemas de reconhecimento de gestos de mãos estáticas, mais propriamente, no reconhecimento baseado na geometria da mão. Podemos afirmar que o sistema funciona com eficiência e precisão com gestos dissemelhantes, mas poderá mostrar alguma fraqueza no uso de gestos idênticos, tais como a linguagem de sinais.

Mostrámos que o método de segmentação por nós utilizado revela alguma imprecisão com a variação das condições do cenário de fundo e da iluminação. Podemos indicar o estudo de outros métodos que respondam melhor não só à variação de cor como à presença de objetos no cenário de captura dos gestos. Algoritmos que deverão responder melhor a estas condições serão os algoritmos baseados em redes neuronais.

Alcançámos a implementação de um sistema que utilizasse as características geométricas da mão para reconhecimento de gestos e procurámos, ao mesmo tempo, superar as limitações encontradas em sistemas já existentes que residiam, principalmente, nas restrições impostas na fase de captura da mão, desempenho do sistema e seleção/precisão das características extraídas.

Assim, em condições de ambiente com algumas limitações, o sistema proposto pode ser caracterizado como uma NUI apta a ser utilizada em aplicações de IHC.

Este trabalho teve o mérito de ser apresentado na Conferencia Internacional, "Human Work Interaction Design", em Dezembro de 2012, será publicado um artigo no livro da Springer Verlag, "Work analysis and HCI" e ainda aceite para publicação no "International Journal of Sociotechnology and Knowledge Development".

## 8.2. Trabalho Futuro

As NUI estão entre as dez tecnologias mais inovadoras do mundo atual (Moore, 2008). Em todos os documentos analisados foi possível encontrar abordagens inovadoras e importantes para o desenvolvimento de uma aplicação que nos ajude a definir novas formas de interação.

No nosso caso, os veículos autónomos assistidos, como sistemas programáveis capazes de executar multitarefas utilizados para transporte de material, combate a incêndios, desarme de explosivos e outras atividades que podem colocar o ser humano em risco, podem e devem, na realidade, ser implementadas.

Constatamos igualmente que a integração de várias destas ferramentas é ainda feita em regime de trabalho experimental, mas que é possível aplicar a teoria à prática.

Atualmente, as NUI revelam uma grande procura por parte do mercado, recebendo um enorme investimento, mas muitas questões estão ainda sem resposta...

Poderemos captar todas as subtilezas da comunicação humana, independentemente do sexo, nacionalidade ou idade?

Quantas diferentes interfaces naturais poderão existir em simultâneo?

Como é que a comunicação acidental será tratada?

Estas e outras questões poderão levar ainda algum tempo a obter uma resposta objetiva.

## 9. Bibliografia

- Acharya, T., & Ray, A. K. (2005). *Image Processing: Principles and Applications*. John Wiley & Sons.
- Aryunto, K., & Limpraptono, F. (2009). A New Color Segmentation Method Based on Normalized RGB Chromaticity Diagram, Seminar on Intelligent Technology and Its Applications. *Proceedings of The 10th Seminar on Intelligent Technology and Its Application (SITIA)*. Indonesia.
- Baek, K., Draper, B., Beveridge, J., & She, K. (2002). PCA vs. ICA. *Proc. of the 4th International Conference on Computer Vision*, (pp. 824-827). Durham.
- Balcells, A. C., & Gonzalez, J. A. (1998). TetraNauta: A wheelchair controller for users with very severe mobility restrictions. *3rd Annual TIDE Congress*. Helsinki.
- Bartlett, M., Movellan, J., & Sejnowski, T. (2002). Face recognition by independent component analysis. *IEEE Transactions on Neural Networks* (pp. 1450-1464). New Jersey: IEEE Press Piscataway.
- Belhumeur, P., Hespanha, J., & Kriegman, D. (1996). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *Proceedings of the 4th European Conference on Computer Vision*, (pp. 45-58). Cambridge.
- Bell, D., Borenstein, J., Levine, S., Koren, Y., & Jaros, J. (1994). An assistive navigation system for wheelchairs based upon mobile robot obstacle avoidance. *IEEE International Conference on Robotics and Automation*, (pp. 2018-2022). San Diego.
- Bersch, R. (2009). *Design de um serviço de tecnologia assistiva em escolas públicas*. Obtido de Lume - Repositório Digital: <http://www.lume.ufrgs.br/handle/10183/18299>

