

TRADUTOR DA LINGUA GESTUAL PORTUGUESA
MODELO DE TRADUÇÃO BIDIRECCIONAL

Odair Roberto Santiago Amarante Oliveira

Dissertação para obtenção do Grau de Mestre em
Engenharia Informática, Área de Especialização em
Sistemas Gráficos e Multimédia

Orientador: Paula Escudeiro

Coorientador: Nuno Escudeiro

Porto, Outubro de 2013

Dedicatória

Pelo apoio incondicional, pela pessoa que é, batalhadora pelos seus filhos, pelo encorajamento durante todos estes anos nunca me deixando desistir, que me mostrou que um dos caminhos mais importantes da nossa vida é o de lutar pelos nossos objetivos. Pelo amor incondicional, pelos ensinamentos que formaram o “homem de amanhã” que sempre quis que fosse e que continuo a fazer de tudo para o conseguir ser.

Dedico esta tese a ti, Mãe, Ana Maria Santiago Amarante.

Aos meus irmãos Fábio e Stephen Santiago, as minhas irmãs Karine e Yasmin Santiago, a minha tia Rosa Santiago, que sempre estiveram comigo neste trajeto, seja em pensamento seja presente fisicamente. Sempre me deram forças para continuar, nunca me deixando desistir do nosso objetivo.

Resumo

O projeto tem como objetivo desenvolver e avaliar um modelo que facilita o acesso para pessoas surdas ou com deficiência auditiva, o acesso ao conteúdo digital - em particular o conteúdo educacional e objetos de aprendizagem – a criação de condições para uma maior inclusão social de surdos e deficientes auditivos.

Pretende-se criar um modelo bidirecional, em que permite a pessoas com deficiências auditivas, possam se comunicar com outras pessoas, com a tradução da Língua Gestual Portuguesa (LGP) para a Língua Portuguesa (LP) e que outras pessoas não portadoras de qualquer deficiência auditiva possam por sua vez comunicar com os surdos ou deficientes auditivos através da tradução da LP para a LGP.

Há um conjunto de técnicas que poderíamos nos apoiar para desenvolver o modelo e implementar a API de tradução da LGP em LP. Muitos estudos são feitos com base nos modelos escondidos de Markov (HMM) para efetuar o reconhecimento. Recentemente os estudos estão a caminhar para o uso de técnicas como o “Dynamic Time Warping” (DTW), que tem tido mais sucesso do que outras técnicas em termos de performance e de precisão.

Neste projeto optamos por desenvolver a API e o Modelo, com base na técnica de aprendizagem Support Vector Machines (SVM) por ser uma técnica simples de implementar e com bons resultados demonstrados em reconhecimento de padrões.

Os resultados obtidos utilizando esta técnica de aprendizagem foram bastante ótimos, como iremos descrever no decorrer do capítulo 4, mesmo sabendo que utilizamos dois dispositivos para capturar dados de descrição de cada gesto.

Toda esta tese integra-se no âmbito do projeto científico/ investigação a decorrer no grupo de investigação GILT, sob a coordenação da professora Paula Escudeiro e suportado pela Fundação para Ciência e Tecnologia (FCT).

Keywords: LGP, LP, LG, Kinect, Data Glove, KDD, SVM

Abstract

The project aims to develop and evaluate a model that facilitates access for people who are deaf or hearing impaired access to digital content - in particular the educational content and learning objects - creating conditions for greater social inclusion of deaf and hearing.

It is intended to create a bi-directional model , which allows people with hearing disabilities to communicate with others , with the translation of Portuguese Sign Language (LGP) for the Portuguese Language (LP) and other people not suffering from any disability hearing may in turn communicate with the deaf or hearing impaired by translating the LP for the LGP .

There is a set of techniques that could support us to develop the model and implement the API translation of LGP to LP. Many studies are done based on hidden Markov models (HMM) to perform recognition. Recently studies are moving towards the use of techniques such as “Dynamic Time Warping “(DTW), which has been more successful than other techniques in terms of performance and precision.

In this project we chose to develop the API and model, based on the technique of learning Support Vector Machines (SVM) to be a simple technique to implement and demonstrated good results on pattern recognition.

The results obtained using this technique were very great learning, as we describe in Chapter 4 of the course, even though we use two devices for capturing data description of each gesture.

All of this thesis integrates within the scientific project / research course in GILT research group under the coordination of Professor Paula Squire and supported by the Foundation for Science and Technology (FCT).

Keywords: LGP, LP, LG, Kinect, Data Glove, KDD, SVM

Agradecimentos

A professora Doutora Paula Escudeiro, minha orientadora, ao professor Nuno Escudeiro, meu co orientador, pelo voto de confiança, pela fundamental contribuição no meu crescimento enquanto pesquisador.

Ao meu colega e pesquisador, Bruno Gouveia, por toda a contribuição que teve no desenvolvimento deste trabalho, pela paciência que teve em me integrar no desenvolvimento do trabalho que já havia feito.

Aos meus amigos irmãos, que sempre me apoiaram e sempre estiveram presentes nos momentos que mais precisava.

A todas as pessoas que, direta ou indiretamente, contribuíram para a concretização deste trabalho.

Índice

1	Introdução	17
1.1	Motivação	17
1.2	Metodologias / Estrutura	17
1.3	Linguagem	18
1.4	A Língua Gestual - Língua	18
1.5	Língua Gestual Portuguesa.....	18
1.6	Fonologia.....	20
1.7	Parâmetros Fundamentais da LGP - Configuração da mão (ou mãos).....	22
1.8	Parâmetros Fundamentais da - LGP Local de articulação (ou localização)	26
1.9	Parâmetros Fundamentais da LGP - Orientação	29
1.10	Parâmetros Fundamentais da LGP - Componentes não - manuais (expressão facial e/ou corporal)	30
1.11	Expressão	31
1.12	Língua Gestual no Mundo	32
2	Estado da Arte de Reconhecimento de Gestos.....	39
2.1	Introdução	39
2.2	Lúvas no Reconhecimento de Gestos da Mão - Reconhecimento da Configuração Manual.....	40
2.2.1	Lúvas 5DT	40
2.2.2	Lúvas CYBER.....	41
2.2.3	Microsoft Digits	41
2.3	Reconhecimento de Gestos Utilizando Visão	42
2.3.1	O Uso de sensores de profundidade.....	42
2.3.2	Microsoft Kinect	43
2.3.3	Mapa de profundidade.....	43
2.3.4	Kinect - Profundidade de foco	44
2.3.5	Kinect - Profundidade de som.....	45
2.3.6	Inferir a Posição do corpo.....	45
2.3.7	Primeiro Subprocesso	46
2.3.8	Segundo Subprocesso	49
2.4	Leap Motion	49
2.5	Aplicações	50
2.5.1	OMEK SDK	50
2.5.2	SoftKinectic SDK	51
2.5.3	3Gear's SDK	52
2.5.4	SigmaNIL SDK.....	53
2.6	Estudos - Reconhecimento de Línguas Gestuais.	53

3	Reconhecimento de Padrões / Mineração de Dados	56
3.1	Descoberta de Conhecimento	56
3.2	Data Mining	57
3.2.1	Inovação Contínua	57
3.2.2	Processo Data Mining	57
3.2.3	Tipos de Data Mining	58
3.2.4	Como funciona?	58
3.2.5	Modos de aprendizagem:	59
3.3	Aprendizagem da Máquina	59
3.4	Como Identificar?	60
3.5	Técnica de Aprendizagem Utilizada	60
4	Arquitetura do Sistema e Desenvolvimento do Software	68
4.1	Modelo Geral da Aplicação	70
4.2	Juntas de Interesse	71
4.3	Implementação baseado em Ficheiro	72
4.4	Método - Gestos Estáticos	73
4.4.1	Luva de dados e Kinect	74
4.4.2	Normalização dos Dados	75
4.4.3	Classificador	76
4.5	Modelo - Gestos em Movimento	80
4.5.1	Normalização dos Dados	81
4.5.2	Invariante a posição do utilizador	82
4.5.3	Invariante ao tamanho do utilizador	83
4.5.4	Classificador	83
4.5.5	Leitura do Ficheiro de Dado	84
4.5.6	Normalização dos Dados - Rapid Miner	85
4.5.7	Processo de Validação e Classificação	85
4.5.8	Resultados	87
4.6	Modelo Tradução Texto - LGP	91
5	Conclusão	93
5.1	Trabalho Futuro	95
6	Bibliografia	96
7	Anexos	98

Lista de Figuras

Figura 1 - Imagem adaptada de Jokinen, 2006, Fonte: (APSurdos_Comunidade).....	20
Figura 2 - Configuração Manual, Alfabeto (Aa à Nn), Fonte: (Baltazar, 2012).....	22
Figura 3 - Configuração Manual, Alfabeto (Oo à Zz), Fonte: (Baltazar, 2012).....	23
Figura 4 - Configuração Manual, Alfabeto (0 à 9), Fonte: (Baltazar, 2012).....	24
Figura 5 - Palavras Simples, Fonte: (Baltazar, 2012)	25
Figura 6 - Local de articulação dos gestos na cabeça e pescoço do praticante de Gestos da LGP - Fonte: (Baltazar, 2012).....	27
Figura 7 - Áreas de articulação dos gestos no tronco e nos membros do praticante de gestos da LGP- Fonte: (Baltazar, 2012).....	27
Figura 8 - Exemplos de gestos realizados no corpo do praticante de Gestos de Língua gestual Portuguesa, Fonte: (Baltazar, 2012).....	28
Figura 9 - Orientações da mão. Fonte: (Baltazar, 2012).....	29
Figura 10 - Demonstração de Orientação da mão num gesto, Fonte:(Baltazar, 2012)	29
Figura 11 - Movimentos essenciais na quirologia da LGP, Fonte: (Baltazar, 2012)	30
Figura 12 - Exemplos de Configuração não Manuais, Fonte:(Baltazar, 2012)	30
Figura 13 - Imagem da luva de dados 5DT (5DT, 5DT Data Glove Ultra Series, 2011).....	41
Figura 14 – Elementos constituintes da Microsoft Kinect.....	43
Figura 15 – Representação da técnica de análise de padrão conhecido, Zhang et al, 3DPVT (2002)	44
Figura 16 – Profundidade de Foco, Watanabe and Nayar, IJCV 27 (3), 1998	44
Figura 17 - Profundidade de Foco, Freedman et al, PrimeSense patent application US 2010/0290698	45
Figura 18 - Shotton et al, CVPR (2011)	45
Figura 19 – Primeiro Subprocesso, Shotton et al, CVPR (2011)	46
Figura 20 – Jogo “twenty questions”, Ntoulas et al, WWW (2006)	47
Figura 21 - Shotton et al, CVPR (2011)	47
Figura 22 - Shotton et al, CVPR (2011)	48
Figura 23 - Shotton et al, CVPR (2011)	48
Figura 24 - Dispositivo Leap Motion.....	49
Figura 25 - OMEK Beckon Fonte: (Beckon).....	51
Figura 26 - Soft Kinect	51
Figura 27 - Sistema 3 Gear's	52
Figura 28 - Processo Data Mining ou Mineração de Dados	56
Figura 29 - Algoritmo SVM	62
Figura 30 - Algoritmo SVM 1.....	62
Figura 31 - Algoritmo SVM - Margens	62
Figura 32 - Algoritmo SVM - HiperPlano	63
Figura 33 - Algoritmo SVM - Separação Vectores	64
Figura 34 - Algoritmo SVM - Divisão Não Linear	64

Figura 35 - Algoritmo SVM - Dimensões	64
Figura 36 - Algoritmo SVM - Interpoladores como o IDW	66
Figura 37 - Algoritmo SVM - Interpoladores RBF	66
Figura 38 - Algoritmo SVM - RBF	67
Figura 39 - Modelo Principal Tradutor Bidirecional de LGP	68
Figura 40 - Modelo Geral - Gesto para Texto.....	70
Figura 41 - Juntas de Interesse.....	71
Figura 42 - Aplicação Desenvolvida - Dados da Mão Direita - Kinect	72
Figura 43 - Modelo Captura Dados e Reconhecimento - Gestos Estáticos.....	73
Figura 44 - Aplicação Desenvolvida Dados da Luva	74
Figura 45 - Aplicação Desenvolvida - Reconhecimento em Tempo Real.....	79
Figura 46 - Modelo Reconhecimento - Gestos em Movimento.....	80
Figura 47 Rapid Miner - Classificação dos Gestos.....	84
Figura 48 - Rapid Miner - Leitura Ficheiro de Dados.....	84
Figura 49 - Rapid Miner - Normalização dos Dados.....	85
Figura 50 - Rapid Miner - Processo de Classificação e Validação.....	85
Figura 51 - Rapid Miner - Processos.....	86
Figura 52 - Rapid Miner - Resultados	87
Figura 53 - Rapid Miner - Matriz de Confusão	88
Figura 54 - Modelo Tradução Texto em LGP.....	91
Figura 55 - Exemplo de Avatar – Letra A (Alfabeto LGP).....	92

Lista de Tabelas

Tabela 1 - Acrónimos e Símbolos	16
Tabela 2 - Locais de Articulação	26
Tabela 3 - Performance do Modelo.....	89
Tabela 4 - Matriz de Confusão.....	90

Acrónimos e Símbolos

API	– Application Programming Interface
SDK	– Software Development Kit
OpenNI	– Standard Natural Interface Infrastructure.
NITE	– Natural Interaction Technology for End-user
Prime Sensor™	– The brand name behind Prime Sense's products. It refers to the reference design for a 3D camera.
Prime Sensor™ IC	– The chip developed by PrimeSense that is implemented in the 3D camera
PSDK	– Prime Sensor™ Development Kit
NUI	– Natural User Interface
IDE	– Integrated Developer Environment
UML	– Unified Modeling Language
WPF	– Windows Presentation Form
NUI	– Natural User Interface
BGR32	– Blue Green Red picture
GUI	– Graphical User Interface
XAML	– Extensible Application Markup Language
WPF	– Windows Presentation Foundation
ASL	– American Sign Language
LbD	– Learning by Demonstration
HOG	– Histogram of Oriented Gradients
kNN	– K Nearest Neighbors
NN	– Neural Network

LSH	–	Locality Sensitive Hashing
eNN	–	Efficient Nearest Neighbors
KDE	–	Kernel Density Estimation
GPU	–	Graphics processing unit
HCI	–	Computer Interaction
HMM	–	Hidden Markov Model
ANN	–	Artificial Neural Networks
DTW	–	Dynamic Time Warping

Tabela 1 - Acrónimos e Símbolos

1 Introdução

1.1 Motivação

O meu primeiro pensamento, desde o momento que vi a tese proposta pela professora e minha orientadora Paula Escudeiro, foi e continua a ser: “Essa é a tese que mais se identifica com a minha forma de estar e ser. ”. Isso porque sempre tive a motivação de fazer algo com o objetivo de ajudar, seja na minha vida pessoal ou profissional.

Tive o meu primeiro contacto com a Língua Gestual Portuguesa, quando numa altura convivi com uma professora da área, que quando viu o meu interesse por aprender a Língua Gestual Portuguesa (LGP), me entregou materiais que me possibilitariam aprender.

O meu objetivo com esta tese é o de encontrar ou construir formas de ajudar na aprendizagem, seja de pessoas com deficiências de audição ou pessoas que simplesmente pretendem aprender, para também se poderem comunicar com algum familiar ou amigo que tenham.

Por também ter tido a oportunidade de trabalhar num projeto integrado no âmbito de investigação que é suportado pela Fundação Ciência e Tecnologia, onde tinha a função de modelizar/implementar um tradutor bidirecional de Língua Gestual Portuguesa (LGP).

1.2 Metodologias / Estrutura

Os métodos utilizados nesta tese, muito têm a ver com os dispositivos que tínhamos disponíveis para efetuar o reconhecimento dos gestos da Língua Gestual Portuguesa (LGP).

Visto que a tese se insere num projeto integrado no âmbito da investigação, encontrou-se a partida desenvolvimentos que utilizavam dois dispositivos – Kinect e Luva de dados 5DT – separadamente. Para termos um maior número de valores que descrevem o gesto decidimos então juntar os dados que os dois dispositivos nos fornecem.

A ideia era construir uma descrição de um gesto com tantos dados quanto possíveis mas que não pusessem em causa a performance e a precisão no reconhecimento dos gestos.

Neste documento, iremos ter neste capítulo 1, uma introdução do que é Língua Gestual Portuguesa, de como é constituído. Iremos ter também neste capítulo um ponto importante que é o da história mundial dos surdos, as dificuldades que tiveram ao longo do tempo e de como as tem superado.

No capítulo 2, iremos ter o estado da arte, no desenvolvimento de aplicações para o reconhecimento da Língua Gestual. Iremos demonstrar também os dispositivos que hoje em

dia estão a ser utilizados. Também falaremos um pouco dos novos dispositivos e a forma como podem ajudar no reconhecimento da Língua Gestual.

No capítulo 3, iremos ter uma introdução sobre descoberta de conhecimento em que um dos pontos – Data Mining – será melhor descrito, visto que utilizamos esse ponto para a classificação dos nossos dados. Dentro deste ponto iremos apresentar a técnica de aprendizagem da máquina que utilizamos para a classificação dos dados da descrição de cada gesto.

No capítulo 4, teremos então o trabalho efetuado com a modelização e implementação da nossa API. Neste capítulo iremos, descrever as abordagens que tivemos perante o reconhecimento dos gestos da Língua Gestual Portuguesa (LGP), a estratégia que utilizamos para formar os dados da descrição dos gestos, a forma como classificamos estes dados e as ferramentas que utilizamos para fazer esta operação. Será também neste capítulo que iremos apresentar os resultados obtidos.

No capítulo 5, teremos então a conclusão do nosso trabalho e teremos também a descrição de trabalhos futuros que podem e devem ser implementados para que a nossa API possa dar melhores resultados na tradução da Língua Gestual Portuguesa em Língua Portuguesa.

1.3 Linguagem

Linguagem é a capacidade inata, localizada no hemisfério esquerdo, que o ser humano tem em utilizar e compreender uma língua, sistemas organizados de signos arbitrários e convencionais partilhados por uma comunidade (Fromkin & Rodman, 1993).

Isso significa que as pessoas têm a capacidade de produzir sons que tem certo significado e compreender ou interpretar sons produzidos por outras pessoas.

1.4 A Língua Gestual – Língua

(ApSurdos) Quando falamos de LGP surge a questão: É uma Língua ou uma Linguagem? Pode desde logo induzir em erro. Portanto devemos saber qual das palavras é a correta a utilizar. Linguagem ou Língua – Gestual Portuguesa. A resposta seria que a melhor palavra a utilizar é a de Língua – Gestual Portuguesa (LGP) e iremos demonstrar o porque nos pontos que se seguem.

1.5 Língua Gestual Portuguesa

(ApSurdos)A Língua Gestual Portuguesa em Portugal, LGP, nasceu com a primeira escola de surdos, em 1823, em Lisboa, na escola Casa Pia, sendo o primeiro educador um sueco que da Suécia trouxe o alfabeto Manual. Apesar de não se notarem semelhanças ao nível do

vocabulário, o alfabeto da LGP e o da Língua Gestual da Suécia, “Svenskt teckensprak”, continuam a revelar a sua origem comum.

O Alfabeto Manual.

O alfabeto manual (ou alfabeto gestual, ou dactilologia) só é utilizado, ocasionalmente, quando há a necessidade de dizer um nome próprio de alguém ou o nome de uma localidade ou uma palavra que não se conhece. Regra geral, na comunicação, os surdos não sentem grande necessidade de recorrer ao alfabeto manual, uma vez que os conceitos têm todos gestos correspondentes.

Reconhecimento.

A investigação da LGP só começou no final dos anos 1970, na Faculdade de Letras da Universidade de Lisboa, sendo que a nível internacional, as descrições linguísticas fizeram-se pela primeira vez com a ASL, Língua Gestual Americana, na Universidade de Gallaudet, nos anos 1960. Em consequência deste enquadramento científico, a LGP foi reconhecida pela Constituição da República, em 1997, numa altura em que apenas 5 países do mundo inteiro o tinham feito.

Diário da República – I Série A – n.º 218 – 20/09/1997 – Lei Constitucional

Artigo 74.º - alínea h) Ensino

h) Proteger e valorizar a língua gestual portuguesa, enquanto expressão cultural e instrumento de acesso à educação e da igualdade de oportunidades.

Em Portugal, estima-se que existam cerca de 120 000 pessoas com algum grau de perda auditiva (incluindo aqui os idosos que vão perdendo a audição gradualmente) e cerca de 30 000 surdos falantes nativos de língua gestual portuguesa (na sua maioria surdos severos e profundos).

Os grupos, (APSurdos_Comunidade), que constituem a comunidade de utilizadores da língua gestual são:

1. Surdos que têm a LGP como sua primeira língua (geralmente filhos de pais surdos);
2. Surdos que a adquirem durante a escolaridade (normalmente com surdez severa a profunda);
3. Familiares, profissionais e amigos que se relacionam com surdos diariamente;
4. Curiosos.

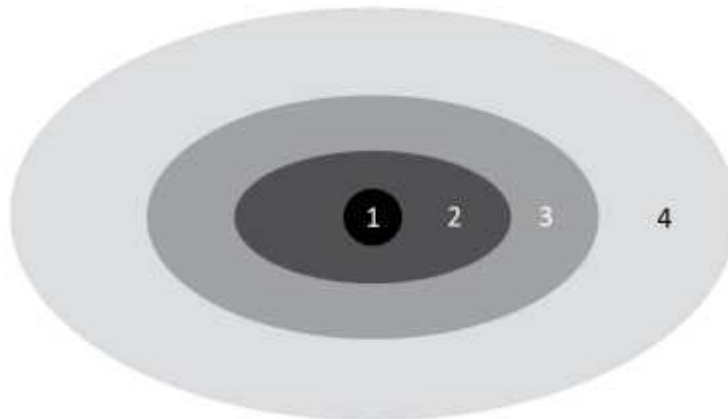


Figura 1 - Imagem adaptada de Jokinen, 2006, Fonte: (APSurdos_Comunidade)

Os três últimos grupos aprendem a língua gestual como língua segunda (ou “estrangeira”). Logo, o número de trinta mil (30 000) falantes pode ser triplicado no sentido de formar a comunidade de utilizadores da LGP. Neste conjunto é importante reter que a maioria (grupo 4 e parte do grupo 3) não é fluente na língua.

1.6 Fonologia

(Correia, 2009) Pelos poucos estudos que existem nesta área, fazer uma abordagem a estrutura fonológica da Língua gestual portuguesa é um desafio complexo.

O problema coloca-se logo na terminologia a utilizar. Como sabemos, a palavra fonologia tem raízes no grego, sendo composta por “phonos” (=som/voz) e “logos” (conhecimento/palavra).

Visto as raízes e o que compõe a palavra fonologia, a primeira parte dessa composição poderá levantar algumas reservas quando aplicamos esse conhecimento a uma língua que não usa o som como entidade pertinente, seja no âmbito da produção ou da perceção.

Como é do conhecimento geral, é utilizado os gestos com as mãos para a produção da linguagem gestual e a visão para a perceção dessa linguagem. Um dos primeiros linguistas que se dedicou ao estudo da estrutura da ASL, Língua de Sinais Americana, William Stokoe, terá sido sensível a esta contradição terminológica, tendo proposto uma nomenclatura distinta para o ramo da linguística da língua gestual que se dedicava aos estudos dos segmentos da LGP.

William Stokoe viria a propor que os estudos das unidades discretas da LGP se pudessem chamar de quirologia, palavra que provem do grego “Khiros”, cujo significado é mão pela simples razão de ter considerado a mão como elemento principal e central da produção dos gestos. As unidades significantes distintivas da LGP seriam designadas por queremas.

Contudo essa posição não foi universalmente aceite e a terminologia foi descartada, continuando-se a usar o termo **Fonologia**. Esta opção justifica-se, segundo Amaral (1994, p. 59) “por uma questão de facilitação de terminologia e também para por em evidência o carácter linguístico deste sistema passou-se a adotar a terminologia da linguística para o estudo da língua gestual”.

Feita esta introdução sobre os critérios de terminologia, mais relevante é perceber a estrutura fonológica ou quirológica deste sistema linguístico.

William Stokoe, poderá ter sido dos primeiros linguistas a afirmar que o signo gestual não tinha um carácter holístico, isto é, não deveria ser encarado como um todo, mas sim, à semelhança do que se considera para as línguas orais, ou seja, constituído por partes discretas e arbitrárias que se combinam para formar a palavra (STOKOE, 1965 referido por Amaral M.A.; Coutinho, A.; Delgado Martins, M.R 1994.

Tentando entender esta ideia, vamos tentar demonstrar através de dois exemplos. Assim, em português, consideramos que a palavra “gato”, é composta por quatro fonemas /g/ /a/ /t/ /u/. Estas unidades mínimas podem combinar-se com outras e formar novas palavras. A ausência, presença ou alteração de uma delas é suficiente para a diferença entre diversos vocábulos. Se mudar o primeiro fonema desta palavra por outro, por exemplo, /p/ obtenho o vocábulo “pato”, que se distancia do anterior pela mudança de um único segmento.