- Birdwhistell, R. (1970). *Kinesics and Context: essays on body motion communication*. Philadelphia: University of Pennsylvania Press.
- Birk, H., Moeslund, T. B., & Madsen, C. B. (1997). Real-Time Recognition of Hand Alphabet Gestures Using Principal Component Analysis. *10th Scandinavian Conference on Image Analysis*.
- Blake, J. (2010). *Multi-touch Development with WPF and Silverlight*. Manning Publications.
- Blake, J. (2010). *Natural User Interfaces in .NET*. Manning Publications.
- Bourhis, G., & Pino, P. (1996). Mobile robotics and mobility assistance for people with motor impairments: rational justification for the VAHM project. *IEEE Transactions on Rehabilitation Engineering*, (pp. 7-12).
- Buxton, B. (06 de Janeiro de 2010). CES 2010: NUI with Bill Buxton. (L. Larsen, Entrevistador)
- Buxton, W., Billingham, M., Guiard, Y., Sellen, A., & Zhai, S. (1994). *Human Input to Computer Systems: Theories, Techniques and Technology*. Obtido de Bill Buxton: <http://www.billbuxton.com/inputManuscript.html>
- Cassinelli, A., Zerroug, A., & Ishikawa, M. (2009). *Virtual Haptic Radar*. Obtido de Ishikawa Oku Laboratory: <http://www.k2.t.u-tokyo.ac.jp/perception/VirtualHapticRadar/>
- Chang, Y., & Li, X. (1994). Adaptive Image Region Growing. *Image Processing, IEEE Transactions on*, 868-872.
- Chathuranga, S., Samarawickrama, K., Chandima, H., Chathuranga, K., & Abeykoon, A. (2010). Hands free interface for Human Computer Interaction. *2010 5th International Conference on Information and Automation for Sustainability (ICIAFs)* (pp. 359-364). IEEE.
- Chhabria, S., & Dharaskar, R. (2012). Multimodal Interface for Disabled Persons. *International Journal of Computer Science and Communication V*, 223-228.
- Clayton, S. (26 de 01 de 2011). *Microsoft is Imagining a NUI future*. Obtido de The Official Microsoft Blog: [http://blogs.technet.com/b/microsoft\\_blog/archive/2011/01/26/microsoft-is-imagining-a-nui-future-natural-user-interface.aspx](http://blogs.technet.com/b/microsoft_blog/archive/2011/01/26/microsoft-is-imagining-a-nui-future-natural-user-interface.aspx)
- Corcoran, P., Nanu, F., Petrescu, S., & Bigioi, P. (2012). Real-Time Eye Gaze Tracking for Gaming Design and Consumer Electronics Systems. *IEEE Transactions on Consumer Electronics*, 347-355.
- Costa, F. L., & Jr., R. M. (2001). *Shape Analysis and Classification Theory and Practice*. CRC Press.
- Dehariya, V., Shrivastava, S., & Jain, R. (2010). Clustering of Image Data Set Using K-Means and Fuzzy K-Means Algorithms. *Computational Intelligence and Communication Networks*, (pp. 386-391).
- Delac, K., Grgic, M., & Grgic, S. (2006). Independent comparative study of PCA, ICA, and LDA on the FERET data set. *International Journal of Imaging Systems and Technology*, (pp. 252-260).