Em LGP o gesto é também composto por diversas unidades discretas que determinam o significado global do gesto, são elas, de acordo com o esquema proposto por Stokoe, a localização, o movimento e a configuração da mão. Assim, se mudássemos qualquer uma delas obteríamos uma palavra diferente o que se constata, por exemplo, nos pares CINCO e SAPATO representados pela mesma configuração de mão, partilhando a localização espaço-corporal, mas alterando o parâmetro movimento, ausente no gesto CINCO, presente no gesto SAPATO. Trabalhos posteriores a Stokoe, como os de Battison (1974, 1978, referido por Sandler e Lillo-Martin, 2006) adicionaram ainda dois outros parâmetros, considerados essenciais na estrutura da LGP, a orientação da mão e os aspetos não - manuais, como a expressão facial. O que nos parece importante nestas investigações, e em todas as que se têm vindo a desenvolver neste âmbito é que eles dão conta de um Universal Linguístico que caracteriza a LGP, o da dupla articulação.

Tal como na Língua Portuguesa, também na LGP se encontram dois níveis: o do significado, constituído pela imagem mental dos conceitos e o do significante de que fazem parte as unidades mínimas com carácter distintivo.

1.7 Parâmetros Fundamentais da LGP - Configuração da mão (ou mãos)

(Baltazar, 2012) De acordo com Ana Bela Baltazar, a mão ou as mãos, podem tomar múltiplas configurações na produção da LGP, sendo possível que a configuração da mão não dominante assuma o papel do local de articulação, em sinais produzidos com as duas mãos.

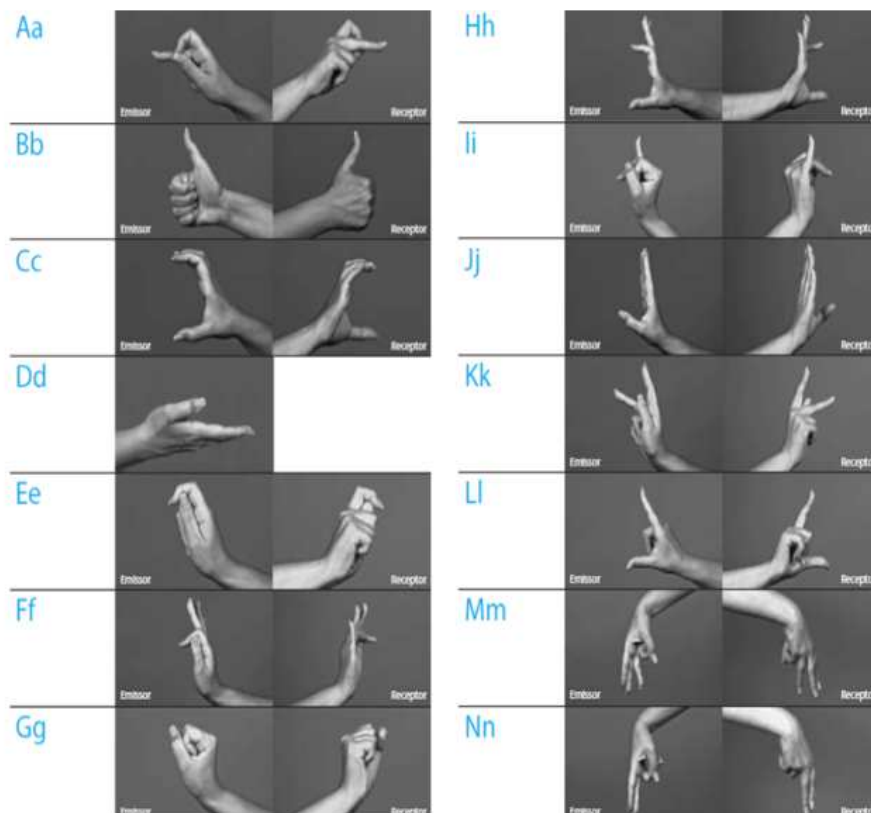


Figura 2 - Configuração Manual, Alfabeto (Aa à Nn), Fonte: (Baltazar, 2012)

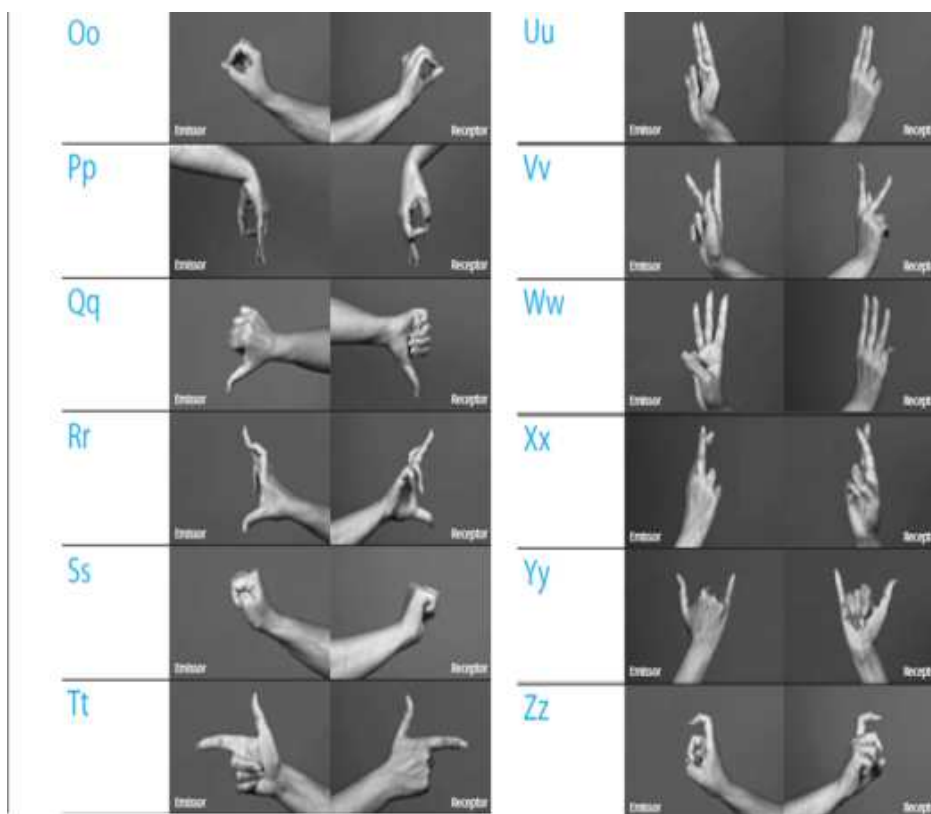


Figura 3 - Configuração Manual, Alfabeto (Oo à Zz), Fonte: (Baltazar, 2012)

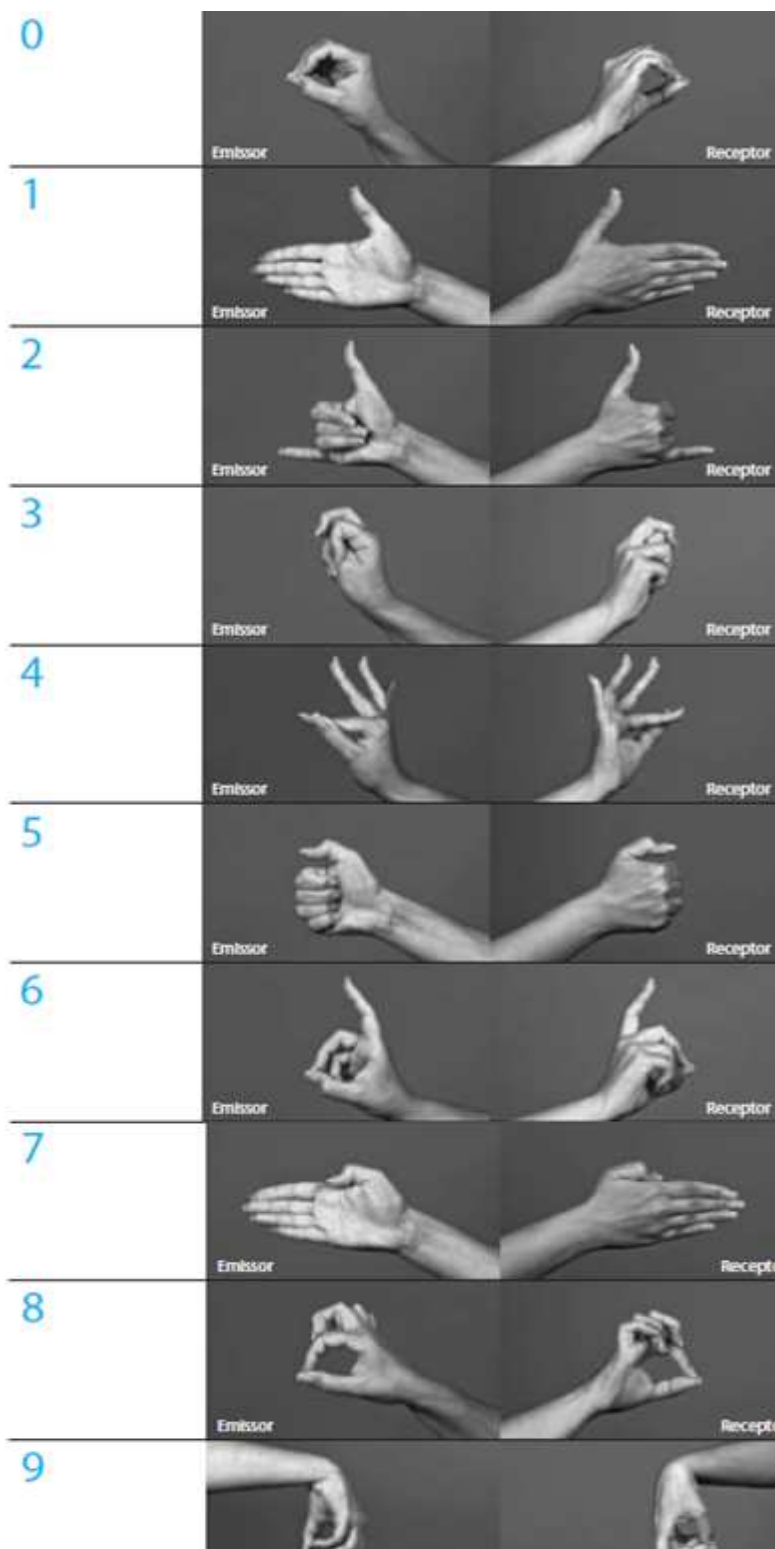


Figura 4 - Configuração Manual, Alfabeto (0 à 9), Fonte: (Baltazar, 2012)



Figura 5 - Palavras Simples, Fonte: (Baltazar, 2012)

1.8 Parâmetros Fundamentais da - LGP Local de articulação (ou localização)

(Baltazar, 2012) Ainda de acordo com a Doutora Ana Bela Baltazar, relativamente a LGP poderemos indicar três grandes espaços articulatórios:

- Espaço que abrange várias partes do corpo;
- Espaço mais ou menos próximos do corpo;
- Espaço de articulação na mão dominante.

Sigla	Significado	Sigla	Significado
CC	Cimo da Cabeça	QU	Queixo
TS	Testa	PE	Pescoço
TE	Têmporas	PM	Ombro
OL	Olho	ES	Esterno
NA	Nariz	MT	Meio do tronco
BH	Bochecha	AB	Abdómen
OR	Orelha	BR	Braço
BO	Boca	NA	Antebraço
LA	Lábios	PR	Perna

Tabela 2 - Locais de Articulação

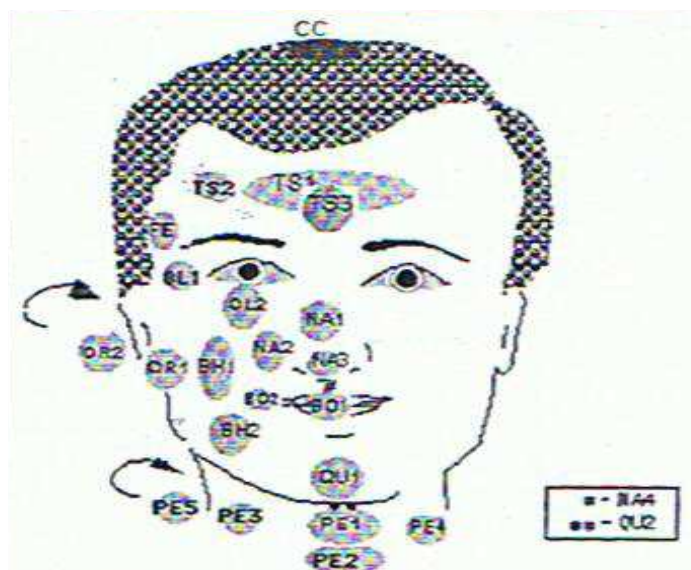


Figura 6 - Local de articulação dos gestos na cabeça e pescoço do praticante de Gestos da LGP - Fonte: (Baltazar, 2012)

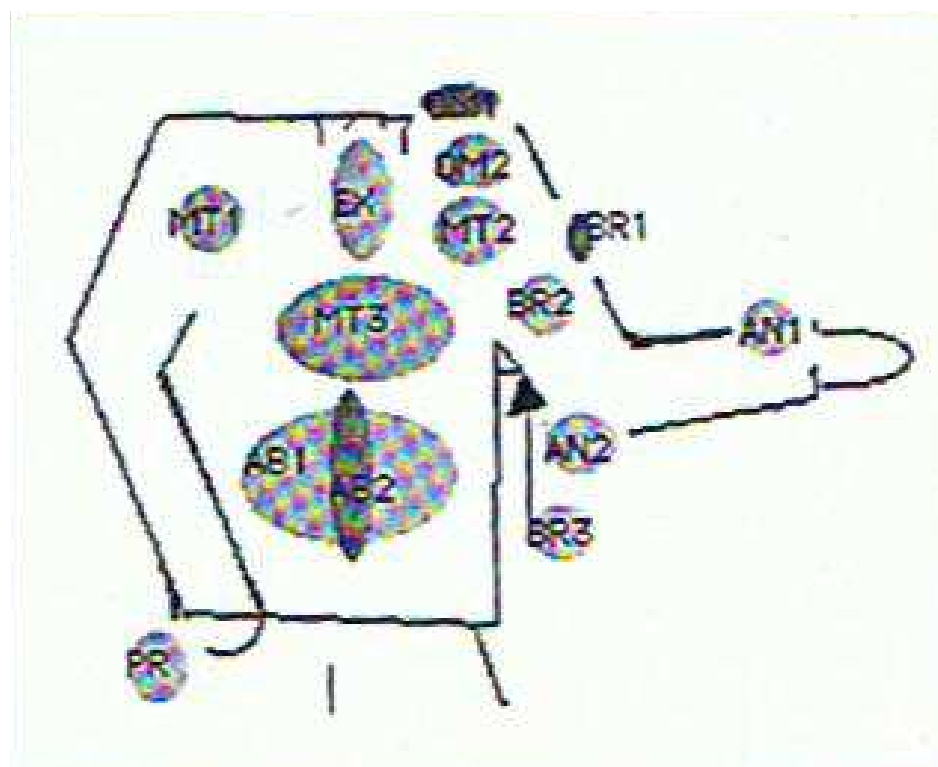


Figura 7 - Áreas de articulação dos gestos no tronco e nos membros do praticante de gestos da LGP- Fonte: (Baltazar, 2012)

Pontos de contacto	Símbolos	Exemplos	
CC cima da cabeça	CC	GALO	
TS testa	— toda a testa	TS1	VACA
	— lado d. ^a	TS2	BOI
	— meio	TS3	INDIANO
TE têmporas	TE	BURRO	
OL olho	— canto	OL1	QUINTA-FEIRA
	— parte inf.	OL2	VER
NA nariz	— lado d. ^a	NA1	MENTIRA
	— cana	NA2	GALINHA
	— ponta	NA3	NAMORADA
	— parte inf.	NA4	CHI-CHI
BH bochecha	— meio	BH1	VINHO
	— inf. dra	BH2	REBUÇADO
OR orelha	— lóbulo	OR1	SURDO
	— atrás	OR2	RETROAURICULAR
BO boca	— os 2 lábios	BO1	SOFA
	— canto	BO2	HOMEM
LA lábio inferior LA	CONTAR	Quadro 28	
QU queixo	— frente	QU1	CAFÉ
	— parte inf.	QU2	AVÓ
PE pescoço	— todo em cima	PE1	FOME
	— todo em baixo	PE2	GOSTAR
	— lado d. ^a	PE3	SEGUNDA-FEIRA
	— lado esq.	PE4	GARGANTA
	— atrás	PE5	NÃO-ME-APETECE
OM ombro	— em cima	OM1	CHEFE
	— frente	OM2	COIMBRA
ES esterno	ES	EU	
MT meio/tronco	— lado d. ^a	MT1	SEXTA-FEIRA
	— lado esq. ^a	MT2	ESPAÑA
	— centro	MT2	ZANGADO
AB abdómen	— linha cen.	AB1	OPERAÇÃO-À-BARRIGA
	— meio	AB2	VOMITAR
BR braço	— face externa	BR1	RAPAZ
	— em cima	BR2	FORTE
	— em baixo	BR3	FRACO
AN antebraço	— pulso	AN1	DOENTE
	— cotovelo	AN2	ÁRVORE
PR perna	PR	PERNA	

Figura 8 - Exemplos de gestos realizados no corpo do praticante de Gestos de Língua gestual Portuguesa, Fonte: (Baltazar, 2012)

1.9 Parâmetros Fundamentais da LGP - Orientação

(Baltazar, 2012) A Doutora Ana Baltazar, também defende que na Língua Gestual Portuguesa, a imagem do que frisaram Liddel e Johnson (1989), a orientação da mão é muito importante, quer no contraste léxico, quer no funcionamento morfológico.



Figura 9 - Orientações da mão. Fonte: (Baltazar, 2012)

Movimento (a direito, para cima, para baixo, circular)



Figura 10 - Demonstração de Orientação da mão num gesto, Fonte:(Baltazar, 2012)

TIPO	DIRECÇÃO	EXEMPLO
Movimento a direito	Horizontal	Semana
Movimento a direito	Vertical	Pessoa
Movimento a direito	Superfície	Chocolate
Movimento a direito	Médio	Perigo
Movimento a direito	Obliquo	Segunda-feira
Movimento indirecto	Horizontal em círculo	Nós
Movimento indirecto	Horizontal em arco	Perguntar
Movimento indirecto	Vertical em círculo	Bola
Movimento indirecto	Em sete	Café
Movimento local	Dedilhar	Futebol de cinco
Movimento local	Esganchar	Associação
Movimento local	Torcer	Árvore
Movimento local	Friccionar	Homem

Figura 11 - Movimentos essenciais na quirologia da LGP, Fonte: (Baltazar, 2012)

1.10 Parâmetros Fundamentais da LGP - Componentes não - manuais (expressão facial e/ou corporal)

A atividade não manual na produção da LGP é muito importante. Pode possuir função morfológica, sintática, lexical, semântica e prosódica, bem como podem ser consideradas como elementos fonológicos.

Trata-se das alterações das posições do corpo, alteração da expressão facial e da posição da cabeça, para além das configurações da boca, as emissões vocais sem relação com os sons da fala que acompanham certos gestos.










	bochecha cheia
	bochecha vazia
	boca cerrada
	expressão facial alegre e aberta
	expressão facial triste
	olhar cerrado/expressão facial carregada
	olhar cerrado/expressão calma
	olhar aberto/expressão exclamativa, de admiração ou surpresa
	olhar aberto/expressão de certeza
	expressão interrogativa ou de dúvida

Figura 12 - Exemplos de Configuração não Manuais, Fonte:(Baltazar, 2012)

1.11 Expressão

O termo “Expressão”, (Correia, 2009), pelo seu uso, se encontra muitas vezes afastado do significado linguístico que pode comportar. Geralmente, esta palavra associa-se à postura corporal, ao maior ou menor uso de linguagem gestual, à expressividade física das emoções, não sendo, à primeira vista, reconhecido o seu valor enquanto unidade pertinente de um sistema linguístico.

Uma das razões que poderá provocar isso será pelo facto de se perspetivar a língua como manifestação oral. Uma correta postura, um uso adequado de gestos expressivos e auxiliares do discurso são fatores que contribuem para uma mais eficaz comunicação oral. Além disso, a expressividade com contornos linguísticos, isto é, enquanto elemento pertinente na formação de sentidos, está reservada ao uso da entoação.

Assim, a nossa curva entoacional varia consoante queiramos exprimir interrogações, exclamações ou até manifestar dúvidas, certezas e outras reações inerentes ao discurso. Por exemplo, a frase “hoje, vamos jantar fora” pode constituir-se como interrogação, exclamação dependendo apenas da curva entoacional que a define.

Na LGP a expressão, adquire funções semelhantes às que acabámos de descrever para a entoação. Estudiosos como Liddell (1986) consideram a expressão apenas como uma reflexão da sintaxe. Para estes linguistas, o uso da expressão facial marcaria o fim de uma frase e o seu tipo.

Na Língua Portuguesa, a entoação depende exclusivamente das cordas vocais, na medida em que estas são responsáveis pela variação do tom de voz. Na LGP, ao contrário do que acontece nas línguas orais, as variações suprasegmentais relacionam-se com vários articuladores, como as sobrancelhas, as pestanas, as faces e os lábios, sendo que podem ocorrer em simultâneo ou independentemente, desempenhando uma ou várias funções.

Não se trata apenas de mera reflexão das emoções do indivíduo ou de auxiliar de comunicação, mas contribui efetivamente para a compreensão da globalidade do texto produzido. Há diversas maneiras, mais ou menos enfáticas, de expressar tristeza, zanga, descontentamento, entre outras emoções ou estados, mas isso deve ser tratado enquanto elemento para linguístico.

O uso da expressão para representar uma pergunta ou uma admiração independente da “expressividade” natural do emissor ou do seu estado de espírito. Sandler e Lillo-Martin (2006), constataam que “signers use facial expression in both non-linguistic and linguistic ways, another parallel with intonation in spoken languages”).

Em português a frase afirmativa, “passei no exame” pode ser dita com maior ou menor efusividade, como neutralidade ou, dependendo do falante e do contexto, com algum descontentamento. O mesmo se passa em LGP para afirmação idêntica. Todavia, não se esperaria que o falante português dissesse essa frase acompanhada de uma curva entoacional

ascendente, típica das estruturas interrogativas, sendo necessário que usasse a frequência correspondente às frases declarativas, independentemente do seu estado de espírito. O mesmo acontece em LGP. Quando o gestuante pretende afirmar algo não é expectável que acompanhe o discurso com movimentos típicos de interrogações como o arquear das sobrancelhas. Assim, a expressão facial assume-se enquanto unidade prosódica codificada contribuindo para o ritmo significativo do enunciado.

1.12 Língua Gestual no Mundo

Para explicar a Língua Gestual no Mundo é importante fazer referência as primeiras perguntas, portanto é importante estudarmos um pouco da história da Língua Gestual e da história dos Surdos. Referencias feitas de acordo com (PorSinal).

Nos anos **368**, o filósofo grego Sócrates, fez esta pergunta ao seu discípulo, Hermógenes, “Suponha que nós não tenhamos voz ou língua, e queiramos indicar objetos um ao outro. Não deveríamos nós, como os surdos-mudos, fazer sinais com as mãos, a cabeça e o resto do corpo?” – Ao que Hermógenes respondeu “Como poderia ser de outra maneira, Sócrates?”

A história dos surdos-mudos começa com muitas restrições que lhes foram impostas. No ano **534**, o Codex Justinianus, do Imperador Bizantino Justiniano I (527-565), vai influenciar as novas nações latinas negando os direitos civis aos surdos congénitos, como o casamento e a propriedade. Estas imposições mudariam ao longo do tempo, já que no ano 1500 o médico Girolamo Cardano, médico e filósofo, reconhecia a habilidade do surdo para a razão, afirmava que “... A surdez e mudez não é impedimento para aprender e o meio melhor é através da escrita... e é um crime não instituir um surdo-mudo.” Este médico já utilizava nesta altura a língua de sinais e escrita com os surdos.

Um dos primeiros alfabetos manuais foi desenvolvido em **1520** por um monge espanhol, Pedro Ponce de León, ao trabalhar com alunos surdos. Permitia que o estudante aprendesse a soletrar (letra por letra) toda a palavra. Não se sabe se este alfabeto foi baseado, unicamente ou parcialmente, em gestos com as duas mãos, diferente do alfabeto uni manual publicado por Juan de Pablo Bonet em **1620**.

A primeira escola para surdos, foi criada, por Abbe' de l'Épée ou Charles Michel de l'Épée, em **1762**, em Paris – França, tinha como objetivos que em França todos os surdos-mudos aprendessem a ler e a escrever. Considerava insuficiente a linguagem natural dos surdos, que tinha aprendido e desenvolveu o método para a língua gestual, o “signes méthodiques” para integrar a gramática da língua francesa. Este método serviu como base para a língua de sinais americana bem como outras línguas de sinais mundiais. Viria a publicar “Instruction de Sourds et Muets par la Voix des Signes Méthodiques”. Abbe' de l'Épée viria a envolver-se numa polémica epistolar sobre os respetivos métodos de ensino com Samuel Heinike, que dirigiu a primeira escola de ensino exclusivamente oral para surdos na Alemanha. Desta polémica, saíam reconhecidos os métodos francês e alemão. Ainda antes da sua morte, Abbe' de l'Épée, fundou 21 escolas para surdos na França e na Europa.

Em **1817**, Gallaudet e Laurent Clerc, abrem em Hartford, Connecticut, o “Asylum for the Education and Instruction of Deaf and Dumb Persons (American School for the Deaf)”, primeira escola permanente nos Estados Unidos da América. Começam por ensinar uma mescla de LSF (*langue des signes française*), francês por gestos e de inglês e acabaram por optar pela ASL (*American sign language*).

Em **1823**, em Portugal, é fundado o Instituto de Surdos-Mudos e Cegos, e, por decisão do Rei D. João VI, chamado o especialista sueco **Pär Aron Borg**, para o orientar. A abertura oficial da educação para surdos-mudos e cegos na Luz foi em 1 de Março de 1824, após portaria de 18 de Fevereiro de 1824 do Intendente Geral de Polícia, mais tarde publicada no Diário do Governo de 16 de Dezembro de 1847. Pär Aron Borg teve um papel muito importante no ensino dos surdos em Portugal, que com ele aprenderam a comunicar através de um alfabeto manual e também da língua gestual de origem sueca. O método de ensino do Instituto de Estocolmo foi adaptado para o ensino em Portugal. O alfabeto manual inventado por Borg foi adotado pela Suécia, Portugal e Finlândia.

Em **1856**, **J. S. Brown** publica "*A Vocabulary of Mute Signs*" e marca nos EUA o início de um estudo da língua gestual que só será retomado mais de cem anos depois. Em França, Pélisser, professor e poeta surdo publica a "*Iconographie des Signes*".

No processo do estudo e do entendimento dos surdos, em **1910**, os psicólogos **Alfred Binet** e **Theodore Simon** publicam um estudo sobre o valor do método oral onde afirmam que, no caso dos surdos, o pensamento se desenvolve antes da linguagem.

Em **1956**, Amílcar Castelo apresenta o primeiro trabalho, em Portugal, de avaliação da inteligência da criança surda. Os resultados do teste do labirinto de Rey apresentam 50% dos casos abaixo do limiar normal e as **matrizes de Raven** atribuem deficiência mental a 60%. A causa estaria na perda ou atraso da linguagem, deficit que implica a forma abstrata do pensamento para concluir que a criança surda não é deficiente mental e que os testes de performance são o meio mais adequado de avaliar a inteligência do surdo.

Em **1958**, em Lisboa é fundada oficialmente a **Associação Portuguesa de Surdos** (APS), por despacho do Ministro da Saúde e Assistência. Em Manchester, o Congresso Internacional sobre o Moderno Tratamento Educativo da Surdez vem pôr fim ao longo monopólio do método oral puro na maior parte dos países europeus. E consagra a sua renovação, o método materno-reflexivo do holandês Van Uden.

Em **1960**, **William Stokoe** (1920-2000), diretor do laboratório de pesquisas linguísticas do Gallaudet College desenvolve o conceito de querema como equivalente gestual do fonema e publica *Sign Language Structure*. Com este trabalho inicia-se o reconhecimento da ASL como língua genuína com uma estrutura complexa que ultrapassa em muito uma imitação rudimentar do discurso oral, como quase todos pareciam acreditar.

Em **1961**, nos EUA, o trabalho de Stephen Patrick Quigley, **D. Robert Frisina**, depois de comparar o rendimento de crianças surdas filhas de pais surdos e de pais ouvintes, encontrou resultados superiores em vocabulário e soletrar dactílico bem como em rendimento educativo nos primeiros, enquanto os segundos eram superiores na inteligibilidade da fala.

Em **1964**, Robert Weitbrecht, surdo, cria o telefone de texto (TTY), um aparelho que permite aos surdos comunicar por escrito através de uma linha telefónica vulgar. O Paget-Gorman Sign System é utilizado pela primeira vez numa experiência de educação de surdos adultos. Trata-se de um sistema gestual artificial que respeita todo o sistema gramatical do inglês. Mais tarde foi usado com crianças surdas, deficientes mentais e motoras. A Zenith fabrica e comercializa a primeira prótese com um circuito integrado. Doyle relata a inserção de um feixe de quatro eletródios na cóclea de um paciente com surdez neuro-sensorial total. Os resultados foram considerados satisfatórios, o paciente conseguia repetir frases. Furth publica *Thinking Without Language*, onde afirma que os surdos podem compreender e aplicar conceitos tão logicamente como as pessoas ouvintes. Os resultados inferiores em certas provas cognitivas devem-se tanto à falta de experiência do mundo como das próprias condições dessas provas que favorecem um background de domínio da linguagem oral. É criado o Instituto de Surdos de Bencanta (Coimbra).

Em **1965**, Surge o Cued-Speech ou Palavra Complementada, técnica criada por **R. Orin Cornett**, então vice-presidente do então Gallaudet College (agora Gallaudet University). É um sistema de apoio à leitura lábio facial que elimina as confusões e pretende tornar totalmente inteligível o discurso falado. Os gestos associados são totalmente desprovidos de significado e destinam-se a esclarecer a informação presente nos lábios, o que faz do CS um sistema oral. Downs, Starrit e Squires, nos EUA, são os primeiros a propor e realizar o teste acústico neonatal para despistagem da deficiência auditiva. A Direcção-Geral da Assistência integra os Estabelecimentos de Educação Especial e cria dois centros para assegurar o seu funcionamento: o COOMP (Centro de Observação e Acompanhamento Médico-Pedagógico), para selecionar as crianças e o CFAP (Centro de Formação e Aperfeiçoamento de Pessoal), para a formação especializada do pessoal desses estabelecimentos. É criado o Instituto de Surdos do Funchal.

Em **1970**, McCay Vernon e Soon D. Koh, em "*Early Manual Communication and Deaf Children's Achievement*", concluem que os filhos de pais surdos são significativamente superiores em rendimento académico, leitura, vocabulário e escrita aos surdos filhos de pais ouvintes. Mas, na sua maioria, esses valores são muito inferiores aos valores estabelecidos para os alunos normoouvintes; Jewett descreve pela primeira vez o exame no homem dos potenciais evocados auditivos do tronco cerebral (PEATC), método que veio tornar possível a avaliação objetiva da surdez em qualquer idade e que é ainda o melhor método disponível.

Em **1972**, é eleito o primeiro presidente surdo da Associação Portuguesa de Surdos, Fernando **Pinto**; Em Portugal, são criadas pelo ME as Equipas de Ensino Especial Básico e Secundário. Os circuitos integrados permitem o fabrico de próteses com processos de compressão e a redução dos ruídos do ambiente.

Em **1973**, é criada em Lisboa a **APECDA, Associação de Pais para a Educação de Crianças Deficientes Auditivas**, que vai introduzir em Portugal o método verbo-tonal, de reabilitação áudio - oral; Um grupo interdisciplinar, constituído por linguistas, terapeutas da fala, engenheiros do laboratório de Fonética da Universidade de Letras de Lisboa, em colaboração com a Associação de Surdos de Lisboa, realizam o 1º estudo sobre a Língua Gestual Portuguesa; A Divisão do Ensino Especial da DGEB/ME apresenta o Programa de Compensação Educativa. Estima-se que têm atendimento 80% dos surdos em idade escolar. O normal é que o ensino primário seja cumprido nos estabelecimentos especiais, sob a tutela do MAS, durante oito anos. Depois são maioritariamente encaminhados para oficinas de treino profissionais. Há 50 surdos profundos a frequentar o ciclo preparatório (com uma média de idade elevada). O programa propõe-se integrar mil crianças com qualquer grau de deficiência no primário e cem no 2º ciclo num período de seis anos. Para tal propõe-se também especializar um professor por cada 20 alunos (primário) ou por cada dez (ciclo preparatório).

Em **1974**, A renovação pedagógica em Portugal leva as Divisões do Ensino Especial da Direção Geral do Ensino Básico (DGEB) e da **Direção Geral de Ensino Superior** (DGES), com apoio do Programa de Cooperação Luso - Sueco a fomentar a formação de docentes e técnicos, o desenvolvimento tecnológico e a investigação na educação dos surdos; A Associação Portuguesa de Surdos promove cursos de alfabetização para jovens e adultos, baseados na Comunicação Total; É criada a Delegação do Porto da Associação Portuguesa de Surdos.

Em **1975**, na Suécia, **Inger Ahlgren** inicia uma experiência envolvendo famílias de pais surdos e ouvintes de crianças surdas de utilização da língua gestual como primeira língua dessas crianças; Primeiro aprovada pelo Congresso Americano em 1975 como Education of All Handicapped Children Act (Lei Pública 94-142), a legislação IDEA exigia que todos os alunos com deficiência até 21 anos de idade deviam ser providos com a educação pública adequada num ambiente menos restritivo e com acomodações.

Em **1979**, **Ursula Bellugi** e Edward Klima publicam "*The Signs of Language*". Julia Maestas Moores estuda a comunicação gestual de crianças surdas filhas de pais surdos e encontra as mesmas sequências de desenvolvimento, interação e de estádios linguísticos que se encontram nas crianças ouvintes; Suzanne Borel-Maisonny, ortofonista, promove a primeira experiência pedagógica baseada no bilinguismo, numa classe com duas professoras, uma ouvinte e outra surda, apesar de tal ser, ao tempo, ilegal em França; Em Varna, Bulgária, o 8º Congresso da WFD reclama ainda a melhoria dos métodos de ensino, melhores meios para o desenvolvimento da capacidades intelectual e da expressão oral das crianças privadas da audição; É criada a **APECDA-Porto**.

Em **1981**, Criação do **SignWriting**, um sistema de escrita das línguas gestuais; Dois surdos portugueses deslocam-se à Universidade de Gallaudet, visando um melhor aperfeiçoamento da Língua Gestual.

Em **1982**, **Danielle Bouvet**, em **La Parole de l'Enfant Sourd**, propõe uma metodologia de ensino bilingue, em que a LG é considerada como língua materna e a língua dos ouvintes como segunda língua; Nos EUA, o acesso a um intérprete foi negado pelo sistema escolar por causa da aluna, Amy Rowley, que era capaz de ler lábios. No primeiro caso sob a égide da lei IDEA, o Supremo Tribunal declarou que a escola não estava obrigada a providenciar um intérprete, porque ela "alcançava a realização educacional, acadêmica e social, sem essa ajuda"; O Secretariado Nacional de Reabilitação organiza o primeiro curso para intérpretes de LGP, orientado pelos monitores José Bettencourt e João Alberto Ferreira.

Em **1983**, primeira experiência portuguesa de bilinguismo verbal-gestual, promovida por **Sérgio Niza**, na escola de A-da-Beja; O governo sueco reconhece oficialmente a língua gestual como a língua nativa dos surdos suecos.

Em **1987**, o **Parlamento Europeu** apela aos estados membros para que reconheçam as línguas gestuais e para que estas façam parte de educação das crianças surdas; A Nicolet constrói a primeira prótese BTE digital. Só foram fabricados três protótipos; Marlee Matlin é a primeira atriz surda a ganhar um óscar pelo seu papel em "Filhos de um Deus menor".

Em **1989**, A **Lei 9/89, Lei de Bases da Prevenção e da Reabilitação e Integração das Pessoas com Deficiência** estabelece que a pessoa surda é uma pessoa com deficiência e, como tal goza do direito à reabilitação. A Associação Portuguesa de Surdos com o apoio do programa Helios, cria o 1º curso de Intérpretes de Língua Gestual Portuguesa.

Em **1990**, em Portugal, o Decreto-Lei 35/90 define a obrigatoriedade da escolaridade básica para os alunos com necessidades educativas especiais.

Em **1991**, o ministério da educação dinamarquês estabelece a possibilidade de todas as crianças surdas poderem estudar a língua gestual nas escolas como primeira língua. Em França, a lei estabelece pela primeira vez como um direito a liberdade de escolha entre uma educação bilingue ou uma educação oral; Em Portugal, o Decreto-Lei 319/91 vem instituir o Regime Educativo Especial, definindo uma orientação para todos os deficientes, que conduziu à integração generalizada dos surdos nas escolas regulares da sua área de residência; É fundada, em Lisboa, a FEPEDA, Federação Europeia **de Pais das Crianças Deficientes Auditivas**; Foi fundada, em Lisboa, a Associação de Intérpretes de Língua Gestual Portuguesa - AILGP.

Em **1992**, primeira edição do **Gestuírio de Língua Gestual Portuguesa**, do Secretariado Nacional da Reabilitação, por protocolo com a DGEB, trabalho coordenado por António Vieira Ferreira; Seis finalistas portugueses do curso de Intérpretes e seis formadores Surdos vão à Universidade de Bristol receber formação; Na Bélgica, o Ministério da Promoção Social nomeia como peritos os primeiros professores surdos e responsabiliza-os pela organização do primeiro curso em língua gestual reconhecido oficialmente com o nível secundário. Na continuidade, foi organizado a formação, de professores surdos, com a atribuição de diploma de regentes.

Em **1993**, nos EUA é aprovado o **Individuals with Disabilities Education Act** (IDEA). O Departamento da Educação estabelece a Política de Inclusão, que garante a todas as crianças com incapacidade o direito a frequentar as escolas regulares locais. As escolas residenciais para surdos são consideradas meios muito restritivos e muitas fecham por falta de alunos. Mas o número de professores surdos aumenta 16%, com maiores oportunidades no ensino público; Intérpretes e Formadores Surdos entram nas escolas. A Língua Gestual Portuguesa começa a ganhar o espaço que lhe é devido; No 1º Congresso Nacional de Surdos, reunido em Coimbra é aprovada a Carta Social da Pessoa Surda, onde se reclama que «seja reconhecido à pessoa surda o verdadeiro direito à igualdade, mantendo o direito de ser diferente, ou diferente mas igual».

Em **1994**, a Conferência Mundial sobre Necessidades Educativas Especiais adota a Declaração de Salamanca, que afirma a «importância da língua gestual como meio de comunicação entre os surdos (...) deverá ser reconhecida e garantir-se-á que os surdos tenham acesso à educação na língua gestual do seu país»; Em Portugal, a profissão de intérprete de LGP passa a constar da tabela nacional de profissões do **Instituto de Emprego e Formação Profissional** (IEFP).

Em **1996**, a **Resolução 48 da ONU** acentua a necessidade de prever a utilização da língua gestual na educação, no seio das famílias e das próprias comunidades e garantir a presença de intérpretes como mediadores da comunicação. Na norma 6 chama a atenção para a especificidade da educação das crianças surdas, que pode aconselhar escolas especiais e classes ou unidades especializadas em estabelecimentos regulares; A Widex lança no mercado a primeira prótese ITE totalmente digital e a Oticon apresenta uma BTE capaz de separar os sons em sete bandas tonais e que utiliza dois processadores de fala, um para as vogais e outro para as consoantes.

Em **1997**, a Língua Gestual Portuguesa foi reconhecida como língua oficial para os Surdos portugueses. A Lei 1/97 introduz na Constituição a incumbência de o estado proteger e valorizar a LGP como expressão cultural e instrumento de acesso à educação e da igualdade de direitos das pessoas surdas. António Barreto, nos jornais, chamou-lhe a constitucionalização do manguito; É criado o curso superior de Língua Gestual Portuguesa em Setúbal.

Em **1998**, em resolução de 17 de Junho, o **Parlamento Europeu** reconhece as línguas gestuais como direito das pessoas surdas, apela ao seu reconhecimento pelos estados membros e à formação de intérpretes de Língua Gestual através dos programas de emprego; Em França, o Congresso da FNSF propõe que a sigla LS passe a significar *langue sourde* em substituição de *langue des signes*; Em Portugal, o despacho n.º 7520/98 da Secretaria de Estado da Educação e da Inovação reconhece a necessidade de um ambiente escolar bilingue e define as condições para a criação e funcionamento das unidades de apoio à educação de crianças e jovens surdos em estabelecimentos públicos do ensino básico e secundário e da organização da competente resposta educativa.

Em 1999, é criado o curso superior de Língua Gestual Portuguesa no Instituto Politécnico do Porto; A Assembleia da República Portuguesa aprova a Lei 89, que define as condições de acesso e exercício da atividade de intérprete de língua gestual.

Em **2004, a lei IDEA/IDEA-2004** garante serviços para estudantes com deficiência. Os tópicos abordados incluem avaliação, desenvolvimento IEP (programa de educação individualizado) e reunião de equipa, disciplina e serviços de apoio pedagógico, ajudando a garantir a responsabilização, equidade e excelência em educação para crianças com deficiência.

2 Estado da Arte de Reconhecimento de Gestos

2.1 Introdução

A pesquisa no reconhecimento de línguas de sinais começou em 1982. Neste sentido há duas grandes abordagens para o reconhecimento distinguidas por diferentes perspectivas:

1 – Abordagem de Luvas de Dados;

2 – Abordagem baseada em Visão.

Destas duas abordagens há mais vantagens na segunda, do que na primeira. O maior problema da abordagem baseado em luvas de dados é ser desconfortável na sua utilização. Enquanto as abordagens baseadas em visão tem pelo menos duas grandes vantagens, sendo a primeira a de ser muito conveniente aos utilizadores pelo facto de não ser preciso a utilização de nenhum acessório. A segunda é a de que a comunicação por gestos utilizando uma câmara, pode oferecer ao surdo-mudo a liberdade de comunicação em sua linguagem gestual natal. Por causa disso essa abordagem é extremamente adequada no seu dia-a-dia. Nos dias de hoje, o reconhecimento da língua gestual inclui principalmente os algoritmos, Hidden Markov Models (HMM) e Artificial Neural Networks (ANN), etc.

Os modelos escondidos de Markov são modelos de representação de distribuição de probabilidades com base em sequências de observações, por outras palavras são modelos não determinísticos com origens em eventos aleatórios não visível, daí ter o nome escondido, mas que pode ser observado através de outro processo não determinístico que produz a sequência de observação. Neste momento são utilizados em praticamente todos os sistemas de reconhecimento de voz, em várias aplicações de biologia molecular computacional, nos algoritmos de compressão de dados, reconhecimento de palavras manuscritas, verificação de assinatura, deteção de falhas em sistemas dinâmicos e também noutras áreas de inteligência artificial e reconhecimento de padrões.

Recentemente tem sido usado em aplicações para a visão, utilizando câmaras, tais como modelação de sequências de imagens e reconhecimento de objetos.

2.2 Luvas no Reconhecimento de Gestos da Mão -

Reconhecimento da Configuração Manual

Análise e reconhecimento de gestos com as mãos têm sido focados pela comunidade de pesquisa nas últimas três décadas. Essa atenção é motivada pela importância da aplicação prática de gestos com as mãos, principalmente no campo da realidade virtual, efetuar o reconhecimento da linguagem, reabilitação médica e animação por computador. A complexidade do problema do reconhecimento do gesto veio trazer diferentes abordagens. Uma de entre as soluções que foram estabelecidas como padrão, a baseada em luvas de dados, que usa luvas especialmente concebidas, com sensores do tipo variável, que resulta em números a registrar a flexão do dedo ou contactos entre eles em tempo real.

A abordagem mais imediata para reconhecer a configuração manual é através do uso de luvas de dados, já que sendo uma luva com vários sensores, normalmente dois por dedo, cujos dados permitem aferir uma dada configuração manual. Os grandes problemas das luvas de dados são a sua capacidade de resolver pequenas variações da configuração manual. Uma luva de dados convencional tipicamente consegue reconhecer configurações bastante diferentes entre si como uma mão fechada e aberta mas pequenas variações são de difícil aferição.

De tal forma que na animação de mãos em avatares na indústria cinematográfica é amplamente usado “motion capture” para todos os membros, exceto nas mãos em que ainda é usada a animação manual.

Outra estratégia para reconhecimento da configuração manual é através de sensores de profundidade onde a informação da profundidade dum certa configuração manual é associada a uma configuração manual previamente definida, podendo se assim perceber qual a configuração manual sem o uso de acessórios.

2.2.1 Luvas 5DT

(5DT, 2011) As luvas de dados 5DT, são otimizadas para o uso em ambientes de Ressonância Magnética, MRI. A luva em si não contém quaisquer peças magnéticas, comunicando-se com uma caixa de controlo, de 5 a 7 cm de distância via fibra ótica. A caixa de controlo se comunica com o computador através de um cabo para a porta serial (RS232 – independente da plataforma). Os modelos disponíveis contem 5 ou 14 sensores.

Esta é uma imagem que ilustra, quantos sensores estão disponíveis na luva 5DT utilizada, como estão disponibilizadas e a sua identificação com por ordem.

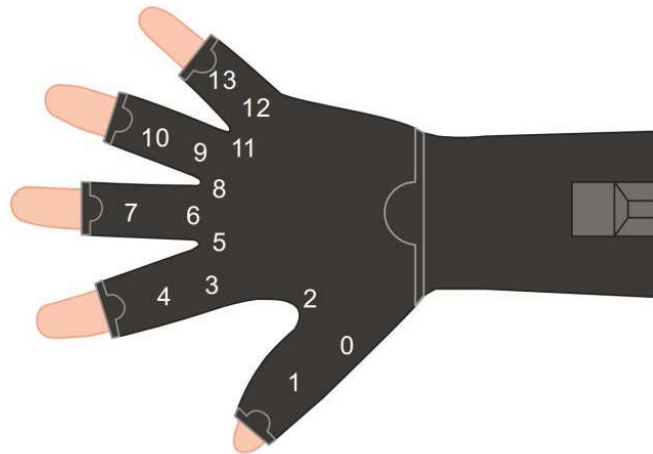


Figura 13 - Imagem da luva de dados 5DT (5DT, 5DT Data Glove Ultra Series, 2011)

2.2.2 Luvas CYBER

Desenhado para ser confortável e funcional como descrevem na documentação do produto, (CyberGlove), é uma luva que nos fornece com alta precisão até 22 sensores, um para cada junta da mão, de diferentes ângulos de medição.

Desenvolvido pela Cyber Glove Systems, a luva Cyber, utiliza a tecnologia curva resistiva de detecção responsável de transformar, com alta precisão, movimentos da mão e dos dedos em tempo real, dados digitais de cada sensor de cada junta.

2.2.3 Microsoft Digits

Criada em laboratório de pesquisa da Microsoft em Cambridge – UK, de acordo com (Microsoft'sDigits), com a ajuda dos investigadores da Universidade de Newcastle e da Universidade de Crete, a “Digits” é uma nova forma de interagir com o mundo digital, utilizando gestos intuitivos da mão. O “Digits” utiliza uma camara infravermelho, um gerador de linhas de raios infravermelhos, iluminação difusa através de raios infravermelhos e uma unidade de medida inercial, para saber as posições de um / cada dedo em relação a cada um dos outros dedos da mão.

Aplicações:

- Uma das aplicações para esta tecnologia, inclui interação espacial 3D com um instrumento móvel. Ao atribuir gestos para certa função, o utilizador pode controlar qualquer especto de um instrumento. Também permite que o utilizador não esteja com os olhos colocados no instrumento, ou seja, por exemplo:

- Imaginemos que tem um telemóvel no bolso e enquanto ouve música pretende aumentar o volume.

- Temos o exemplo do vídeo utilizado pela Microsoft onde eles atribuem o polegar para cima para controlar o recebimento de uma chamada num telemóvel, atendendo-a.

- Outra aplicação que esta luva da Microsoft poderá ter é o dos vários tipos de jogos que se encontram no mercado, dispensando o uso do teclado, rato ou outros dispositivos de controlo.

- Uma outra aplicação que o “Microsoft Digits” tem, como nos mostra no fim do vídeo de apresentação é ao do objetivo da tese, que é Reconhecimento de Linguagens Gestuais. Sendo que o dispositivo, tanto pode nos mostrar os sinais que fazemos em um modelo de mão no computador, mas também podemos caracterizar cada gesto do alfabeto da Língua Gestual com uma letra do alfabeto da Língua Portuguesa. Assim que essa classificação acabar, a pessoa pode fazer estes gestos e vai ser reconhecido e mostrado a letra correspondente.

O objetivo da construção da luva da Microsoft Digits, passa por substituir, o uso das luvas atuais que podem ser desconfortáveis e com melhor precisão e rapidez atingir os objetivos que pelas luvas atuais eram propostos.

2.3 Reconhecimento de Gestos Utilizando Visão

2.3.1 O Uso de sensores de profundidade

O desenvolvimento de um dispositivo de entrada de sensor de movimento da Microsoft, o Kinect, sensor de profundidade, foi uma revolução na tecnologia semelhante aos que moldaram as descobertas mais fundamentais do século 20. Embora este desenvolvimento pode parecer ampla e diversificada, que pode ser resumido simplesmente: pela primeira vez, os computadores podem ver.

A imagem que o sensor de profundidade Kinect pode dar é tão importante porque é muito mais fácil para um computador para interpretar do que uma imagem de cor convencional. Se tentar processar uma imagem plana 2D, pixéis com cores semelhantes que estão próximos um do outro pode não pertencer ao mesmo objeto. Se você tiver informações 3D, em seguida, os pixéis que correspondem a locais fisicamente próximos uns dos outros tendem a pertencer a um mesmo objeto, independentemente da sua cor. Tem sido dito muitas vezes que o reconhecimento de padrões tem sido feito artificialmente difícil porque a maioria dos sistemas dependem de dados 2D.

Com o uso de sensores de profundidade, tem provado ser um método de controlo mais robusto do que os métodos baseados em visão. Outra grande vantagem da Microsoft Kinect SDK é que ele retorna as posições e orientações da mão facilmente com uma boa precisão.