- Fourney, A., Terry, M., & Mann, R. (2010). Gesturing in the wild: understanding the effects and implications of gesture-based interaction for dynamic presentations. *Proceedings of BCS HCI*, (pp. 230-240 ).
- Gonzalez, R. C., & Woods, R. E. (2002). *Digital Image Processing (2nd Edition)*. New Jersey: Prentice Hall.
- Gribble, W. S., Browning, R. L., Hewett, M., Remolina, E., & Kuipers, B. J. (1998). Integrating vision and spatial reasoning for assistive navigation. *Assistive technology and artificial*, 179-193.
- Hu, M.-K. (1962). Visual pattern recognition by moment invariants. *IRE Trans. on Information Theory*, (pp. 179-187).
- Jia, P., Hu, H., Lu, T., & Yuan, K. (2007). Head Gesture Recognition for Hands-free Control of an Intelligent Wheelchair. *Industrial Robot: An International Journal*, 34, 60-68.
- Jiang, X., Zhang, R., & Nie, S. (2009). Image Segmentation Based on PDEs Model: a Survey. *Bioinformatics and Biomedical Engineering*, (pp. 1-4). Shanghai.
- Kain, A. J. (1989). *Fundamentals of Digital Image Processing*. Prentice Hall.
- Kang, W., Yang, Q., & Liang, R. (2009). The Comparative Research on Image Segmentation Algorithms. *IEEE Conference on ETCS*, (pp. 703-707 ). Wuhan.
- Katevas, N., Sgouros, N., Tzafestas, S., Papakonstantinou, G., Beattie, P., Bishop, J., . . . Koutsouris, D. (1997). The autonomous mobile robot SENARIO: a sensor aided intelligent navigation system for powered wheelchairs. *Robotics & Automation Magazine*, 60-70.
- Kobayashi, Y., Kinpara, Y., Shibusawa, T., & Kuno, Y. (2009). Robotic wheelchair based on observations of people using integrated sensors. *IROS'09 Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems*, 2013-2018.
- Kobayashi, Y., Kinpara, Y., Shibusawa, T., & Kuno, Y. (2009). Robotic wheelchair based on observations of people using integrated sensors. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, (pp. 2013-2018).
- Krueger, W. M. (1991). *Artificial Reality II*. Addison-Wesley.
- Kumar, K. M., & Dinesh, M. (2012). Hand Gesture Recognition For Wheelchair Controlled by Digital Image Processing. *International Journal of Communications and Engineering*, (pp. 67-71).
- Lina, C.-S., Huana, C.-C., Chana, C.-N., Yehb, M.-S., & Chiu, C.-C. (2004). Design of a computer game using an eye-tracking device for eye's activity rehabilitation. *Optics and Lasers in Engineering*, (pp. 91-108).
- Liu, C., & Wechsler, H. (1999). Comparative assessment of independent component analysis (ICA) for face recognition . *In Audio- and Video-Based Biometric Person Authentication*, (pp. 212-216). Washington D.C.
- Luo, R., Chen, T. M., & Lin, M. H. (1999). Automatic guided intelligent wheelchair system using hierarchical grey-fuzzy motion decision-making algorithms. *Proceedings. 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (pp. 900-905).
- Madarasz, R. L., Heiny, L. C., Crompton, R. F., & Mazur, N. M. (1986). The design of an autonomous vehicle for the disabled. *IEEE J. Robotics Automation RA-2(3)*, 117-126.