2.3.2 Microsoft Kinect

Criada inicialmente para jogos, para concorrer com a Wii e com a Move da Sony para o Playstation 3, o Kinect é um avanço nas recentes câmaras 3D de profundidade. Os sensores da Microsoft Kinect vieram criar muitas oportunidades para a computação multimédia. Foi construído para revolucionar a forma como as pessoas jogam e mudar a experiência do entretenimento.

Com o Kinect, as pessoas são capazes de interagir com os jogos com o próprio corpo de uma forma natural, sendo a chave para a tecnologia a compreensão da linguagem de corpo humano. O computador primeiro terá que entender o que o utilizador está a fazer antes de conseguir dar uma resposta. Este foi sempre um objetivo na investigação no campo da visão humana mas essa ideia tal como já foi provada é extremamente difícil com câmaras de vídeo.

Como funciona?

O Kinect utiliza luz estruturada e máquina de aprendizagem.

Inferir a posição do corpo é um processo de duas etapas onde primeiro calcula um mapa de profundidade utilizando a luz estruturada que emite e depois infere a posição do corpo utilizando a máquina de aprendizagem, obtendo bons resultados.

O sistema utiliza muitos conceitos de matemática de nível superior e demonstra os avanços notáveis feitos em visão por computador nos últimos 20 anos.

2.3.3 Mapa de profundidade

Na primeira etapa, o mapa de profundidade é construída por meio da análise de um padrão salpico de luz laser infravermelho (MacCormick, 2010).



Figura 14 – Elementos constituintes da Microsoft Kinect

Importa dizer que a Microsoft licenciou esta tecnologia através de uma empresa chamada PrimeSense. O cálculo de profundidade é toda feita pelo hardware PrimeSense construído em Kinect. Os detalhes não estão disponíveis publicamente, esta descrição é especulativa (baseada principalmente em pedidos de patentes PrimeSense) e poderá estar errado.

Técnica de análise de padrão conhecido.

Esta técnica chamada de luz estruturada tem por princípio geral, projetar um padrão conhecido para uma cena inferindo a profundidade na deformação do padrão.

O Kinect combina a luz estruturada com duas técnicas clássicas de visão por computador:

- Profundidade de foco;
- Profundidade de som.



Figura 15 – Representação da técnica de análise de padrão conhecido, Zhang et al, 3DPVT (2002)

2.3.4 Kinect – Profundidade de foco

A profundidade de foco utiliza o princípio de que o material que está mais borrado é o material que se encontra mais longe, nesse sentido o Kinect melhora em muito a precisão da profundidade do foco tradicional. (MacCormick, 2010)

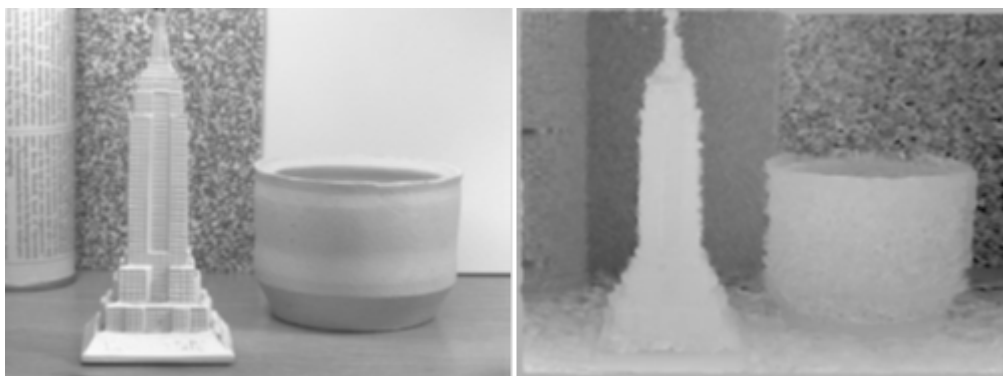


Figura 16 – Profundidade de Foco, Watanabe and Nayar, IJCV 27 (3), 1998

Usando uma lente especial, “astigmatismo”, com diferentes distâncias focais nas direções X e Y, sendo assim um círculo projetado se torna numa elipse cuja orientação depende da profundidade. (MacCormick, 2010)

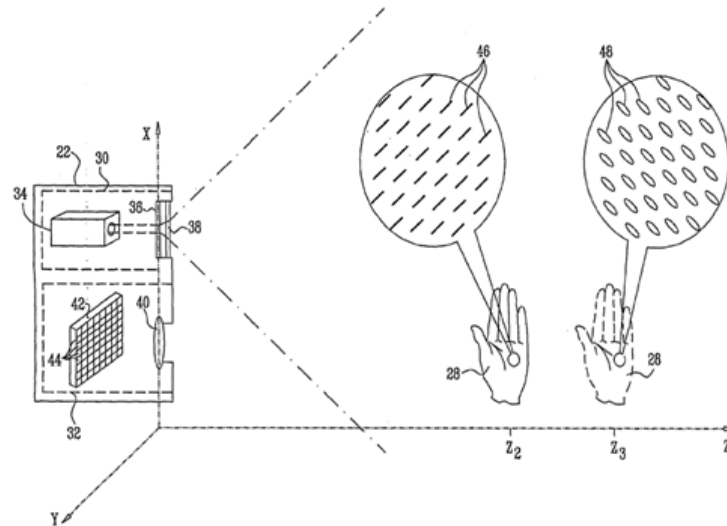


Figura 17 - Profundidade de Foco, Freedman et al, PrimeSense patent application US 2010/0290698

2.3.5 Kinect – Profundidade de som

A profundidade de som utiliza paralaxe, em que se uma pessoa olhar para uma cena de um ângulo diferente, os materiais que estão perto ficam deslocados para o lado, mais do que os materiais que estão mais longe.

O Kinect analisa essa mudança do padrão de salpico pela projeção de um local e ao mesmo tempo observando do outro. (MacCormick, 2010)

2.3.6 Inferir a Posição do corpo

Este segundo processo, Inferir a posição do corpo, consiste em dois subprocessos, utilizando representações intermediárias de “partes do corpo”. (MacCormick, 2010)

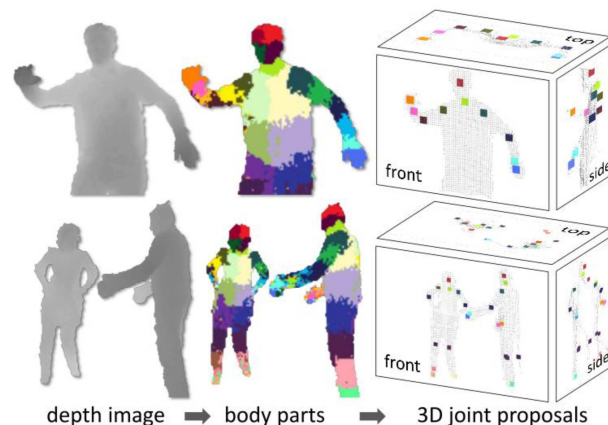


Figura 18 - Shotton et al, CVPR (2011)

2.3.7 Primeiro Subprocesso

O primeiro processo dos dois constituídos pela segunda parte de Inferir a posição do corpo começa com cem mil imagens de profundidade detalhadas com esqueletos conhecidos (a partir de um sistema de captura de movimento). (MacCormick, 2010)



Figura 19 – Primeiro Subprocesso, Shotton et al, CVPR (2011)

Para cada imagem real, processa e trata a imagem retornando dezenas de imagens, utilizando técnicas de computação gráfica. Usa a computação gráfica para processar todas as sequências de 15 diferentes tipos de corpo e vários outros parâmetros obtendo desta forma mais de um milhão de exemplos de treino.

Aprende através do algoritmo de floresta de decisão aleatório, mapeando as imagens de profundidade para as partes do corpo.

Uma floresta de decisão aleatória é uma versão mais sofisticada da clássica árvore de decisão que é por sua vez como um jogo de “vinte perguntas” pré planejadas. (MacCormick, 2010)

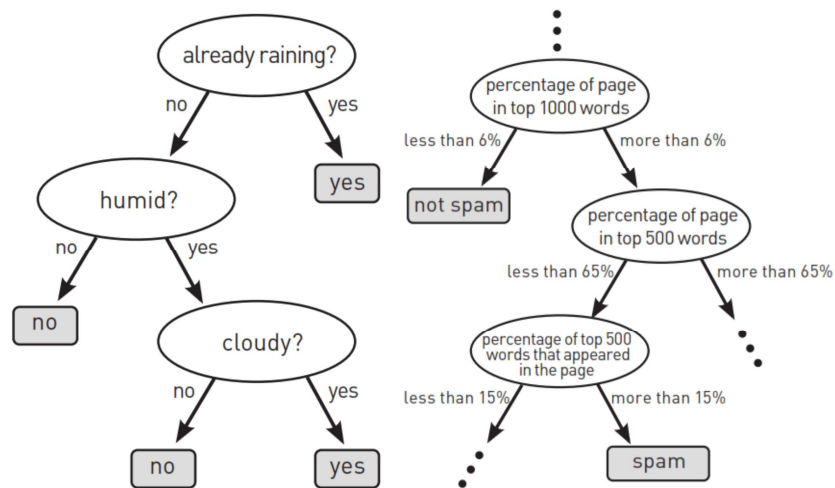


Figura 20 – Jogo “twenty questions”, Ntoutas et al, WWW (2006)

O jogo de vinte perguntas é um jogo onde para começar, se escolhe uma pessoa e esta pessoa deve pensar num objeto. Um jogador faz perguntas sobre o objeto a pessoa escolhida, enquanto este responde sim ou não as perguntas do jogador. Depois de ouvir a resposta, é permitido a pessoa que faz as perguntas, adivinhar qual é o objeto. Cada jogador está autorizado a fazer vinte perguntas. Se o jogador que fizer as perguntas adivinhar o objeto, ganha o jogo e começa o jogo de início pelo jogador que ganhou.

Em relação ao Kinect, que tipo de “perguntas” pode fazer.

Na versão simplificada, um exemplo pode ser, “É o *pixel* naquele offset no fundo?” Enquanto na versão real, uma pergunta pode ser, “Como é a profundidade (normalizada) naquele *pixel* em comparação a este *pixel*?”

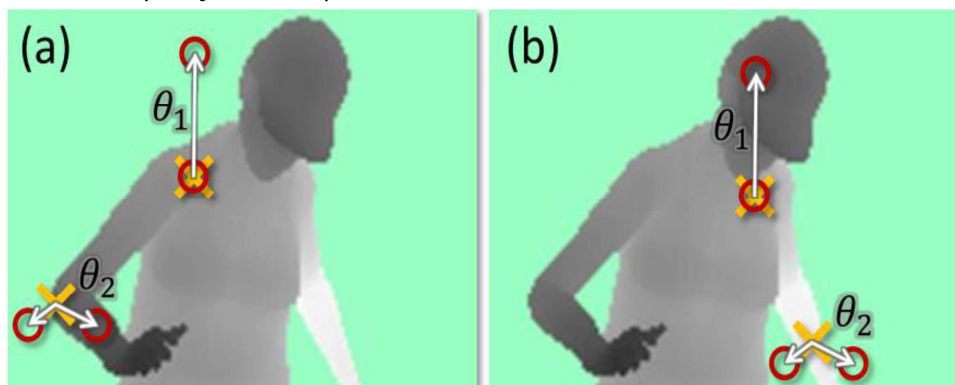


Figura 21 - Shotton et al, CVPR (2011)

Para aprender uma árvore de decisão, se escolhe como a próxima questão, aquele que é o mais útil (a parte relevante) no treino de dados. Para uma árvore de decisão de utilização de um guarda-chuva, qual é o mais útil, “chovendo?”, ou “nublado?”?

Na prática, “útil” é igual ao ganho de informação G (que é derivado de entropia H).

$$G(\phi) = H(Q) - \sum_{s \in \{l,r\}} \frac{|Q_s(\phi)|}{|Q|} H(Q_s(\phi))$$

Figura 22 - Shotton et al, CVPR (2011)

O Kinect utiliza uma floresta de decisão aleatória:

- Aleatória:
 - Porque existe muitas perguntas possíveis e por isso é feita uma seleção de 2000 perguntas de cada vez.
- Floresta
 - Aprendizagem através de múltiplas árvores;
 - Para classificar, adiciona o resultado das árvores;
 - Sendo que os resultados são na realidade, probabilidades distribuídas e não decisões únicas.

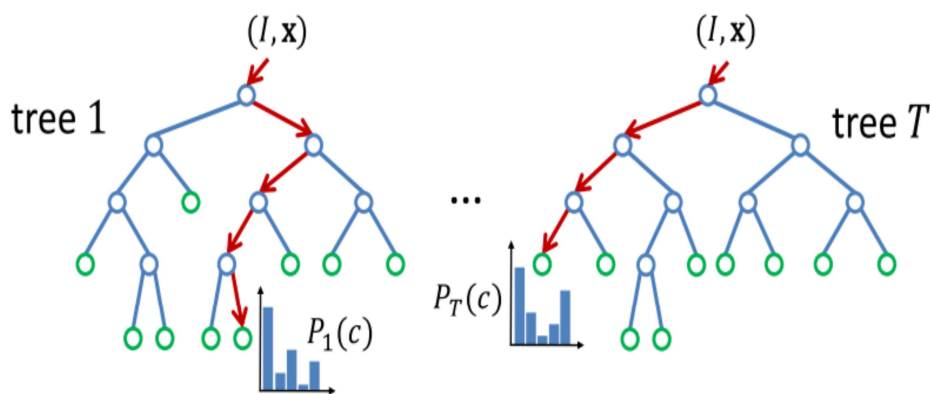


Figura 23 - Shotton et al, CVPR (2011)

A aprendizagem da floresta de decisão requer vinte e quatro mil horas de processador mas toma apenas um dia utilizando centenas de computadores simultaneamente.

“Para manter os tempos de treino baixos, empregamos uma aplicação distribuída. Treinando três árvores de profundidade em vinte imagens de um milhão de imagens demora um dia num *cluster* de 1000 núcleos”. — Shotton et al, CVPR (2011), (MacCormick, 2010)

2.3.8 Segundo Subprocesso

O Segundo subprocesso transforma a imagem de cada parte do corpo num esqueleto. O algoritmo de deslocamento significativo é utilizado para calcular, de forma robusta, modos de distribuições de probabilidades. Este algoritmo é simples, rápido e eficaz. (MacCormick, 2010)

2.4 Leap Motion

Um outro desenvolvimento importante é o sensor **Leap Motion**, um sensor de profundidade feito em especial para controlar todos os recursos das mãos.

A visão de David Holz, Director Técnico da LEAP, e Michael Buckwald, cofundador, que achavam que a iteração Humano – Computador era francamente curta, foram de criar um sistema capaz de permitir ao utilizador controlar um ambiente digital da mesma forma que se controla objetos no mundo real.

Pergunta:

Se pode-se pegar numa bola no mundo real, porque é que não se pode pegar numa bola no ambiente digital?

O primeiro produto desenvolvido por esta equipa, como prova de conceito, foi um dispositivo de pequeno porte, iPod, que pode ser conectado a porta USB de um computador e fornecer deteção de movimento, de modo que os movimentos exatos dos dedos individuais e a rotação do pulso pudessem ser detetados com precisão e processados sem latência.

Como funciona:

O controlador Leap Motion junto com a API atual oferece posições no espaço cartesiano de objetos pré definidos, como as pontas dos dedos, ponta da caneta, etc. As posições retornadas são em relação ao ponto central do controlador de movimento Leap que está localizado na posição do segundo emissor infravermelho centrado.

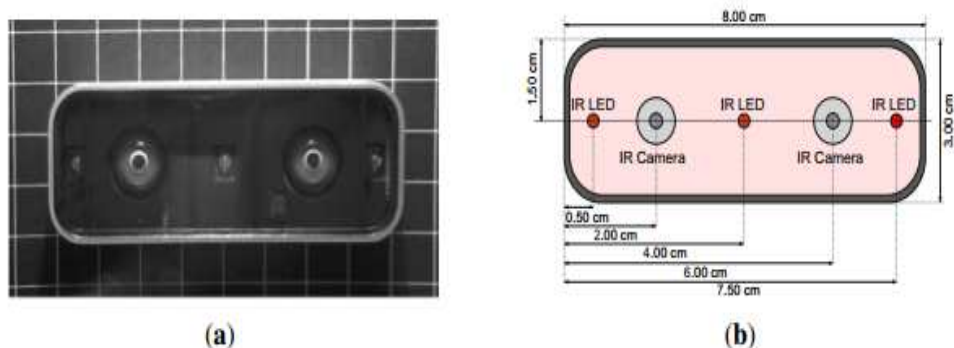


Figura 24 - Dispositivo Leap Motion

O controlador, tal como demonstrado na figura, **Figura 24**, consistem em três emissores de infravermelho e duas câmaras também de infravermelho. O **Leap Motion** é por isso categorizado de sistemas de rastreamento ótico baseado em visão estéreo.

2.5 Aplicações

Muitos projetos foram desenvolvidos utilizando as tecnologias acima referidos, sendo que muitos foram feitos especialmente para a utilização do Kinect porem destaca-se os seguintes:

2.5.1 OMEK SDK

O OMEK SDK trabalha com compreensão dos processos de entrada a partir de um novo tipo de câmara de profundidade, otimizado para operações de curto alcance. Essas novas câmaras estão sendo desenvolvidas em módulos que irão permitir que sejam incorporados diretamente nos computadores e dispositivos Tablet.

O produto, que tem como principal objetivo ser precisa e robusta, faz um acompanhamento ágil e preciso de ambas as mãos e os dedos, mesmo com obstruções.

Os principais destaques alem dos já apontados:

- Modelo 3D completa das mãos;
- Identificação da mão direita versus esquerda com posição, orientação e vinte e duas articulações nomeadas;
- Suporta a interação com as duas mãos;
- Não é necessária calibração;
- Suporte de interfaces tão perto quanto 10 centímetros do monitor;
- Inclui um pacote de Gestos, "*Gesture Pack*", com gestos pronto a utilizar.

OMEK BECKON

É um conjunto de ferramentas e *middlewares* da OMEK que acompanha os movimentos dos jogadores em ambientes arbitrários, isolando os jogadores na cena e, em seguida, a construção de modelos 3D dos ambientes. (Beckon)

Os algoritmos são desenhados robustamente para segmentar os "jogadores" de outros objetos na cena. Beckon faz a auto calibração do modelo do esqueleto para duma forma otimizada se enquadrar com as dimensões do corpo de cada jogador para assegurar um rastreamento suave e consistente em uma ampla variação do tamanho e medidas do utilizador. A medida que os utilizadores se movem livremente pela cena, os seus movimentos são mapeados para o modelo do esqueleto em tempo real, providenciando um rastreamento robusto mesmo em casos onde partes do corpo não estão sendo captadas ou estão escondidos da câmara. (Beckon)



Figura 25 - OMEK Beckon Fonte: (Beckon)

2.5.2 SoftKinectic SDK

O SoftKinectic é uma plataforma de desenvolvimento e implantação de reconhecimento de gestos para a criação de aplicações de várias áreas que se beneficiam da interação natural com conteúdo digital. É compatível com todas as câmaras de sensores de profundidade em 3D e permite que os desenvolvedores de aplicações possam desenvolver interfaces intuitivas, transparente e envolventes. (SoftKinectic)

É possível com apenas uma câmara de profundidade 3D, capturar movimentos dos utilizadores em tempo real, com grande precisão e resolução. Cada *pixel* produzido pelo sensor 3D é caracterizado pela sua profundidade – A distância entre a câmara e o ponto do objeto correspondente a aquele pixel. Dessa forma, com a informação dessa profundidade através de um vetor, há possibilidades para desenvolver novas aplicações interativas baseados nos pontos já identificados. Interfaces Intuitivas, transparente e envolventes.



Figura 26 - Soft Kinect

Partes do Corpo

Identifica e rastreia movimentos de partes do corpo dos utilizadores e providencia um rastreamento robusto e superior e também a gestão de oclusões.

Sistemas Avatar 3D – Extração do Esqueleto do Corpo Inteiro

O sistema avatar 3D permite a criação de avatares de terceira pessoa de forma realística. Esse sistema pode ser utilizado por desenvolvedores para animar modelos 3D em tempo real utilizando métodos *standards* tais como “mesh skinning”.

2.5.3 3Gear's SDK

O sistema 3Gear consiste nos sensores de curto alcance da PrimeSense, um suporte para esse sensor e a plataforma 3Gear SDK. (3Gear)

Utiliza o sensor para o rastreio das mãos e com o suporte ajusta a câmara na melhor altura possível.

A plataforma combina as informações da posição transmitida pelas câmaras a 30 *frames* por segundo com uma base de dados trinta mil registos de configuração dos dedos e da mão.

O processo de correspondência leva apenas trinta e três milissegundos criando a impressão de que o computador responde instantaneamente.



Figura 27 - Sistema 3 Gear's

2.5.4 SigmaNIL SDK

SigmaNIL é uma plataforma *open source*, dirigida a interfaces de utilizadores naturais, com reconhecimento das mãos e elevado sistema de precisão dos dedos, reconhecimento de gestos e rastreamento do esqueleto da mão.

Foi desenhado para suportar todos os dispositivos de sensores de profundidade e pode tanto ser utilizado com o KinectSDK ou o OpenNI.

É constituído pelo SigmaNIL Core que contem o código fonte disponibilizado por módulos, segmentação da mão, esqueleto da mão, forma da mão e Gestos com a Mão. Contem ferramentas de treino para adaptar os módulos, criando ficheiros de dados relevantes.

Há portanto, a fase de treino dos dados, onde se identifica qual o gesto ou a forma que estamos a produzir com as mãos. No caso da língua gestual portuguesa, se queremos identificar a letra A, no módulo de gestos estáticos criamos um ficheiro com inúmeros exemplos da configuração manual da letra A para depois de terminada essa fase de treino podermos testar o sistema. O sistema vai comparar os dados introduzidos com os ficheiros de dados de varias letras no caso e nos retorna a letra correspondente ao gesto efectuado. (SigmaNil-Framework)

2.6 Estudos – Reconhecimento de Línguas Gestuais.

Foram feitos já muitos trabalhos no que tem a ver a configuração estática da mão humana que é chamada de postura da mão, “Hand Posture” e as suas aplicações para a interação humano computador.

Os pesquisadores têm encontrado e utilizado diferentes procedimentos e técnicas de aquisição destas configurações e comportamentos.

Entre os muitos artigos há a realçar:

2013 - [Xiujuan Chai, Guang Li, Yushun Lin, Zhihao Xu, Yili Tang, Xilin Chen] Sign Language Recognition and Translation with Kinect

No trabalho que desenvolveram e que ainda desenvolvem, trabalham com Alinhamento de trajetória 3D para o reconhecimento da língua Gestual. Em que utilizam a Microsoft Kinect SDK para o rastreamento da mão que gera a descrição da trajetória 3D de uma palavra correspondente em LG.

Considerando a diferença da velocidade da mão, fazem uma recolha de nova amostra linear para terem a trajetória normalizada pela média acumulada do comprimento de todo o vetor. Com esta operação eles normalizam a trajetória de cada palavra no mesmo ponto de amostragem.

Para o reconhecimento dos movimentos, são necessários alinhamentos entre a trajetória do gesto efetuado e a galeria de vetores existentes. Os resultados correspondentes são calculados utilizando a medição de distância euclidiana para se saber o resultado do reconhecimento.

Nos seus testes, obtêm uma taxa de acerto de 83,51% a 96,32%, utilizando 5 amostras de cada um dos 293 gestos gravados.

[F. Trapero Cerezo] 3D Hand and Finger Recognition using Kinect

Incluído no campo de HCI, o reconhecimento de dedos e da mão, no trabalho que eles desenvolveram quiseram descrever um sistema eficiente e robusto de rastrear a mão e os dedos utilizando o dispositivo Kinect. Para isso, tentaram calcular o contorno da mão e do interior e em seguida procurar neste conjunto de dados, pontos são os candidatos a ser a ponta dos dedos ou o centro da mão. Para encontrar as pontas dos dedos e o centro da palma da mão eles criaram um conjunto de passos que o seu sistema tem que percorrer.

Conjunto de passos criado:

- A geração de uma matriz com pixéis próximos;
- Diminuição de ruídos;
- Classificação do contorno ou de pixéis dentro do conjunto;
- Diferenciar as mãos e calcular o seu contorno;
- Alocar pontos dentro da mão;
- Encontrar o centro da palma da mão;
- Encontrar as pontas dos dedos;
- Alocar pontos num espaço 3D.

O trabalho de F. Trapero Cerezo, pode ser considerado muito importante no reconhecimento da Língua Gestual pois ajudaria e muito na descrição dos gestos. Muitos gestos similares poderão ser classificados corretamente se na sua configuração houver um dedo que aparece e no outro gesto parecido este estiver escondido da câmara Kinect.

[Daniel James Ryan] Finger and Gesture recognition with Microsoft Kinect

A visão deles é a de criar uma API que consegue rastrear os dedos das mãos e efetuar reconhecimento de gestos com bons resultados e coloca-lo disponível para um grande número de pessoas. Para isso também consideraram o rastreamento do contorno da mão, algoritmo que eles mesmos desenvolveram. Para isso o algoritmo de rastreamento trabalha de forma a ler / varrer a imagem de profundidade para encontrar um *pixel* de contorno valido. Quando encontrado começa a rastrear o contorno do objeto que este *pixel* pertence. Isso de acordo com o autor é feito procurando num conjunto de pixéis ao redor do mesmo. Depois de terminado o rastreamento retorna uma lista ordenada de posições de pixéis de contorno:

- O processo de reconhecimento dos dedos deles consiste em três passos;
- O primeiro é o de encontrar curvas no contorno da mão;

O segundo é o de encontrar quais as curvas são ponta de dedo e o ultimo passo é o de encontrar os contornos que são os meios dos dedos.