- Madarasz, R. L., Heiny, L. C., Crompton, R. F., & Mazur, N. M. (1986). The design of an autonomous vehicle for the disabled. *IEEE J. Robotics and Automat*, (pp. 117-126).
- Mazo, M., Rodriguez, F., Lazaro, J. L., Urena, J., Garcia, J. C., Santiso, E., . . . Garcia, J. J. (1995). Wheelchair for physically disabled people with voice, ultrasonic and infrared sensor control. *Autonomous Robots*, v.2, no.3, 203.
- McNeill, D., & Levy, E. (1982). *Conceptual Representations in Language Activity and Gesture*. John Wiley and Sons Ltd.
- Merrill, D., Kalanithi, J., & Maes, P. (2007). Siftables: Towards Sensor Network User Interfaces. *Proceedings of the First International Conference on Tangible and Embedded Interaction (TEI'07)*. Louisiana: ACM.
- Mingqiang, Y., Idiyo, K., & Joseph, R. (2008). A Survey of Shape Feature Extraction Techniques. *Pattern Recognition Techniques, Technology and Applications*, (pp. 43-90).
- Mitra, S., & Acharya, T. (2007). Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, (pp. 311-324).
- Monson-Haefel, R. (15 de 01 de 2010). *Proposed Definition of Natural User Interface (NUI)*. Obtido de Discussing the Business of App Books: <http://theclevermonkey.blogspot.pt>
- Moore, S. (8 de Maio de 2008). *Gartner Identifies Top Ten Disruptive Technologies for 2008 to 2012*. Obtido de Gartner: <http://www.gartner.com/it/page.jsp?id=681107>
- Mulder, A. (1996). *Hand gestures for hci*. Vancouver: Simon Fraser University.
- Murakami, Y., Kuno, Y., Shimada, N., & Shirai, Y. (2001). Collision avoidance by observing pedestrians' faces for intelligentwheelchairs. *IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 2018-2023). Maui: IEEE.
- Nakanishi, S., Kuno, Y., Shimada, N., & Shirai, Y. (1999). Robotic wheelchair based on observations of both user and environment. *Proceedings. 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (pp. 912-917).
- Nespoulous, J.-L., Perron, P., & Roch, A. (1986). *The Biological Foundations of Gestures: Motor and Semiotic Aspects*. Hillsdale: Lawrence Erlbaum Associates.
- Nisbet, P. D., Craig, J., Odor, J. P., & Aitken, S. (1995). "Smart" wheelchairs for mobility training. *Technol Disabil*, 49-62.
- Oviatt, S. (2003). Multimodal interfaces. In J. Jacko, & A. Sears, *The Human-Computer Interaction Handbook* (p. 286.304). New Jersey: Lawrence Erlbaum and Associates.
- Pavlovic, V. I., Sharma, R., & Huang, T. S. (1997). Visual Interpretation of Hand Gestures for Human Computer Interaction: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (pp. 677-695). IEEE Computer Society Washington.
- Porto Editora. (18 de 07 de 2012). *Dicionário da Língua Portuguesa - com Acordo Ortográfico*. Obtido de Infopédia - Enciclopédia e Dicionários Porto Editora: <http://www.infopedia.pt/lingua-portuguesa/gesto>
- Radabaugh, M. P. (07 de 29 de 2012). *NIDRR's Long Range Plan - Technology for Access and Function Research*. Obtido de National Center for the Dissemination of Disability Research - Section Two: NIDDR Research Agenda - Chapter 5: TECHNOLOGY FOR ACCESS