No processo de gravação de gestos eles guardam a posição dos dedos e a direção de todas as frames dentro de um período de tempo especificado pelo utilizador ou então até que for alcançado um numero de frames máximos.

No processo de reconhecimento dos gestos, utilizaram uma variante do algoritmo DTW. O algoritmo DTW reconhece as similaridades entre duas serieis temporais. As series não necessitam de ser sincronizados dessa forma o utilizador pode fazer o gesto mais rápido ou mais lento que o tempo do gesto guardado.

3 Reconhecimento de Padrões / Mineração de Dados

Para reconhecer um gesto efetuado através de uma câmara Kinect, este envia uma luz estruturada para a cena, e deteta a deformação desse padrão e nos envia esses dados, precisamos saber como é feita a classificação dos dados que nos é enviado pela Kinect. Antes de falar sobre a classificação importa termos um conhecimento básico do que está por trás.

3.1 Descoberta de Conhecimento

Descoberta de conhecimento é um processo de desenvolvimento de métodos e técnicas de extração de conhecimento de alto nível a partir de informação guardada em base de dados. É constituído por um meio físico, computadores e redes de computadores e utiliza base de dados (Data Warehouses).

O processo de descoberta de conhecimento, que vem da sigla KDD (Knowledge Discovery in Data Base), permite identificar padrões e ou modelos sejam eles novos, modelos ou padrões úteis e compreensíveis.

O processo, segundo FAYYAD, é um processo, interativo, iterativo, distinguidas em 5 etapas.

- Seleção dos Dados;
- Pré – processamento;
- Transformação;
- Data Mining;
- Interpretação e Avaliação.

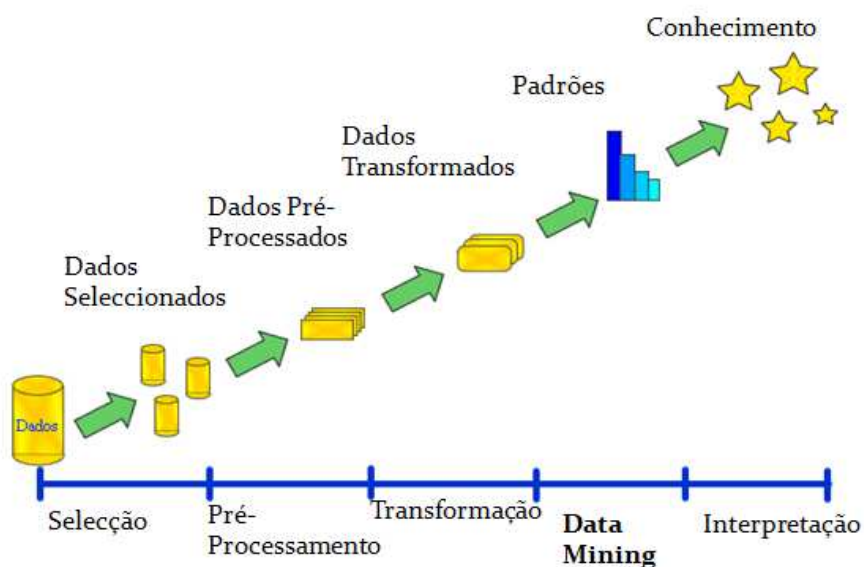


Figura 28 - Processo Data Mining ou Mineração de Dados

Na figura, **Figura 28**, mostra-se bem a forma como esse processo de Descoberta de Conhecimento é feito, quais são as etapas e qual é a ordem, que nesse caso é importante.

A partir dos dados que nos é enviado pelo Kinect ou que recolhemos do Kinect, é feito um pré processamento destes e transformados de forma que seja possível a um classificador identificar padrões e classifica-los.

O ponto mais importante depois de transformar os dados de forma que seja passível de classificação é o ponto Data Mining do processo de descoberta de conhecimento.

3.2 Data Mining

(Frاند, 1996) O objetivo do Data Mining, muitas vezes chamado de Descoberta de Conhecimento, é o processo da análise dos dados de diferentes perspectivas e resumi-los em informação útil.

A ferramenta do Data Mining é uma das muitas ferramentas analíticas para analisar os dados. Permite que os utilizadores façam a análise dos dados de diferentes ângulos e dimensões, categorizando-os e resumindo as relações identificadas.

Tecnicamente Data Mining é um processo de deteção de correlações ou padrões entre os muitos campos numa base de dados relacional.

3.2.1 Inovação Continua

Embora Data Mining ou mineração de dados seja um termo relativamente novo, a tecnologia que está por trás não o é. Muitas empresas têm utilizado processos semelhantes a Data Mining, que utilizam computadores poderosos para filtrar grandes volumes de dados através dos leitores dos supermercados e analisando relatórios de pesquisa dos supermercados por vários anos.

No entanto, as inovações contínuas no processamento dos dados por parte dos computadores, no armazenamento de dados e nos programas de estatísticas, aumentam drasticamente a precisão da análise, reduzindo muitos custos.

3.2.2 Processo Data Mining

O uso do processo de Data Mining hoje em dia é feito por empresas com grande interesse nos consumidores – retalho, financeiro, comunicações e organizações de mercado.

Dá a possibilidade as empresas destes ramos de determinar as relações entre fatores internos, tais como preços, posicionamento de produto ou capacidades dos empregados e fatores externos tais como indicadores de economia, competição e dados demográficos do cliente. Ainda dá a possibilidade de determinar os impactos nas vendas, na satisfação dos clientes e nos lucros das empresas.

Finalmente permite-lhes ter a informação resumida para visualizar dados detalhados de transações.

Com Data Mining, um vendedor pode utilizar os registos de ponto de venda das compras do consumidor para enviar promoções personalizadas baseados no histórico de compra de cada um.

3.2.3 Tipos de Data Mining

Encontramos dois tipos de Data Mining que tem como objetivo detetar padrões e regularidades implícitas nos dados.

Tipos:

- Data Mining Preditivo:
 - Orientado por um objetivo – é um processo direto de descoberta de conhecimento.
- Data Mining Descritivo
 - Não existe um objetivo definido – é um processo indireto de descoberta de conhecimento.

3.2.4 Como funciona?

O processo de Data Mining nos dá a ligação entre os sistemas analíticos e os sistemas transacionais. O programa de Data Mining analisa as relações e os padrões em transações armazenadas em base de dados e permite que os utilizadores façam consultas abertas.

Existem muitos tipos de programas analíticos: programas de estatísticas, de aprendizagem da máquina e redes neuronais.

Geralmente qualquer destes tipos de relacionamento é procurado entre os dados:

- **Classes:** Os dados armazenados são usados para localizar em grupos pré determinados.
- **Cluster:** Os registos são agrupados de acordo com as relações lógicas ou com preferências dos utilizadores.
- **Associações:** Os dados podem ser extraídos para identificar associações.
- **Padrões Sequenciais:** Os dados podem ser extraídos para antecipar padrões nos comportamentos e tendências.

O processo consiste em cinco elementos principais:

- Extrair, transformar e carregamento de dados para um sistema de Armazenamento de Dados.
- Guardar e gerir os dados num sistema de base de dados multidimensionais.
- Providencia acesso de dados para analistas de negócios e profissionais de tecnologias de informação.
- Analisa os dados pelo programa.

- Disponibiliza os dados num formato legível, tais com gráficos ou tabelas.

São disponibilizados diferentes níveis de análise:

- **Redes Neurais Artificiais:** São modelos preditivos não lineares que aprendem por meio de treino e se assemelham a redes neuronais de estrutura biológica.
- **Algoritmos Genéricos:** São técnicas otimizadas que utilizam processos tais como combinações genéticas, mutações e seleção natural num projeto baseado em conceitos de evolução natural.
- **Árvores de Decisão:** Estruturas em forma de árvore que representam conjuntos de decisões. Estas decisões geram regras para a classificação de um conjunto de dados. Métodos específicos de árvores de decisão incluem árvores de classificação e regressão.
- **Método do vizinho mais próximo:** é uma técnica que classifica cada registo num conjunto de dados com base em uma combinação das classes de registos K mais parecido com o mesmo num conjunto de histórico de dados. Muitas vezes é chamado de “K - Técnica do Vizinho mais próximo”.

3.2.5 Modos de aprendizagem:

Há três tipos de modos de aprendizagem associada ao **Data Mining**:

Aprendizagem Supervisionada:

- Os algoritmos são treinados com amostras de dados previamente classificados.

Aprendizagem Não Supervisionada:

- Os algoritmos processam os dados autonomamente

Aprendizagem Semi Supervisionada:

- Desconhecem as classes de alguns pontos de treino.

3.3 Aprendizagem da Máquina

Aprendizagem da máquina é a ciência que faz com que os computadores agem sem serem explicitamente programados para o fazerem. Nas últimas décadas, a aprendizagem da máquina nos deu várias coisas, entre elas temos os carros de auto condução, o reconhecimento da voz, pesquisas eficazes na internet.

Ao longo destas últimas duas décadas, tornou-se num dos pilares da tecnologia de informação, que faz parte da nossa vida. Com quantidades de dados sempre em crescente disponível e sendo necessária a análise desses dados há razões para acreditar que a análise inteligente se tornará mais difundida como um fator necessário para o progresso tecnológico.

3.4 Como Identificar?

A aprendizagem de máquina pode aparecer em muitas formas. Por exemplo, muitos de nós estamos familiarizados ao conceito de ranking da página web. Ou seja, o processo de envio de uma consulta a um motor de busca que, em seguida, encontra páginas web relevantes para a consulta e que lhes retorna em sua ordem de relevância.

Ou seja, dada uma determinada consulta o motor de busca retorna uma lista ordenada de páginas. Para atingir este objetivo, um motor de busca precisa “saber” quais as páginas que são relevantes e quais as páginas que correspondem a consulta. Esse conhecimento pode ser ganho ou conseguido através de várias fontes:

As estruturas de hiperligação das páginas Web; os seus conteúdos; a frequência com que os utilizadores irão seguir as hiperligações sugeridas numa consulta ou de exemplos de consultas em combinação com páginas classificadas manualmente.

Cada vez mais se utiliza aprendizagem de máquina ao invés de suposições e de engenharia inteligente, para automatizar o processo de conceção de um bom motor de busca.

Uma das aplicações bastante relacionada é a técnica de filtragem colaborativa. Grandes livrarias da internet, tais como Amazon ou aplicações web de aluguer de vídeos como o NetFlix, usam esta informação de forma extensiva para motivar os seus utilizadores a comprar produtos adicionais ou alugar mais filmes.

O problema é bastante semelhante ao de uma classificação da página da web. Como antes, queremos obter uma lista ordenada, neste caso de artigos.

A principal diferença é que não é feita uma consulta explícita, em vez disso, utiliza-se apenas compras dos utilizadores e as suas decisões de visitas para prever hábitos de compras e de visitas. A chave da informação é as decisões feitas por utilizadores semelhantes daí a natureza colaborativa do processo.

É evidente que é melhor para o funcionamento deste processo que se disponha de um sistema automático, evitando assim conjeturas e poupando muito tempo.

3.5 Técnica de Aprendizagem Utilizada

Implementação do SVM no R.

(Karatzoglou, et al., 2006) A primeira implementação do SVM no R foi introduzida no pacote `e1071`, que foi a utilizada no projeto, pacote esse que foi implementado por Dimitriadou, Hornik, Leisch, Meyer, e Weingessel em 2005.

A função `svm()` no pacote `e1071`, nos providencia uma interface rígida a biblioteca `libsvm`, juntamente com métodos de visualização e de possíveis configurações de parâmetros.

A biblioteca **libsvm** é a mais rápida das implementações do SVM no R e muito fácil de utilizar, sendo uma das mais populares formulações – classificação “C” e “V” regressão “E” e “V” e detecção de novidades. Esta biblioteca inclui as funcionalidades de *kernel* mais comuns – linear, polinomial, RBF, e sigmoid – que só é extensível, alterando o código - fonte C++.

A implementação no R é baseada em mecanismos de classe S3. Basicamente, nos providencia uma função de treino com interfaces padrão e formula, e também nos providencia um método de previsão – “predict()”.

Em adição, o método “plot()” para a visualização dos dados, vetores de suporte e os limites de decisão são fornecidos. A afinação é feita usando a `Framework tune()` no pacote `e1071`, realizando uma pesquisa por uma gama de parâmetros especificados.

SVM – Support Vector Machines

(DTREG)De acordo com a Teoria Estatística da Aprendizagem, as Support Vector Machines, que se designa pelo termo Maquinas de Vetores de Suporte (MVS’s) em português, surgiram pela primeira vez na década de setenta do século passado com o trabalho de Vapnik em 1979.

Trata-se de algoritmos que podem ter algumas variações mediante a aplicação, criados inicialmente para classificação de dados, começam a ser utilizados também em regressão.

Portanto, define-se por uma técnica de aprendizagem que realiza a classificação através da construção de N hiperplano dimensional, separando otimamente os dados em duas ou mais categorias.

Os resultados dessas técnicas são comparados as Redes Neurais Artificiais (RNA’s), outro algoritmo de aprendizagem, tendo mesmo muitas vezes melhores resultados.

É uma técnica que trabalha com varias funções *Kernel*, entre elas a função *Kernel* RBF – Radial Basics Function, utilizada para a construção do modelo do projeto.

Nesta técnica, uma variável de previsão é chamado de atributo e um atributo transformado utilizado para definir hiperplano é chamado de características. O processo de escolha da representação mais adequada é conhecido como seleção característica, enquanto um conjunto de características que descreve um caso é denominado de vetor.

Objetivo do SVM ou MVS

O objetivo da técnica de aprendizagem SVM é o de encontrar um hiperplano ótimo que separa os conjuntos de vetores de uma forma que casos de uma categoria da variável de interesse estão num lado do plano e casos de outra categoria estão no outro lado do plano.

Os vetores que ficam mais perto do hiperplano são os vetores de suporte.

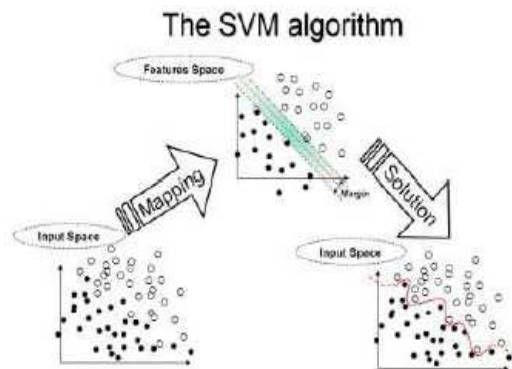


Figura 29 - Algoritmo SVM

Para melhor entender, neste exemplo, **Figura 30**, os casos de uma categoria na parte esquerda da figura e os casos de outra categoria na direita sendo que os casos aqui estão completamente separados.

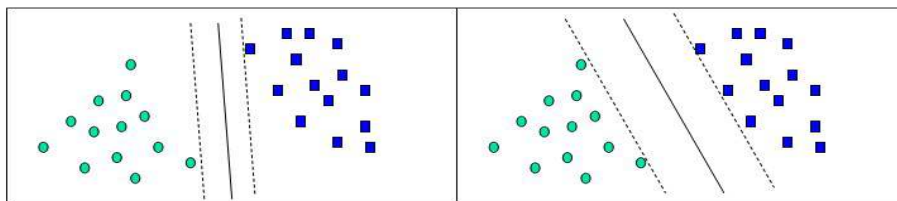


Figura 30 - Algoritmo SVM 1

O SVM tenta encontrar um hiperplano de uma dimensão, sendo uma linha que separa os casos consoante a categoria.

Podemos ter um numero infinito de linhas, neste caso temos duas linhas candidatas na figura, a questão que se põe é qual delas é a correta e como podemos definir uma linha ótima. As linhas com pontos são desenhadas em paralelo a linha de separação marcando a distância entre esta linha e os vetores mais próximas da mesma.

As distâncias entre as linhas com pontos são chamadas de margens. Os vetores que limitam a largura das linhas são os chamados vetores de suporte.

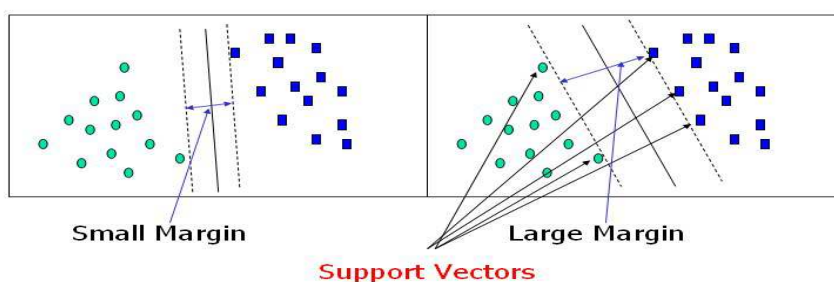


Figura 31 - Algoritmo SVM - Margens

Uma análise do SVM encontra a linha – ou o hiperplano – que está orientado de modo que a margem entre os vetores de suporte é maximizada. A margem da linha da imagem a direita é superior a margem da linha da imagem a esquerda.

Geralmente o SVM não lida com análises que se consistem em variáveis de interesse de apenas duas categorias, com duas variáveis de previsão nem que o conjunto de dado pudesse ser dividido por uma linha reta.

Em geral, o SVM trata:

- Mais de duas variáveis de previsão;
- Separação dos pontos com curvas não lineares;
- Tratamento de casos onde os conjuntos não podem ser completamente separados.
- Tratamento de classificação onde existem mais do que duas categorias.

No exemplo anterior tínhamos apenas duas variáveis de previsão e conseguimos demonstrar num plano de duas dimensões. Ao adicionarmos mais uma variável de previsão podemos então utilizar o seu valor para uma terceira dimensão e demonstrar os pontos num cubo de 3 dimensões.

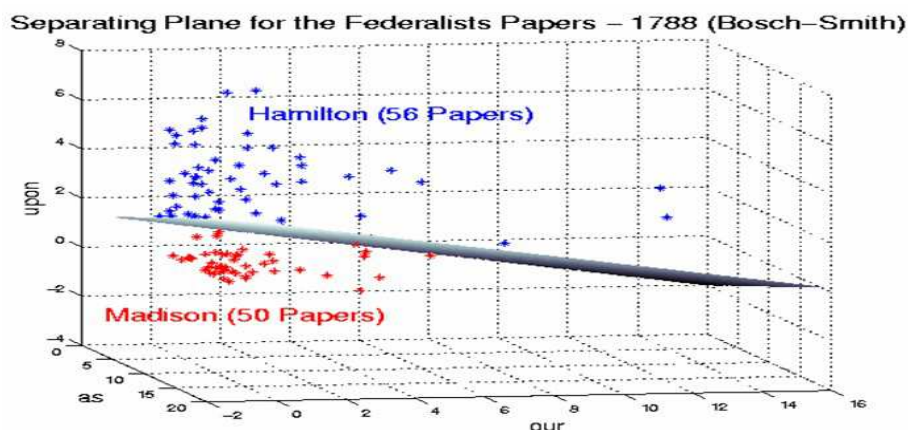


Figura 32 - Algoritmo SVM - HiperPlano

Da mesma forma que pontos de um plano de duas dimensões podem ser separados por uma linha de uma dimensão, pontos num plano de três dimensões podem ser separados um plano de duas dimensões.

Ao adicionar uma variável de previsão adicional os pontos de dados podem ser representados num espaço n – dimensional e um hiperplano n-1 dimensional pode separar esses pontos.

A forma mais simples de dividir dois grupos é com uma linha reta, com uma superfície plana ou com um hiperplano de dimensão N. Mas e se os pontos são separados por uma região não linear?

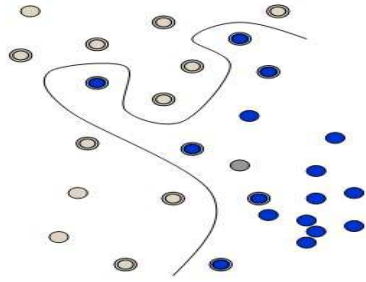


Figura 33 - Algoritmo SVM - Separação Vectores

Neste caso precisamos de uma linha de divisão não linear.

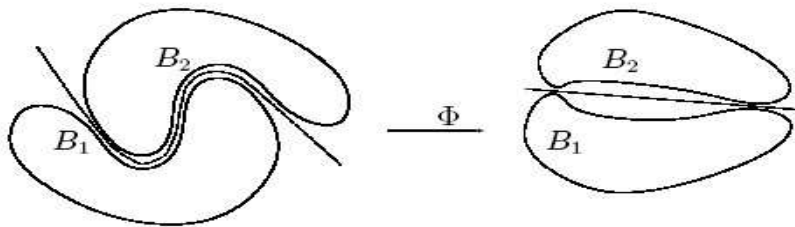


Figura 34 - Algoritmo SVM - Divisão Não Linear

Ao invés de traçar uma linha curva não linear sobre os dados, o SVM trata disso utilizando as funções *kernel* para mapear os dados em diferentes espaços em que um hiperplano pode ser utilizado para a separação.

A função *kernel* pode transformar os dados num espaço dimensional superior para que seja possível uma separação.

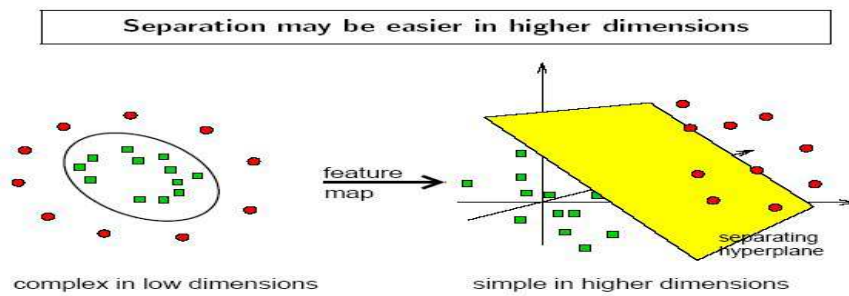


Figura 35 - Algoritmo SVM - Dimensões

Pontos Importantes na utilização do SVM

- **Como utilizar eficazmente e eficientemente o SVM**
 - Fazer a normalização dos vetores de entrada;
 - Selecionar o C – que é o parâmetro que controla o conflito da escolha entre a margem e o erro;
 - Efetuar a manipulação de dados desequilibrados;
 - Selecionar uma função *Kernel*;
 - Classificação multi-classe;
 - Selecionar um algoritmo de treino.
- **Como atribuir os valores de Recurso? Pontos a ter em consideração:**
 - A importância do recurso é monótona em seu valor absoluto;
 - Quanto maior for o valor absoluto do recurso mais influencia este ganho.
 - Problemas típicos: valores como por exemplo número de porta [0-5] e valores como preço [0-100000]
 - Se quer características relevantes cada vez maiores e características irrelevantes cada vez menores.
 - Normalização dos valores para que os recursos sejam igualmente importantes:
 - Por média e variância;
 - Por outra distribuição.
 - Normalização para que os vetores estejam na mesma escala:
 - Dados direcionais: classificação de texto;
 - Normalizando o comprimento do vetor de acordo com algumas normas;
 - Muda se um problema é (linearmente) separável ou não.
 - Dimensionar todos os vetores para um comprimento que permite um treino numericamente estável.
- **Selecionando uma função *Kernel*. Pontos a ter em consideração:**
 - A função *kernel* pode ser considerada como uma medida de similaridade:
 - Exemplos numa mesma classe devem ter um valor de *kernel* significante;
 - Exemplos em classes diferentes devem ter baixo valor de *Kernel*;
 - Valor de *kernel* ideia é calculado pela relação de equivalência.
 - A normalização também se aplica ao *Kernel*:
 - Os recursos implícitos têm um peso relativo;
 - Para dados direcionais deve ser normalizada;
 - Potenciais problemas com números grandes, por exemplo, para o *kernel* polinomial para um d grande.

Função *Kernel* RBF

(ArGis, 2007)As funções *kernel* RBF são uma série de técnicas exatas de interpolação, isto é, a superfície tem que passar por cada valor de amostra medido.

Há cinco diferentes funções base:

- Thin-plate spline;
- Spline with tension;
- Completely regularized spline;
- Multiquadric function;
- Inverse multiquadric function.

Cada função de base tem uma forma diferente e resulta numa superfície de interpolação ligeiramente diferente. Métodos RBF são na verdade uma forma de redes neuronais artificiais.

Sendo interpoladores exatos, as funções RBF diferem dos interpoladores polinomiais globais e locais, que são os dois interpoladores não exatos que não exigem que a superfície passe por todos os valores medidos.

As funções do RBF podem prever dessa forma valores acima do máximo e abaixo dos valores mínimos medidos ao contrário de outros interpoladores, como mostra nas figuras abaixo.

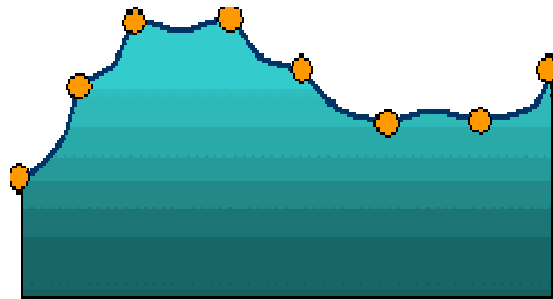


Figura 36 - Algoritmo SVM - Interpoladores como o IDW

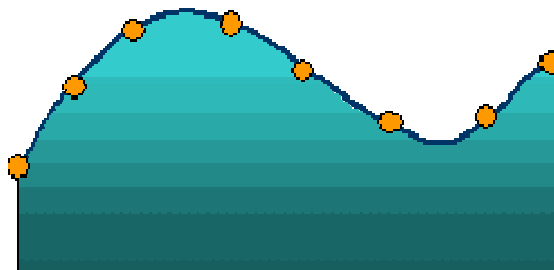
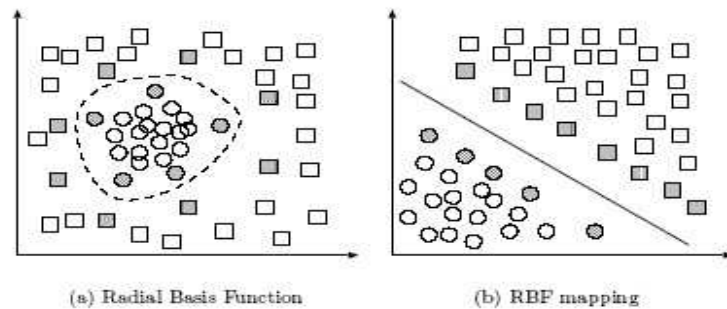


Figura 37 - Algoritmo SVM - Interpoladores RBF



Separable classification with Radial Basis kernel functions in different space. Left: original space. Right: feature space.

Figura 38 - Algoritmo SVM - RBF

Quando utilizar RBF's?

As funções RBF são utilizadas para calcular superfícies lisas de uma grande número de valores. As funções produzem bons resultados para diferentes superfícies, tais como elevação.

Essas técnicas são inadequadas quando há grandes mudanças nos valores de superfície dentro de uma curta distancia horizontal e / ou quando se suspeita de que os dados de amostra são sujeitos a erros ou incerteza.

As diferenças entre as técnicas da função RBF não são grandes porem pode se ter razões para escolher um deles ou pode-se experimentar vários e utilizando validação cruzada selecionar o melhor.

4 Arquitetura do Sistema e Desenvolvimento do Software

O projeto foi de início dividido em duas partes distintas. A primeira parte tinha como objetivo conseguir o reconhecimento dos gestos estáticos, ou seja, os gestos que simbolizam o alfabeto da LGP. A segunda parte já mais composta que a primeira, tinha como objetivo o reconhecimento de gestos em movimento. Ou seja reconhecer uma palavra ou um conjunto de palavras efetuado por um utilizador.

Nos pontos a seguir, Gestos Estáticos e Gestos em Movimento, irei explicar os procedimentos executados durante o desenvolvimento do programa "Virtual Sign".

Em primeiro lugar teremos uma visão geral de cada uma dessas partes, indicando bem os processos seguidos na figura, **Figura 39**, sendo que depois cada um desses processos será detalhado.

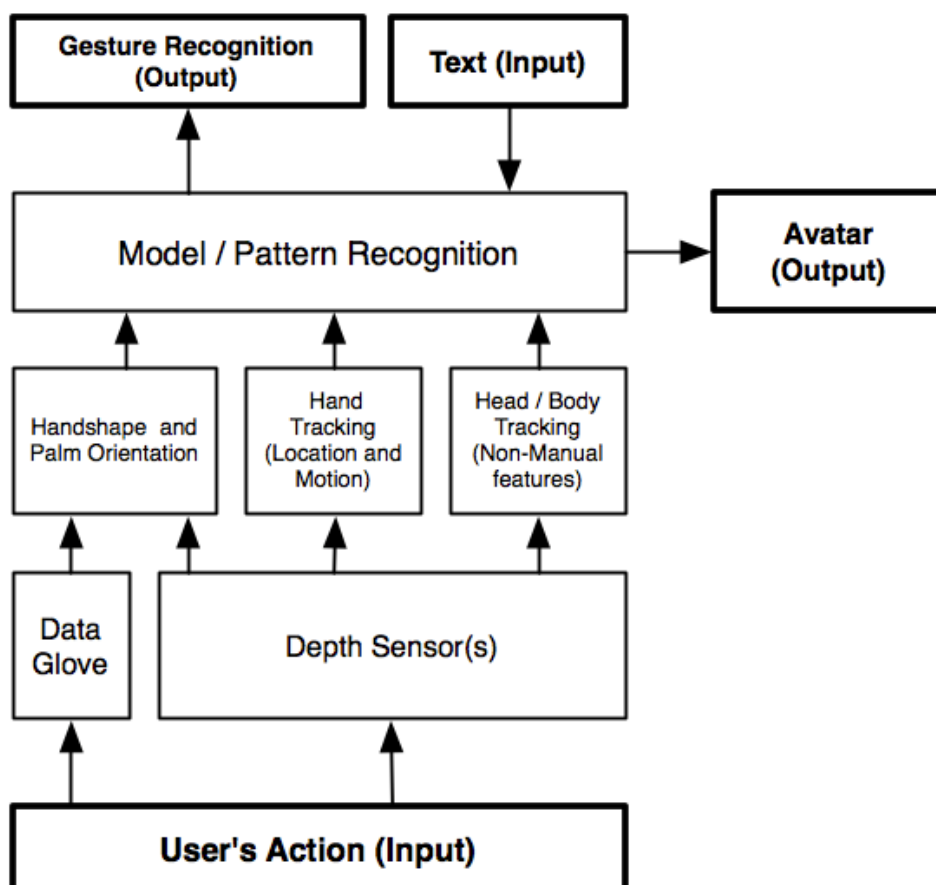


Figura 39 - Modelo Principal Tradutor Bidirecional de LGP

Dispositivos de Interação

A interação do utilizador será feita a partir de dois tipos de dispositivo utilizado:

- Sensores de Profundidade;
- Luva de Dados.

Estes dispositivos, Sensores de Profundidade e Luva de Dados – sendo o primeiro, Kinect XBOX 360, e o segundo a 5DT – já se encontram detalhados sobre a forma como trabalham no capítulo 2 pelo que me irei debruçar sobre a forma como estão a ser utilizados mais a frente. São dispositivos que já se encontravam no projeto integrado no âmbito de investigação que é suportado pela Fundação Ciência e Tecnologia.

***Hand Shape and Palm Orientation* – Forma da Mão e sua Orientação**

Sendo uma característica importante da LG é importante para o reconhecimento do gesto, estático ou dinâmico, a forma da mão e a sua orientação, por ser uma característica que difere muito de gestos para gestos e de pessoa para pessoa. Com isto quero dizer que uma pessoa que faça por exemplo o gesto estático para a letra A pode conter valores que deferem com valores de outra pessoa que faça o mesmo gesto, embora os dados não sejam tão diferentes.

A nível de gestos em movimento já se nota muitas diferenças, que são gestos que contem muitas formas e orientação diferentes o que permite a nível de classificação de dados uma melhor classificação.

Hand Tracking (Location and Motion) – Rastreamento da Mão (Localização e Movimento)

Para os gestos em movimento, é importante fazer o rastreamento da mão, ou seja, para cada gesto, num determinado espaço de tempo, onde este gesto irá ser efetuado, guardamos por onde a mão passa e qual é a sua configuração manual nessa altura. Visto que cada gesto começará nesta primeira fase pelo estado inicial que é com as duas mãos em baixo irá ser explicado a forma como se contornou o problema de boa parte do rastreamento da mão e o movimento inicial seja levantar a mão para começar o gesto.

Head / Body Tracking (Non Manual Features) – Rastreamento da Cabeça e do Corpo

Tendo como das características mais importantes da Língua Gestual, seria importante ter a oportunidade de as considerar na implementação da API de reconhecimento de gestos da LGP. Um gesto pode ser feito com o corpo inclinado para a frente e o corpo inclinado para trás e ter um significado diferente. Como também, um gesto complementado com o uso da expressão facial e do movimento da cabeça ter um significado diferente se for feito apenas o gesto em si. Por essas e outras razões que já foram explicadas no capítulo 2, as características não manuais são pontos importantes no reconhecimento da Língua Gestual.

Model / Pattern Recognition – Modelização e Reconhecimento de Padrão

Depois de se ter considerado, os dispositivos a utilizar, as características que se acha importante para o reconhecimento, falta então a construção do modelo que irá fazer o reconhecimento dos gestos que em seguida nos irá mostrar o resultado.

4.1 Modelo Geral da Aplicação

Tratando de forma diferente, os gestos estáticos dos gestos em movimento, se definiu um modelo geral que irá conter a possibilidade de escolha por parte do utilizador. Essa escolha se incide em se irá utilizar o modelo estático do modelo dinâmico.

Essa escolha permite uma classificação mais rápida, pois os modos de reconhecimento dos gestos são diferentes bem como os dados que são enviados pelo modelo as ferramentas que utilizamos para a classificação dos mesmos.

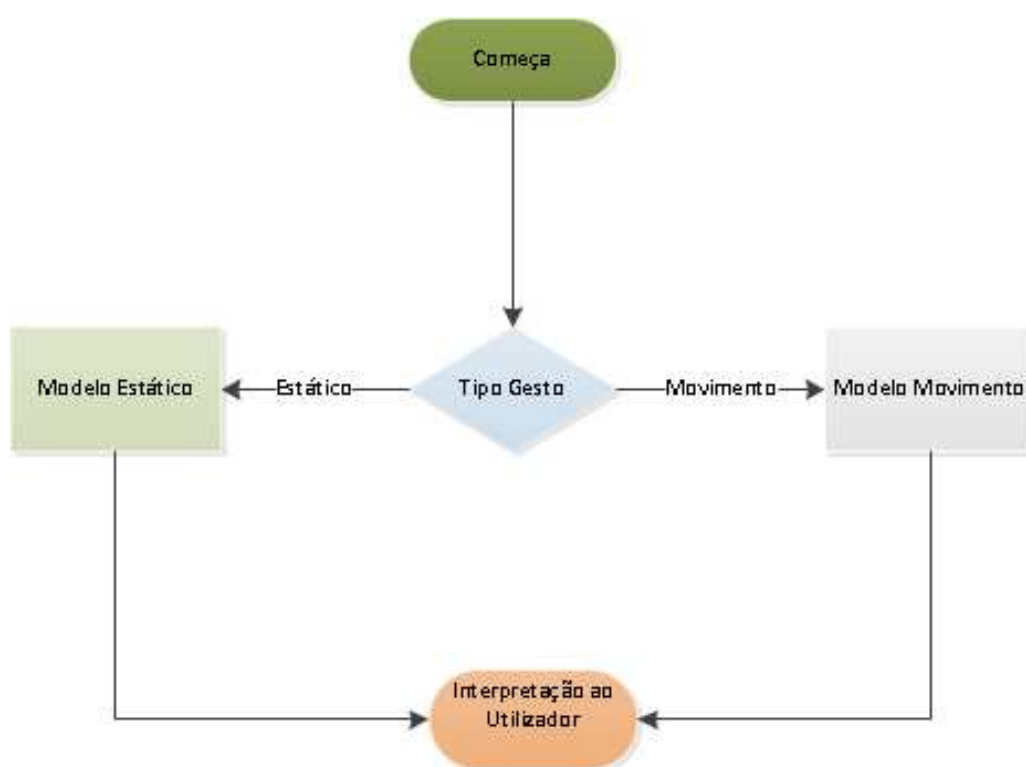


Figura 40 - Modelo Geral - Gesto para Texto

Portanto, como mostra a **Figura 40**, o utilizador, ao entrar na aplicação, terá a possibilidade de dizer se pretende gravar dados de treino ou se quer classificar novos gestos. Se passar pela gravação de novos gestos, se identifica o gesto – estático ou em movimento – que se quer fazer e começa a gravação.

4.2 Juntas de Interesse

O Xbox 360 Kinect pode controlar até 20 posições conjuntas. Sendo que com a finalidade que se tem de reconhecimento da linguagem de sinais temos demasiados pontos, portanto escolheu-se apenas aqueles que tem um peso significativo para a tarefa que se tem em frente.

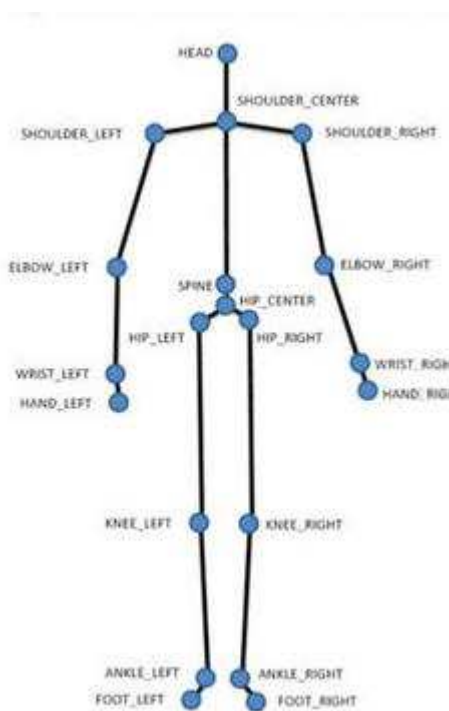


Figura 41 - Juntas de Interesse

Após um estudo cuidadoso dos sinais do dicionário proposto para o sistema, de gestos estáticos ou seja de letras do alfabeto, apenas quatro juntas foram selecionadas das vinte possíveis para que os resultados destas mesmas tenham um peso significativo para a descrição de um sinal. Portanto nesse sentido foram escolhidos as juntas da mão, do cotovelo e do ombro.

Não há nenhum sentido em acompanhar as outras articulações como joelhos, pés, anca, etc., porque são juntas que permanecem quase estáticas durante a execução do gesto. De outra forma, adicionando estas juntas, para descrever um sinal, estaríamos a adicionar informação redundante.

Para os gestos estáticos, ou seja, gestos das letras do alfabeto da LGP, não será importante adicionar as juntas da cabeça ao contrário nos poderia ajudar muito se as utilizássemos na configuração de gestos em movimento.

4.3 Implementação baseado em Ficheiro

O sistema é construído na sequência de uma implementação baseada em ficheiros, neste caso ficheiros de texto. Isto significa que cada gesto é armazenado como um conjunto de valores dos pontos selecionados.

No caso, se o gesto que foi feito pelo utilizador tiver a indicação que pertence ao conjunto de dados de teste, é gravado então o gesto – conjunto de valores do Kinect e da luva de dados – no ficheiro de teste. Caso o gesto efetuado pelo utilizador seja feito para a classificação é despoletado então processo, podendo ser um gesto estático ou um gesto em movimento

Portanto de cada junta escolhida – mão (duas juntas), cotovelo, e ombro – para a configuração do gesto retiramos as coordenadas do gesto como mostra a **Figura 42**.

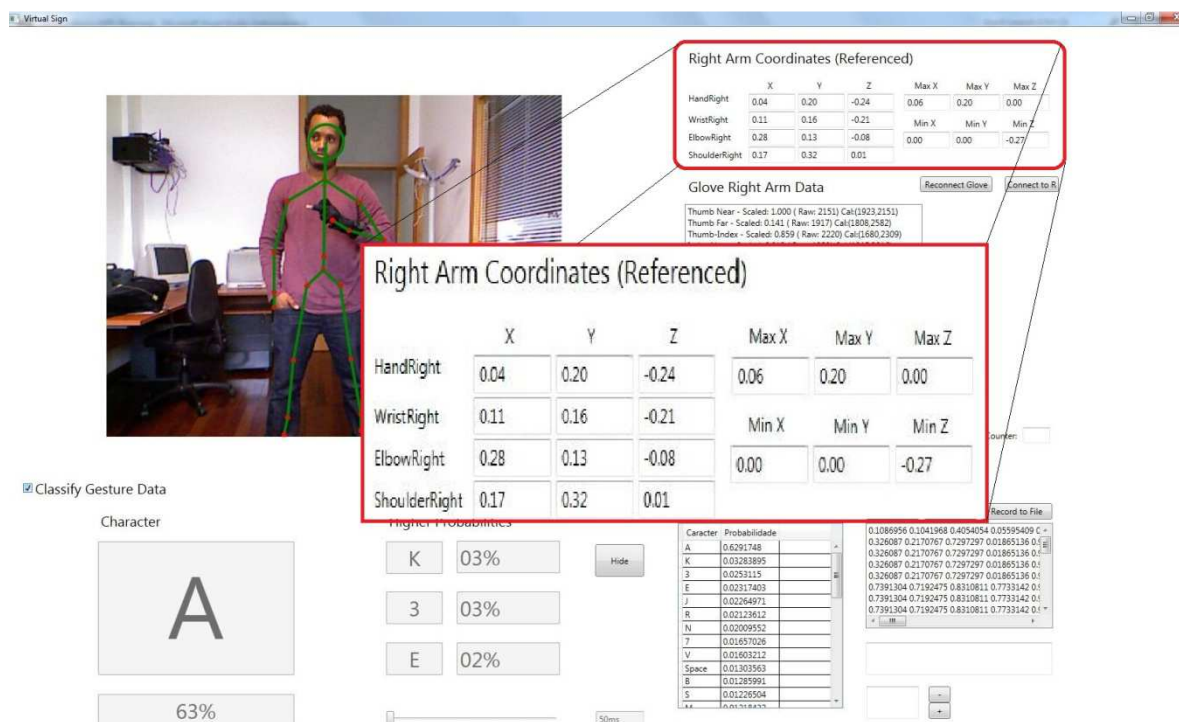


Figura 42 - Aplicação Desenvolvida - Dados da Mão Direita - Kinect

A imagem nos ajuda a perceber quais os dados retirados, de que mão é retirada os gestos e quais os pontos de interesse de cada junta.

Os dados de valor máximo de cada uma dos valores de X Y e Z das juntas de interesse foram também consideradas como mostra a figura, **Figura 42**, mas não foram aplicadas por não acrescentarem valor ao conjunto de dados retirados do gesto.

No sistema o utilizador é o único que pode escolher em que modo quer trabalhar, sendo que os dados do utilizador são gravados numa pasta pré configurada, como já foi explicado:

Fase de Calibração: O utilizador pode ficar em frente ao Kinect a ver os gestos que estão a ser feitos, sem se preocupar se estão sendo gravados, apenas com a intenção de treinar os gestos da LGP, seja ela estático ou em movimento. Tem também a possibilidade nessa fase de efetuar a calibração da Luva de Dados utilizada. Esta calibração é necessária e é feita efetuando um conjunto de ações para que os dados retirados da luva de dados sejam o mais preciso possível sobre a configuração manual.

Fase de Gravação: Um utilizador, ajudante, irá então clicar no botão de gravação de dados que irá permitir efetuar a gravação do gesto efetuado pelo outro utilizador. O gesto começa a partir do momento que o utilizador ajudante clica no botão de gravação de gestos e acaba quando o utilizador que faz o gesto coloca a mão na posição inicial, ou seja com a mão esticada em direção ao chão.

Fase de Teste: O utilizador antes de efetuar o gesto e já depois de ter feito a calibração da luva de dados, indica na aplicação que quer efetuar a classificação. É então despoletado a classificação do gesto que neste caso dividimos em dois métodos. Método dos Gestos Estáticos e o Método dos Gestos Dinâmicos.

4.4 Método – Gestos Estáticos

O reconhecimento de gestos estáticos é a primeira parte do nosso projeto que consiste num conjunto de processos que serão detalhados neste capítulo.

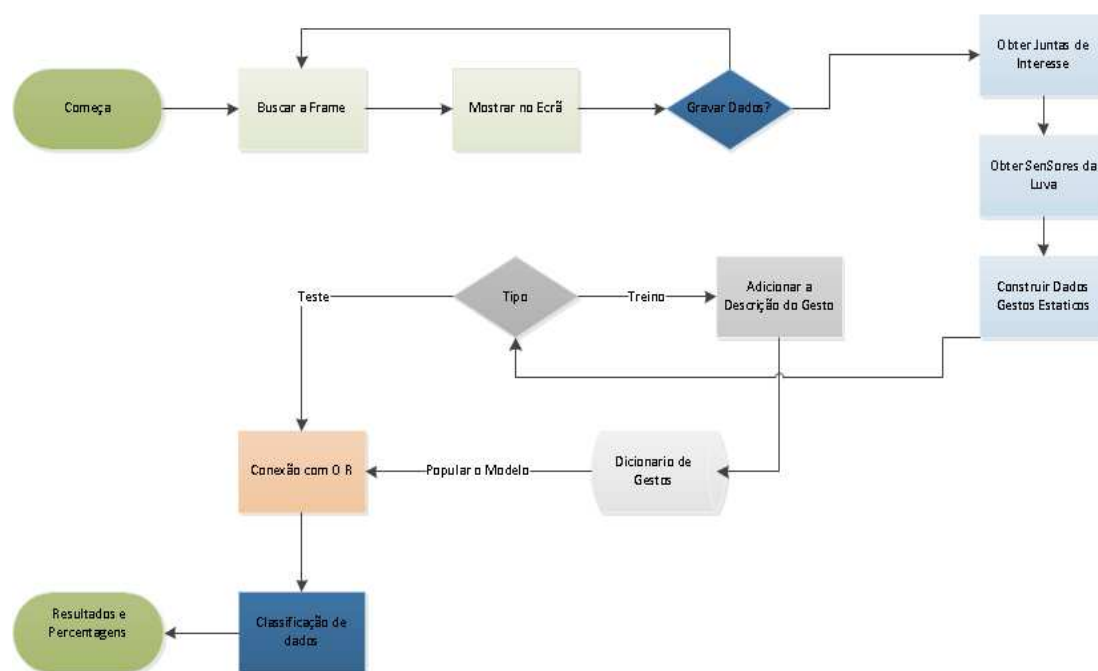


Figura 43 - Modelo Captura Dados e Reconhecimento - Gestos Estáticos

Nesta primeira parte, tal como na segunda, é recolhido os dados através de um utilizador que vai fazendo os gestos enquanto, outro utilizador lhe indica qual o gesto a gravar. O utilizador irá gravar uma serie de gestos estáticos de uma mesma letra do alfabeto, repetindo este procedimento por todas as letras.

Depois da gravação destes dados e depois de serem normalizados, é portanto importado, estes registos para um classificador.

O processo a seguir é o de reconhecimento destes gestos, ou seja, com os dados para classificação já no classificador, o utilizador irá fazer um gesto para ser reconhecido e o programa irá nos mostrar o resultado desse procedimento.

4.4.1 Luva de dados e Kinect

O sistema criado deve ser robusto para que possa fornecer o resultado de reconhecimento para qualquer gesto efetuado. Neste sentido, utilizou-se as duas ferramentas já faladas que nos dá uma maior segurança.

A luva de dados utilizada nos veio trazer de uma forma mais rápida e fácil de integrar com os dados do Kinect, valores que nos ajudaram a classificar os dados. Para isso trabalhamos com a Luva 5DT que como já explicamos tem 14 sensores, dois para cada dedo e mais um para a separação de cada um dos mesmos.

Dessa forma, conseguimos retirar da luva de dados, valores referentes aos tais falados 14 sensores que nos ajudarão a identificar a configuração manual de cada dedo. Retiramos esses valores que estão já escalados de forma a poderem ser utilizados, com oito casas decimais cada.

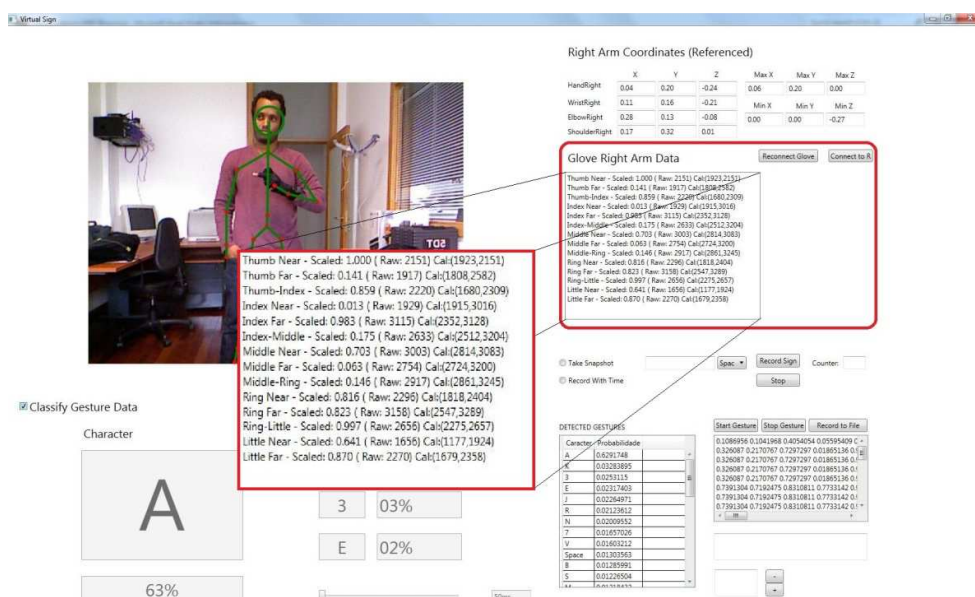


Figura 44 - Aplicação Desenvolvida Dados da Luva

Conseguimos dessa forma, retirar os valores de todos os dedos da mão referenciados desta forma:

- Valores dos sensores que estão mais perto da ponta do dedo:
 - Thumb Near – Polegar mais perto;
 - Index Near – Indicador mais perto;
 - Middle Near – Médio mais perto;
 - Ring Near – Anular mais perto;
 - Little Near – Mínimo mais perto.
- Valores dos sensores que estão mais perto do pulso:
 - Thumb Far – Polegar mais longe;
 - Index Far – Indicador mais longe;
 - Middle Far – Médio mais longe;
 - Ring Far – Anular mais longe;
 - Little Far – Mínimo mais longe.
- Valores dos sensores que estão entre os dedos:
 - Thumb Index – Entre o Polegar e o Indicador;
 - Index Middle – Entre o Indicador e o Médio;
 - Middle Ring – Entre o Médio e o Anular;
 - Ring Little – Entre o Anular e o Mínimo.