- AND FUNCTION: [http://www.ncddr.org/new/announcements/lrp/fy1999-2003/lrp\\_techaf.html](http://www.ncddr.org/new/announcements/lrp/fy1999-2003/lrp_techaf.html)
- Radicchi, A., Nunes, A., & Botega, L. (2010). Proposta de Desenvolvimento de Interface Tangível para Aplicações de Gerenciamento de Emergência. *XI Symposium on Virtual and Augmented Reality*. Natal.
- Rautaray, S., & Agrawal, A. (2010). A Vision based Hand Gesture Interface for Controlling VLC Media Player. *International Journal of Computer Applications*.
- Rocchetti, M., Marfia, G., & Semeraro, A. (2011). A Fast and Robust Gesture Recognition System for Exhibit Gaming Scenarios. *Proceedings of the 4th International ICST Conference on Simulation Tools and Techniques* (pp. 343-350). Barcelona: ICST/ACM.
- Roefer, T., & Lankenau, A. (2000). Architecture and Applications of the Bremen Autonomous Wheelchair. *Information Sciences*, 1-20.
- Rokeby, D. (24 de 11 de 2010). *Very Nervous System (1986-1990)*. Obtido de David Rokeby - Artist: <http://www.davidrokeby.com/vns.html>
- Sarter, N. B. (2006). Multimodal information presentation: Design guidance and research challenges. *International Journal of Industrial Ergonomics*, 439-445.
- Shan, C., Tan, T., & Wei, Y. (2007). Real-time hand tracking using a mean shift embedded particle filter. *Journal Pattern Recognition* (pp. 1958-1970). New York: Elsevier Science Inc.
- Shapiro, L., & Stockman, G. (2001). *Computer Vision*. Prentice-Hall.
- Smola, A. (1999). Geometry and invariance in kernel based methods. In B. Schölkopf, C. Burges, & A. Smola, *Advances in Kernel Methods: Support Vector Learning* (pp. 89-116). MIT Press.
- Solomon, C., & Breckon, T. (2011). *Fundamentals Of Digital Image Processing: A Practical Approach with Examples in Matlab*. John Wiley & Sons, Ltd.
- Tahboub, K., & Asada, H. H. (1999). A semi-autonomous control architecture applied to robotic wheelchairs. *Proceedings. 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (pp. 906-911).
- Tse, M. C., & Luo, R. C. (2000). Multilevel multi-agent based team decision fusion for mobile robotbehavior control. *3rd World Congress on Intelligent Control and Automation* (pp. 489-494). Hefei: IEEE.
- Tsui, K. M., Kimb, D.-J., Behal, A., Kontak, D., & Yanco, H. A. (2011). "I Want That": Human-in-the-Loop Control of a Wheelchair-Mounted Robotic Arm. *Applied Bionics and Biomechanics*, 127-147.
- Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 71-86.
- Varshney, S., Rajpal, N., & Purwar, R. (2009). Comparative study of image segmentation techniques and object matching using segmentation. *Proceeding of International Conference on Methods and Models in Computer Science*, (pp. 1-6).
- Wakaumi, H., Nakamura, K., & Matsumura, T. (1992). Development of an automated wheelchair guided by a magnetic ferrite marker lane. *Journal of Rehabilitation Research and Development*, 27-34.

- Wang, H., Kang, C., Ishimatsu, T., & Ochiai, T. (1997). Auto-navigation of a wheelchair. *Artificial Life and Robotics*, 141-146.
- Weiser, M. (1991). The Computer for the 21st Century. *Scientific American*, vol.265, (pp. 94-104).
- Weiser, M., & Brown, J. (21 de Dezembro de 1995). *Designing Calm Technology*. Obtido em Janeiro de 2012, de Designing Calm Technology: <http://www.ubiq.com/weiser/calmtech/calmtech.htm>
- Wigdor, D., & Wixon, D. (2011). *D. Brave NUI world: designing natural user interfaces for touch and gesture*. Burlington: Morgan Kaufmann.
- Yanco, A. H. (1998). Wheellesley: a robotic wheelchair system: Indoor navigation and user interface. In V. O. Mittal, H. A. Yanco, J. Aronis, & R. Simspon, *Lecture Notes in Artificial Intelligence: Assistive Technology and Artificial Intelligence* (pp. 256-268).
- Yoder, J. D., Baumgartner, E. T., & Skaar, S. (1996). Initial results in the development of a guidance system for a powered wheelchair. *IEEE Trans Rehabil Eng* 4, 143-151.
- Zang, X., Chen, X., Wang, W., Yang, J., Lantz, V., & Wang, K. (2009). Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors. *Proceedings of the 13th international conference on Intelligent user interfaces* (pp. 401-406). New York: Springer.
- Zhang, Y., Qu, H., & Wang, Y. (2006). Adaptive Image Segmentation Based on Fast Thresholding and Image Merging. *Artificial Reality and Telexistence--Workshops*, (pp. 308-311).

## 10. Anexos

No Anexo A - Gestos da BD - expomos as imagens utilizadas na BD. Contém quinhentas imagens capturadas apenas por um indivíduo.

No Anexo B - Gestos utilizados para teste - exibimos as imagens utilizadas para testes. São correspondentes à captura da mão de cinco indivíduos com idades compreendidas entre os 14 e os 35 anos que perfizeram um total de três mil imagens.

Salientamos o facto de que todas as imagens apresentadas foram escolhidas aleatoriamente.

10.1. Anexo A - Gestos da BD



## 10.2. Anexo B - Gestos utilizados para teste