4.4.2 Normalização dos Dados

Depois de retirar os dados da luva e para que o sistema fique mais robusto e possa dar uma melhor resposta, aproveitamos os dados do Kinect e fizemos uma concatenação aos dados da luva.

Conseguimos então para um único gesto estático construir um vetor de 14 valores dos sensores da mão mais os valores das quatro juntas que consideramos importante para o reconhecimento do gesto. Aproveitamos portanto valores das coordenadas da mão (mão e pulso), do cotovelo e do ombro.

Desta forma conseguimos construir um vetor de um total de 23 coordenadas, sendo 9 do Kinect e 14 da luva de dados que depois ao serem gravados na base de dados, neste caso num ficheiro “txt”. A essas coordenadas serão adicionados mais dois, o utilizador que fez os gestos e a designação dos gestos.

Em termos de normalização dos dados tendo em conta a altura e a posição ou distanciamento do utilizador em relação a localização do Kinect não a consideramos por pensarmos que os dados da luva neste caso de um gesto estático serão os mais importantes visto que são esses dados que nos vão ajudar mais na configuração da mão. Os dados do Kinect irão ajudar mais em saber a posição da mão do utilizador que será mais importante num gesto em movimento.

4.4.3 Classificador

Para a classificação dos dados, depois de estes serem normalizados, utilizamos através da aplicação, uma conexão feita ao R.Net que permite a “.Net Framework” colaborar com a ferramenta de computação de estatística R.

Antes, vamos abordar esta ferramenta que nos vai ajudar com a utilização dos SVM's – máquinas de vetores de suporte – que já foi descrito no capítulo 3 – para a classificação dos nossos dados utilizando a função *kernel* RBF.

Ferramenta R

R, de acordo com R – Project.org, é uma linguagem e ambiente para computação estatística e de gráficos. É um projeto GNU, que é similar a linguagem e ambientes S, que foi desenvolvido nos Laboratórios BELL.

R fornece uma ampla variedade de técnicas estatísticas – modelagem linear e não linear, testes estatísticos clássicos, classificação, *clustering* ... – e técnicas gráficas e é altamente extensível.

A sintaxe desta linguagem tem alguma similaridade com a linguagem C mas a semântica é a de FPL que traduzido para português significa linguagem de programação funcional. Em particular, permite “computação na linguagem”, o que por sua vez é possível escrever funções que recebem como parâmetro expressões, algo que muitas vezes é útil para a modelação estatística e para fazer gráficos.

Sendo possível fazer muito mais utilizando o R interactivamente executando expressões simples a partir de uma linha de comando.

Em cada linguagem de programação, variáveis providenciam um meio de acesso aos dados guardados em memória. O R não providencia acesso direto a memória do computador, em vez disso providencia uma lista de estrutura de dados especializados que se referem como objetos que podem ser depois utilizadas em expressões para interpretação do R.

Depois desta breve introdução sobre a ferramenta R importa dizer que a utilizamos no sistema **Virtual Sign** através de uma “DLL” de conexão R.NET.

E utilizamos da seguinte forma. No código dizemos qual é pasta onde se encontra a biblioteca com a DLL que possibilita a utilização da ferramenta R. Depois disso, criamos uma nova instância do RDOTNET. Inicializamos esta instância e lhe dizemos para nos abrir o modelo já feito no R com os dados do treino. Fazemos isso utilizando a função Load e enviando como parâmetro o nome do modelo que queremos, neste caso o modelo criado tem o nome de SVMModelGloveKinect.rdata.

Este ficheiro é um conjunto de funções previamente executadas no R e guardadas para que possam ser chamadas na utilização do mesmo.

É chamada a função do R para avaliar o texto que passamos em parâmetro para fazer o carregamento do ficheiro que por sua vez contem as funções para a classificação dos novos dados.

```
REngine.SetDllDirectory (@"C:\Program Files\R\R-3.0.0\bin\i386");  
  
engine = REngine.CreateInstance ("RDotNET");  
  
engine.Initialize ();  
  
string rdata = "load (" + "\" + "SVMModelGloveKinect.rdata" + "\" + ")";  
  
try  
  
{ engine.Evaluate (rdata);  
  
} catch (RDotNet.ParseException exe) {}
```

Depois de se ter feito o carregamento do modelo é preciso dizer qual o pacote que contem o algoritmo de aprendizagem que se vai utilizar na gravação dos dados. Nesta aplicação trabalhamos com o pacote e1071 que é o pacote para utilização do SVM.

```
SymbolicExpression s_express2 = engine.Evaluate ("require (e1071) ");
```

O objectivo era o de conseguir a classificação dos dados através da ferramenta R e para isso utilizamos as seguintes linhas de comando:

- *SymbolicExpression s_y = engine.Evaluate ("abc<-c(" + str + ")");*

O que estamos a fazer com esta linha de código é dizer a ferramenta R para avaliar a expressão que estamos a inserir. A avaliação desta expressão nos retorna um "SymbolicExpression". Mas antes vamos interpretar a expressão que acabamos por inserir.

A expressão, *abc <-c (" +str+ ")* , leva como parâmetro o texto que compõe o gesto feito pelo utilizador em valores. O que passamos como parâmetro é uma "string", contendo todos os dados da luva de dados, com a designação de rg0 à rg13 que são os valores dos 14 sensores da luva. Ainda nesta "string", adicionamos os valores que retiramos do **Kinect**, das juntas de interesse do esqueleto. Para relembrar enviamos os valores da mão, do pulso, do cotovelo e do ombro. Este conjunto de dados do **kinect** nos dá um total de doze coordenadas X Y e Z que adicionamos ao novo gesto, em forma de texto com valores, para a classificação.

Depois disso esse texto é transformado num vector de dados com a função "c ()" para que possa ser utilizado numa variável a que chamamos de "abc". Esta variável irá conter o vector de dados do gesto.

- *SymbolicExpression s_ex = engine.Evaluate ("abc");*

Esta linha de comando é utilizado para fazer a avaliação do nosso vector de dados que depois de executado retorna uma expressão. Se essa expressão estiver vazia é porque houve algum problema com os dados e o nosso gesto não pode ser utilizado para classificação. Ao contrário se nos enviar valores de sucesso na avaliação poderemos utilizar o modelo previamente carregado onde dizemos que tem que nos retornar a decisão sobre que gesto mais se assemelha ao gesto feito e a probabilidade do nosso gesto ser algum dos outros gestos de treino que foram previamente carregados com o modelo.

Para fazer isso é necessário utilizar a função “predict” que recebe como parâmetros o nosso modelo, o gesto que fizemos e as duas variáveis para saber a decisão e a probabilidade, carregando o resultado para a variável “previsions”.

- *SymbolicExpression s_exp3 = engine.Evaluate("previsions <- predict(model, rbind(abc), decision.value=TRUE, probability=TRUE)");*

Depois de ter executado a função “predict” a função Evaluate da nossa instância de Engine nos envia uma resposta para desta forma sabermos se a variável “previsions” foi carregada com dados ou não.

Desta forma executamos a linha de código seguinte que nos faz a avaliação da variável “previsions” para a confirmação de que tem resultados.

- *SymbolicExpression s_ess4 = engine.Evaluate("previsions");*

A próxima linha de código nos vai servir para dividirmos o resultado entre a previsão e o resultado da probabilidade do nosso gesto ser qualquer um dos outros gestos, visto que na última linha de código acabamos por colocar todos esses valores apenas na variável “previsions”.

- *SymbolicExpression s_es4 = engine.Evaluate ("attr(previsions, \"probabilities\")");*

Depois de ter dividido o resultado em dois conjuntos colocamos os dados de avaliação numa tabela para que possam ser organizados e disponibilizados para o utilizador.

Apenas para nos retornar os resultados como um vector de dados adicionamos esta linha de código para ser mais fácil de interpretar:

- *SymbolicExpression s_express4 = engine.Evaluate("predicted <- as.vector(previsions)");*

Os resultados são então transportados para o interface do utilizador, onde apresentamos qual o gesto reconhecido, qual a sua percentagem e quais os três gestos mais próximos do gesto reconhecido.



Right Arm Coordinates (Referenced)

	X	Y	Z	Max X	Max Y	Max Z
HandRight	0.04	0.20	-0.24	0.06	0.20	0.00
WristRight	0.11	0.16	-0.21	Min X	Min Y	Min Z
ElbowRight	0.28	0.13	-0.08	0.00	0.00	-0.27
ShoulderRight	0.17	0.32	0.01			

Glove Right Arm Data

Reconnect Glove Connect to R

Thumb Near - Scaled: 1.000 (Raw: 2151) Cal:(1923,2151)
 Thumb Far - Scaled: 0.141 (Raw: 1917) Cal:(1808,2582)
 Thumb-Index - Scaled: 0.859 (Raw: 2220) Cal:(1680,2309)
 Index Near - Scaled: 0.013 (Raw: 1929) Cal:(1915,3016)
 Index Far - Scaled: 0.983 (Raw: 3115) Cal:(2352,3128)
 Index-Middle - Scaled: 0.175 (Raw: 2633) Cal:(2512,3204)
 Middle Near - Scaled: 0.703 (Raw: 3003) Cal:(2814,3083)
 Middle Far - Scaled: 0.063 (Raw: 2754) Cal:(2724,3200)
 Middle-Ring - Scaled: 0.146 (Raw: 2917) Cal:(2861,3245)
 Ring Near - Scaled: 0.816 (Raw: 2296) Cal:(1818,2404)
 Ring Far - Scaled: 0.823 (Raw: 3158) Cal:(2547,3289)
 Ring-Little - Scaled: 0.997 (Raw: 2656) Cal:(2275,2657)
 Little Near - Scaled: 0.641 (Raw: 1656) Cal:(1177,1924)
 Little Far - Scaled: 0.870 (Raw: 2270) Cal:(1679,2358)

Gesto Reconhecido

Outras Probabilidades

Classify Gesture Data

Character

A

63%

Higher Probabilities

K	03%
3	03%
E	02%

50ms

Take Snapshot Spac Record Sign Counter: Stop

Record With Time

DETECTED GESTURES

Caracter	Probabilidade
A	0.6291748
K	0.03283895
3	0.0253115
E	0.02317403
J	0.02264971
R	0.02123612
N	0.02009552
7	0.01657026
V	0.01603212
Space	0.01303563
B	0.01285991
S	0.01226504
M	0.01184322

Start Gesture Stop Gesture Record to File

0.1086956 0.1041968 0.4054054 0.05595409 C
 0.326087 0.2170767 0.7297297 0.01865136 0.0
 0.326087 0.2170767 0.7297297 0.01865136 0.0
 0.326087 0.2170767 0.7297297 0.01865136 0.0
 0.326087 0.2170767 0.7297297 0.01865136 0.0
 0.7391304 0.7192475 0.8310811 0.7733142 0.0
 0.7391304 0.7192475 0.8310811 0.7733142 0.0
 0.7391304 0.7192475 0.8310811 0.7733142 0.0

Figura 45 - Aplicação Desenvolvida - Reconhecimento em Tempo Real

4.5 Modelo – Gestos em Movimento

O reconhecimento de gestos em movimento é a segunda parte do nosso projeto que consiste num conjunto de processos que serão detalhados neste capítulo.

Na segunda parte, os dados são recolhidos através de um utilizador que vai fazendo os gestos enquanto outro utilizador lhe indica qual o gesto a gravar. O utilizador irá gravar uma serie de gestos em movimento de uma mesma palavra, repetindo este procedimento por todas as palavras que definimos para classificação.

Depois da gravação destes e já depois dados serem normalizados, é portanto importado, estes registos para um classificador.

O processo a seguir é o de reconhecimento destes gestos, ou seja, com os dados para classificação já no classificador, aqui o utilizador irá gravar os gestos em movimento de testes. Estes dados também irão ser importados para o classificador e irão ser indicados como dados de teste, diferente dos muitos registos de dados de treino que já foram inseridos para classificação.

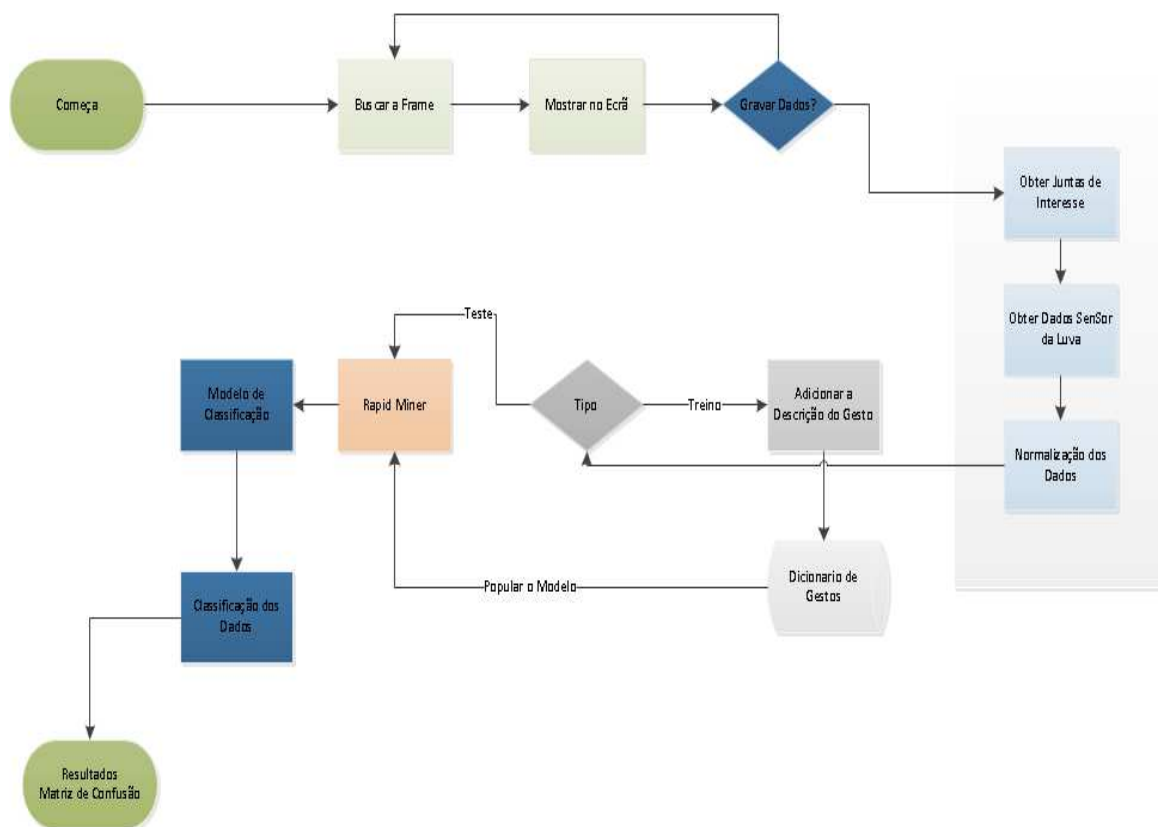


Figura 46 - Modelo Reconhecimento - Gestos em Movimento

4.5.1 Normalização dos Dados

A aplicação deve ser robusta para capturar o movimento de cada gesto efetuado pelo utilizador. Desta forma se tornando num dos pontos indispensáveis na descrição do gesto e de futuramente considerar o tamanho do utilizador e a sua posição.

Contudo para a aplicação que temos no momento, dado um gesto, a normalização consiste em:

- Duração do Gesto;
- Divisão do Gesto Efetuado;
- Construção do vetor de dados (descrição do gesto).

Duração do Gesto

O primeiro ponto, a duração do gesto, é um dos aspetos que consideramos importante na descrição do gesto. Isso porque, nos muitos gestos que existem na LGP, há os gestos mais simples que podem ser feitos em pouco espaço de tempo e os gestos que nas primeiras vezes nos levaria mais tempo a fazer. Sendo que mesmo o utilizador experiente, o tempo que leva a fazer um gesto que significa “Olá” levará menos tempo do que um gesto que significa “Floresta”. Isto porque há gestos com um movimento da mão como o gesto que significa “Olá” e há outros gestos que consistem em vários movimentos da mão ou das mãos como é o caso da palavra “Floresta”.

Pelas razões já explicadas consideramos então a duração do gesto como um das características que faz parte da descrição de cada gesto.

Divisão do Gesto Efetuado

O segundo ponto, a divisão do gesto efetuado por um utilizador tem a ver com a simplicidade do gesto e a forma como essa descrição é feita. Por exemplo para um gesto que significa “Olá”, que é apenas um movimento da mão direita, importa guardar as *frames* que compõem o gesto apenas. Mas para um gesto que signifique “Floresta”, ou seja um gesto mais complexo que implica mais movimentos por parte do utilizador, é importante dividir onde o gesto começa a diferenciar-se dos restantes e guardar os dados dos movimentos mais importantes para o mesmo.

Dessa forma, para cada gesto, optamos então por gravar as trinta *frames* por segundo, durante o tempo que o gesto é efetuado. Se por exemplo, um gesto tem uma duração de três segundos, nesses três segundos teremos um total de noventa *frames* guardados. Esse conjunto de *frames* será dividido em duas partes iguais. Cada uma dessas partes será tratada de forma diferente. A primeira parte que não irá ter a parte principal do gesto, contribuirá com cinco *frames*, enquanto a segunda parte ficará com os restantes quinze *frames* de um total de vinte *frames* que consideramos ser importante para a classificação do gesto.

Da primeira parte, iremos retirar a primeira e a última *frame* e os restantes três *frames* são retirados do restante que ficar do total de *frames* desta parte.

Da segunda parte, é onde consideramos que devemos retirar mais dados por ser onde se começa a detalhar o gesto efetuado. Desta forma, retiramos quinze *frames* deste conjunto de dados.

Retirando a primeira e a última *frame* ficamos com treze *frames* para retirar do conjunto de dados que serão retirados dividindo o restante das *frames* por treze, desta forma encontrando um número que nos serve de paragem e coleção dos dados da *frame* ao percorrer o conjunto de *frames*.

Portanto neste exemplo temos:

- 3 Segundos – duração do gesto;
- 90 *Frames* – descrição do gesto:
 - Primeiro conjunto 45 *Frames*:
 - 5 *Frames*, retirados para a descrição.
 - Segundo conjunto 45 *frames*:
 - 15 *Frames*, retirados para a descrição.

Construção do vetor de dados (descrição do gesto)

A descrição de cada gesto efetuado pelo utilizador irá conter as informações que vem da luva de dados mais a informação que vem do Kinect. Portanto teremos por *frame* a informação da luva de dados onde retiramos os valores de 14 sensores e ainda teremos os valores do Kinect com as juntas de interesses como já foi explicado no capítulo Gestos Estáticos.

Essa informação, ou seja, estes vinte e três registos serão multiplicadas depois pelas vinte *frames* que consideramos importantes para termos uma descrição que irá contribuir para uma boa classificação do gesto.

Portanto o nosso vetor de dados – descrição do gesto – irá conter um total de características de quatrocentos e sessenta. Irá conter mais uma característica adicional que é a duração que o utilizador demorou a fazer o gesto.

4.5.2 Invariante a posição do utilizador

O modelo deve contar com um ponto muito importante na recolha dos dados do utilizador. Este dado é o de que o gesto efetuado pelo utilizador, ou seja, a descrição do gesto ou dos gestos feitos a uma longa distância deve ser idêntico do gesto feito mais perto possível do dispositivo Kinect. O utilizador pode estar a em diferentes posições numa sala e portanto os dados devem ser gravados de acordo com essa posição.

Apesar de não estarmos considerando esta informação na aplicação desenvolvida pelo pouco tempo que tivemos, esta informação deverá ser considerada. E deverá ser considerada também para que não aconteça ter que fazer cada um dos gestos de diferentes posições e enviar essa informação para o classificador desta forma atrasando a performance do mesmo. A descrição dos gestos deve conter no seu valor a normalização para a posição do utilizador numa sala.

4.5.3 Invariante ao tamanho do utilizador

Dado um gesto, a sua descrição deve ser o mesmo, não importando para isso se o utilizador é mais pequeno ou mais alto e o modelo deve ser capaz de traduzir a palavra certa em qualquer caso. Mesmo que o dicionário de dados seja feito para permitir muitos exemplos para o mesmo gesto, ou seja iremos ter neste caso signos iguais mas com descrições que são diferentes pelo tamanho do utilizador, é muito difícil de guardar dados para todos os tamanhos de utilizadores possíveis em termos de tamanho. Se também fosse possível, a classificação dos dados, ou seja o processo de classificação iria ser cada vez mais lento e com menos precisão.

O tratamento da informação que deve sair para o utilizador deve ser portanto invariante ao tamanho do utilizador, ou seja, devemos de alguma forma considerar o tamanho e adicionar aos dados de classificação. Este é um trabalho que não foi feito para os dados de classificação que temos no momento, mas que deveria ser um ponto importante no tratamento dos dados do utilizador.

4.5.4 Classificador

Ferramenta Rapid Miner

A ferramenta Rapid Miner, “rapid-i”, é descrita como um aplicativo, livre, independente para análise de dados e como um mecanismo de mineração de dados, Data Mining.

Portanto, permite:

- Integração e Análise dos dados;
- Execução de Relatórios sobre a análise e integração dos dados;
- Diferente do “R”, tem um especto gráfico que permite desenhar como será feita a análise e a execução do processo.

O processo de classificação dos dados, utilizando a ferramenta Rapid Miner compõe-se em 3 fases, como mostra as figuras seguintes:

1. Leitura do Ficheiro de dados;
2. Normalização dos dados;
3. Processo de Validação e Classificação.

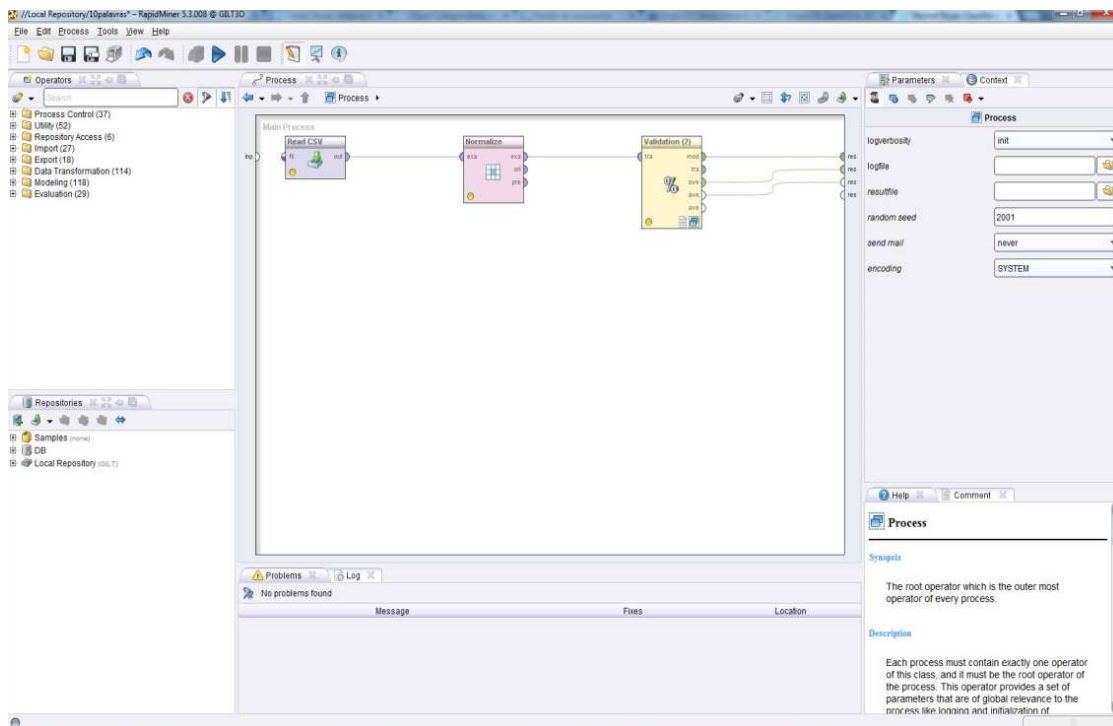


Figura 47 Rapid Miner - Classificação dos Gestos

4.5.5 Leitura do Ficheiro de Dado

O primeiro passo é escolher o ficheiro de dados a que vamos fazer a classificação. Esse processo consiste em ir buscar o ficheiro de dados, configurada de forma que possa ser interpretada pelo Rapid Miner. Esse ficheiro deve ser construído de forma que cada linha do ficheiro – CSV – tenha o mesmo número de colunas. De outra forma o Rapid Miner nos indica que o ficheiro se encontra com erros.

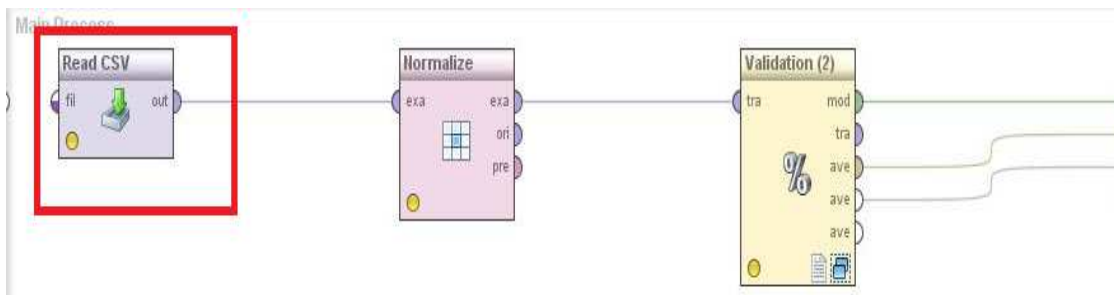


Figura 48 - Rapid Miner - Leitura Ficheiro de Dados

Estes dados que estão no ficheiro CSV, tem como ultima coluna a palavra ao que os dados se referem. Inicialmente o nome dessa coluna era “Label” mas ao carregar para o Rapid Miner, tivemos que mudar o nome da coluna para “Classe” visto que essa designação já é utilizada pela ferramenta.

4.5.6 Normalização dos Dados – Rapid Miner

Este passo do processo de classificação dos dados é uma técnica utilizada de forma que os valores de cada coluna do ficheiro “csv” estejam sempre dentro de uma determinada escala, no caso de 0 a 1.

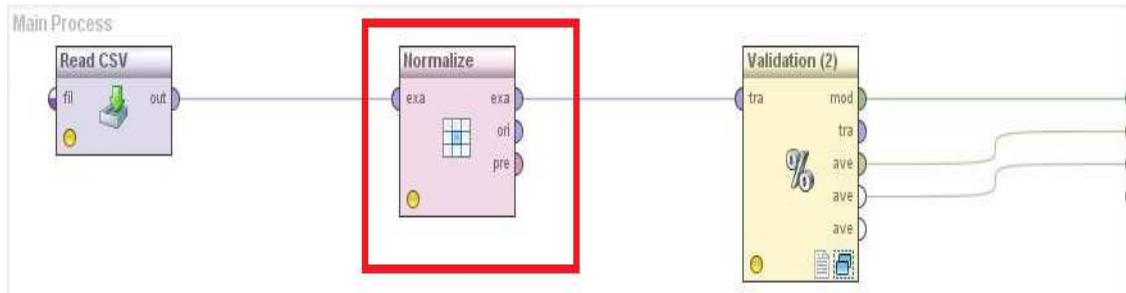


Figura 49 - Rapid Miner - Normalização dos Dados

A normalização dos dados é muito importante quando se trata de dados de diferentes unidades e escala. Por exemplo, algumas das técnicas de mineração de dados utiliza a distância Euclidean. Por isso todos os dados, ou no nosso caso todos os exemplos de dados letras ou palavras, devem ter a mesma escala para que a comparação seja a mais justa possível.

Noutras palavras a normalização é uma técnica utilizada para nivelar o campo de jogo quando se olha para os atributos que variam amplamente em tamanho como resultado das unidades selecionadas para a representação.

4.5.7 Processo de Validação e Classificação

O processo de validação de dados é onde aplicamos o algoritmo de classificação propriamente dita. Neste capítulo de Gestos em movimento também utilizamos o mesmo algoritmo de classificação que utilizamos para a classificação dos gestos estáticos.

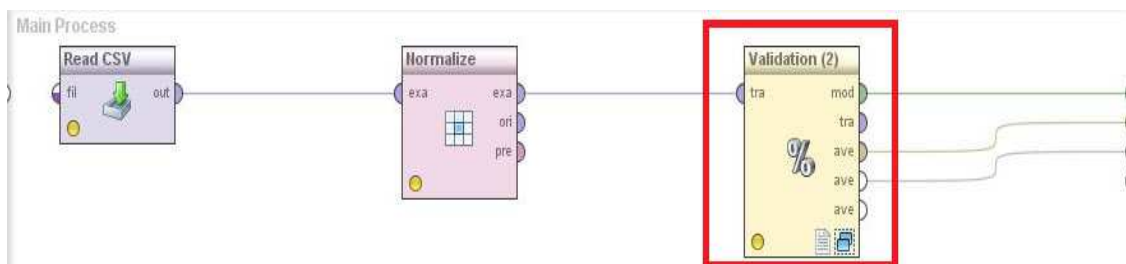


Figura 50 - Radpi Miner - Processo de Classificação e Validação

Neste caso o Rapid Miner aplica uma validação cruzada de forma a estimar a performance estatística da aprendizagem, permitindo estimar o quão preciso é o modelo em reconhecer os gestos na prática.

Este processo de validação consiste em dois subprocessos:

1. Processo de Treino – utilizada para efetuar o treinamento do modelo.
2. Processo de Teste – é onde o modelo treinado no processo de treino é aplicado. A performance do modelo escolhido é também medida neste subprocesso.

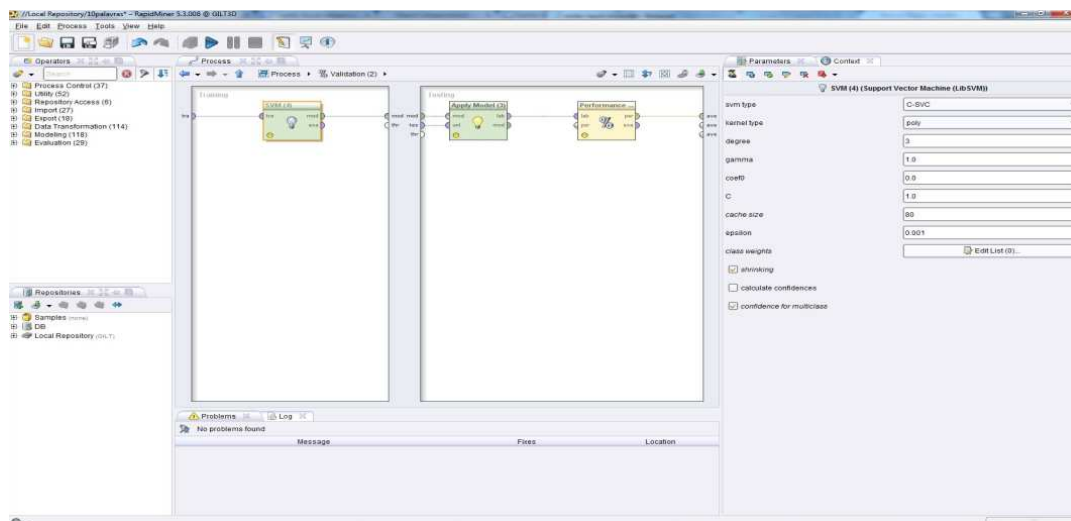


Figura 51 - Rapid Miner - Processos

Os dados inseridos para classificação são particionados em K subconjuntos de tamanho igual. Desses subconjuntos, apenas um é retido para ser testado sendo que os restantes K-1 subconjuntos são utilizados para treino.

O processo de validação cruzada é repetido o mesmo número de vezes que compõem os subconjuntos, com cada um destes sendo utilizados uma vez como sendo dados de teste.

Através dos resultados que vem das iterações pode então ser calculada uma média para produzir uma única estimativa.

Na fase de treino, no modelo escolhido, podemos configurar através de alguns parâmetros valores que melhor classificam os dados. Depois de vários testes, para os dados inseridos e para o modelo escolhido – SVM – os parâmetros ficaram com estes valores:

- Tipo de SVM: C-SVC – classificação por SVM regularizada;
- Tipo de *Kernel*: Polinomial – *kernel* que utiliza a fórmula “ $(\text{gama} * u' * v + \text{coef0})^{\text{Grau}}$ ”.
- Grau: 3 – o grau da função polinomial;
- Gama: 1.0 – um parâmetro multiplicativo para a função polinomial;
- Coef0: 0.0 – um parâmetro aditivo para a função polinomial;
- C: 1.0
- Tamanho de cache: 80
- Épsilon: 0.001
- Encolher: verdadeiro

- Confiança para várias classes: verdadeiro.

4.5.8 Resultados

Tendo construído todo o processo de classificação de dados, o próximo passo é o de executar o mesmo, ou seja, no Rapid Miner há a opção de executar o processo em que estamos. Esta opção está identificada com uma imagem clara de execução, uma seta verde apontada a direita.

Isso nos irá fornecer os resultados da execução do processo, onde também nos irá informar sobre o desempenho do modelo de classificação. Ao referir o desempenho do modelo, estamos interessados na capacidade do modelo para prever corretamente ou fazer a separação das classes.

Ao olhar para os erros cometidos pelo modelo que adotamos, a matriz de confusão dá a imagem completa.

Contudo antes teremos uma imagem completa sobre como funciona o “Rapid Miner” nesta questão.

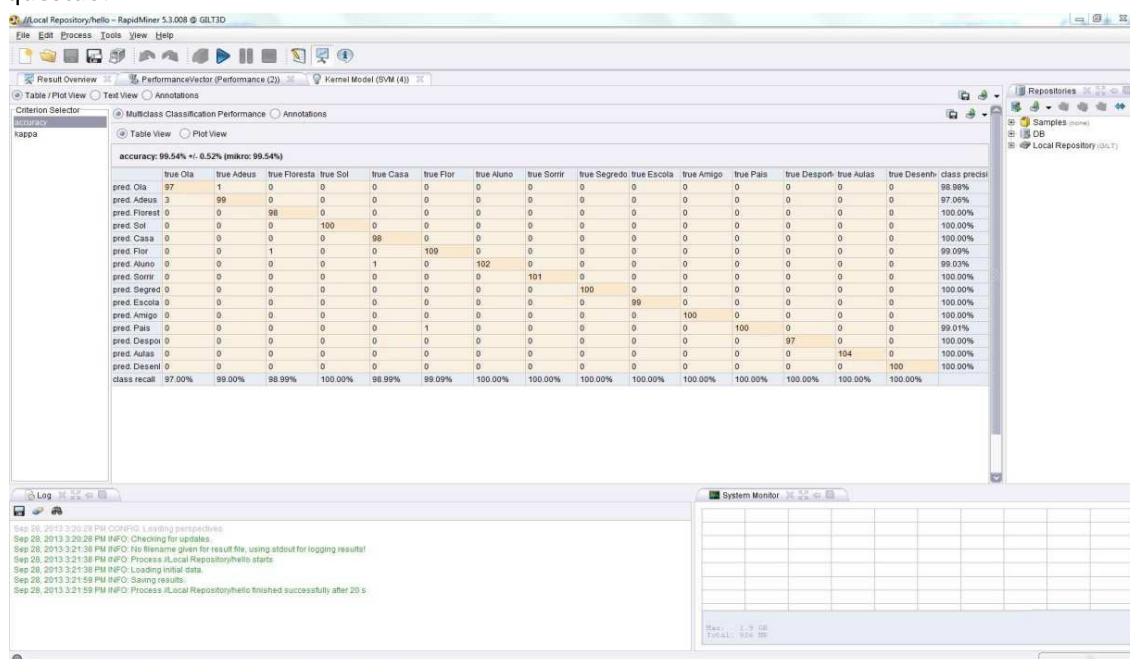


Figura 52 - Rapid Miner - Resultados

A matriz de confusão representada na **Figura 53**, para as palavras que selecionamos para testar o nosso modelo de gestos em movimento nos dá a informação sobre o modelo adotado. Relembrando, o modelo adotado utiliza a técnica de aprendizagem SVM's.

Esta matriz de confusão tem por base, 15 classes diferentes, representadas pelas palavras:

- Olá; Adeus; Sorrir; Segredo;
- Floresta; Sol; Flor;
- Aluno; Escola; Casa; Aulas; Desenho;
- Amigo; Pais; Desporto.

accuracy: 99.54% +/- 0.52% (mikro: 99.54%)																
	true Ola	true Adeus	true Floresta	true Sol	true Casa	true Flor	true Aluno	true Sorrir	true Segredo	true Escola	true Amigo	true Pais	true Desporti	true Aulas	true Desenh	class predisi
pred. Ola	97	1	0	0	0	0	0	0	0	0	0	0	0	0	0	98.98%
pred. Adeus	3	99	0	0	0	0	0	0	0	0	0	0	0	0	0	97.06%
pred. Florest	0	0	98	0	0	0	0	0	0	0	0	0	0	0	0	100.00%
pred. Sol	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	100.00%
pred. Casa	0	0	0	0	98	0	0	0	0	0	0	0	0	0	0	100.00%
pred. Flor	0	0	1	0	0	109	0	0	0	0	0	0	0	0	0	99.09%
pred. Aluno	0	0	0	0	1	0	102	0	0	0	0	0	0	0	0	99.03%
pred. Sorrir	0	0	0	0	0	0	0	101	0	0	0	0	0	0	0	100.00%
pred. Segred	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	100.00%
pred. Escola	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	100.00%
pred. Amigo	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	100.00%
pred. Pais	0	0	0	0	0	1	0	0	0	0	0	100	0	0	0	99.01%
pred. Despoi	0	0	0	0	0	0	0	0	0	0	0	0	97	0	0	100.00%
pred. Aulas	0	0	0	0	0	0	0	0	0	0	0	0	0	104	0	100.00%
pred. Desenh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	100.00%
class recall	97.00%	99.00%	98.99%	100.00%	98.99%	99.09%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	

Figura 53 - Rapid Miner - Matriz de Confusão

Cada gesto foi efetuado um número de vezes que dá para fazer uma comparação correta dos dados. Esses dados foram extraídos de gestos produzidos por diferentes utilizadores, 3 neste caso, e tentamos ter uma divisão dos registos de cada gesto igual para cada um desses utilizadores. Como se vê na **Figura 53**, utilizou-se para a palavra “Olá”, 100 registos. Para a palavra “Adeus” 102 registos; palavra “Floresta” 100 registos; palavra “Sol” 100 Registos; 100 registo para a palavra “Casa”; 110 para “Flor”; 102 para a palavra “Aluno”; “Sorrir” 101 Registos; “Segredo” 100 Registos; “Escola” 99 registos; 100 registo efetuados para a palavra “Amigo”; palavra “Pais” com 100 registos; Palavra “Desporto” com 97 Registos; Aulas com 104 Registos e Desenho com 100 registos.

A informação que a matriz nos dá é das medidas de “Recall” e “Precision” em relação a cada uma das classes e a medida “Accuracy” para o geral da classificação.

Sendo que “Recall” uma medida da capacidade de um modelo para selecionar casos de uma determinada classe de um conjunto de dados. Ela é definida pela fórmula:

$$“Recall = TP / (TP + FN) ”$$

Onde TP e FN são os números de previsões positivas verdadeiras e de falsos negativos para a classe considerada. O total dos registos de cada classe é identificado pela soma dos registos de TP e FN.

A outra medida, “Precision” para cada uma das classes é uma medida de precisão fornecida de uma classe específica que tem sido prevista. Esta medida é definida pela fórmula:

“Precision = TP/ (TP + FP) ”

Onde TP e FP são números verdadeiros de previsões positivas e falsas para a classe considerada.

“Accuracy” é a medida de precisão global do modelo e é calculada como a soma das classificações corretas dividido pelo número total de classificações.

Neste caso, tal como representado na **Figura 53**, temos uma precisão global de 99.54 % com uma taxa de erro de 0.52% para as 15 classes representadas, ou seja para os 15 gestos efetuados.

Para cada uma das palavras os resultados como mostram a **Figura 53**, são aceitáveis tendo em conta que utilizamos dois dispositivos para a captura dos dados.

Teremos uma tabela com o detalhe da performance do modelo utilizado.

Performance do Modelo	Valores Percentagem	Taxa Erro	Mikro
Accuracy:	99,54%	0,52%	99.54%
kappa:	0,995	0,006	0.995

Tabela 3 - Performance do Modelo

Classes	Olá	Adeus	Floresta	Sol	Casa	Flor	Aluno	Sorrir	Segredo	Escola	Amigo	Pais	Desporto	Aulas	Desenho	Class Precision
Olá:	97	1	0	0	0	0	0	0	0	0	0	0	0	0	0	98,98%
Adeus:	3	99	0	0	0	0	0	0	0	0	0	0	0	0	0	97,06%
Floresta:	0	0	98	0	0	0	0	0	0	0	0	0	0	0	0	100,00%
Sol:	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	100,00%
Casa:	0	0	0	0	98	0	0	0	0	0	0	0	0	0	0	100,00%
Flor:	0	0	1	0	0	109	0	0	0	0	0	0	0	0	0	99,09%
Aluno:	0	0	0	0	1	0	102	0	0	0	0	0	0	0	0	99,03%
Sorrir:	0	0	0	0	0	0	0	101	0	0	0	0	0	0	0	100,00%
Segredo:	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	100,00%
Escola:	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	100,00%
Amigo:	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	100,00%
Pais:	0	0	0	0	0	1	0	0	0	0	0	100	0	0	0	99,01%
Desporto :	0	0	0	0	0	0	0	0	0	0	0	0	97	0	0	100,00%
Aulas:	0	0	0	0	0	0	0	0	0	0	0	0	0	104	0	100,00%
Desenho:	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	100,00%
Class Recall	97,00 %	99,00%	98,99%	100 %	98,99 %	99%	100,00 %	100,00 %	100,00%	100,00 %	100,00 %	100,00 %	100,00%	100,00 %	100,00%	

Tabela 4 - Matriz de Confusão

4.6 Modelo Tradução Texto – LGP

O modelo geral da nossa aplicação inclui a tradução do texto para LG que futuramente será implementado. Tal como é extremamente importante a tradução dos gestos da Língua Gestual é também importante a tradução de texto para a Língua Gestual, visto que possibilita uma melhor comunicação para utilizadores que não saibam LG com utilizadores surdos que tem essa Língua como língua principal.

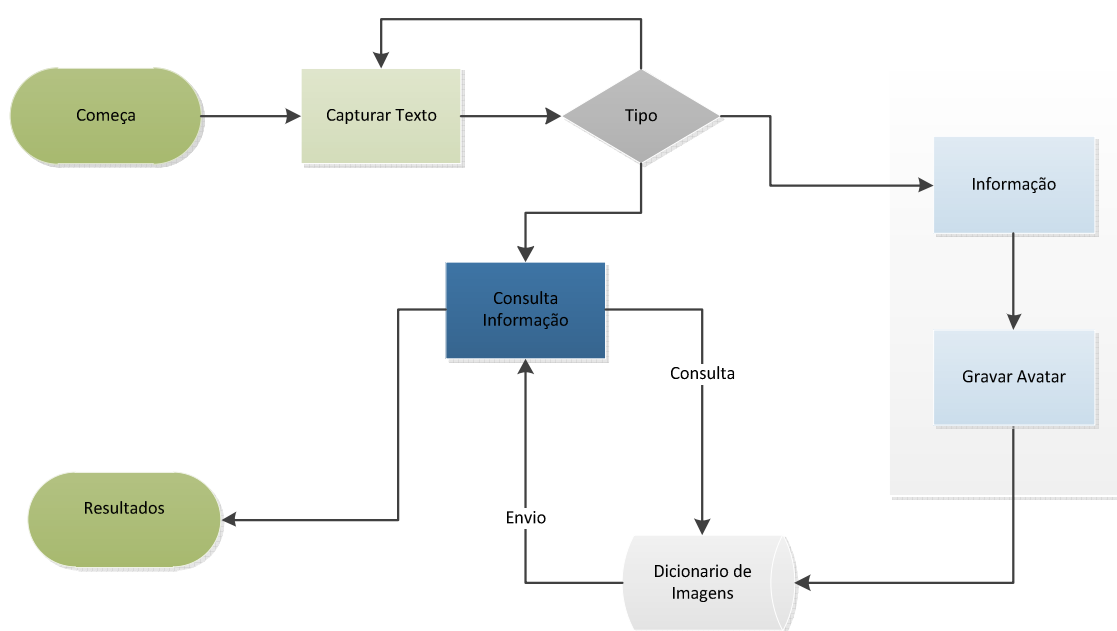


Figura 54 - Modelo Tradução Texto em LGP

Este modelo implicaria num momento inicial uma caixa de texto na nossa aplicação e um espaço para se poder ver o gesto utilizando um avatar para ver o gesto em movimento ou estático correspondente a uma palavra escrita por um utilizador.

1ª Fase – Guardar Informação e Avatar

Esta primeira fase será onde os utilizadores poderão gravar os dados numa base de dados e o processo inclui:

- Texto - Significado do Gesto (Avatar);
- Gesto (Avatar).

Texto – Significado do Gesto (Avatar)

O texto escrito pelo utilizador como significado de uma imagem não poderá conter espaços para que possamos conter para um gesto o seu significado de forma a evitar que para palavras diferentes termos o mesmo gesto erradamente.

Gesto (Avatar)

Será possível ao utilizador fazer o carregamento do Avatar correspondente a palavra inserida. Podendo depois através de uma lista de palavras inseridas confirmar esses registos abrindo em detalhe ou mesmo em lista.



Figura 55 - Exemplo de Avatar – Letra A (Alfabeto LGP)

5 Conclusão

O objetivo inicial do projeto era o de criar e implementar um Modelo de Tradutor Bidirecional de Língua Gestual Portuguesa utilizando o dispositivo Microsoft Kinect Xbox 360TM e a Luva de Dados 5DT.

Depois de ter sido aprovado o respetivo trabalho no campo da visão por computador, que tem como objetivo o de modelar e implementar um sistema capaz de reconhecer qualquer gesto da Língua Gestual Portuguesa o ponto principal seria decidir como utilizar os dados que nos fornece o dispositivo do Kinect juntamente com a luva de dados.

A partida, na análise que fizemos do estado de arte, ficou claro que em geral todas as abordagens utilizam descritores complexos ou técnicas de aprendizagem complexos tais como: Modelos Escondidos de Markov – Hidden Markov Model (HMM); Redes Neurais Artificiais – Artificial Neural Network (ANN).

Alguns pesquisadores, neste momento, já utilizam algoritmos tais como DTW. No nosso caso, o grupo de pesquisa havia demonstrado o interesse pela utilização do SVM.

Começamos então por iniciar o projeto dividindo entre o modelo entre os gestos estáticos e os gestos em movimento. Dessa forma definindo que para o modelo de gravação e reconhecimento de gestos estáticos precisaríamos de apenas uma *frame* gravado pelo Kinect, ou seja os dados das juntas do esqueleto providenciado pelo Kinect mais os valores da luva de dados utilizada.

Para os gestos em movimentos o modelo teria que ser muito mais complexo visto que a definição de um gesto pode levar algum tempo dependendo do praticante que a estiver o executando.

Dessa forma decidimos utilizar para os gestos estáticos um modelo e outro diferente para os gestos em movimento. A forma de gravação de gestos estáticos difere da gravação de gestos dinâmicos no número de frames por gesto. Enquanto no modelo de gestos estáticos basta-nos uma *frame*, no modelo de gestos dinâmicos definimos que seriam necessários vinte vezes os dados que retiramos dos dados estáticos.

Portanto se numa *frame* retiramos vinte e três características – valores dos catorze sensores da luva de dados mais os valores X Y e Z das juntas de interesse definidas - em vinte frames, retiramos essas características, juntando a característica do tempo a ser efetuado o gesto dessa forma fazemos com que a descrição do nosso mesmo em movimento tenha um conjunto de quatrocentos sessenta e uma características.

Dessa forma conseguimos as classificações dos gestos estáticos e dinâmicos. Obtivemos resultados encorajadores que a nível de gestos estáticos conseguimos que a API reconhecesse todas as letras do alfabeto da Língua Gestual Portuguesa (LGP) e a nível de gestos dinâmicos em 15 palavras conseguimos resultados muito positivos com percentagens de acerto de

valores elevados. Nos gestos dinâmicos dividimos os conjuntos de palavras entre aqueles que consideramos mais difíceis de efetuar e aqueles mais fáceis mas que podiam dificultar o reconhecimento por parte da aplicação. Palavras como “Olá” e “Adeus” são das palavras que sendo bastante parecidas podia levar a confusão por parte do reconhecimento da aplicação. Mas da forma como foi construído o modelo, conseguimos obter bons resultados.

Com a palavra “Olá”, com 98, 98% de precisão, obtivemos 97 % de identificação da descrição do gesto como sendo a palavra “Olá”, para cem gestos efetuados, sendo que os restantes 3%, o modelo confunde com a palavra “Adeus”.

Já no reconhecimento da palavra “Adeus”, o modelo com uma precisão de 97,06 % acerta em 99% do total das descrições da palavra, confundindo apenas uma vez como sendo a palavra “Olá”.

Portanto pela definição de palavras que são semelhantes o modelo nos apresenta resultados considerados bons.

Para palavras, que não sendo semelhantes, tem um nível de dificuldade maior do que as duas palavras “Olá “ e “Adeus”, conseguimos resultados ainda melhores. Para palavras como “Aluno”, “Desporto”, “Floresta”, etc., conseguimos resultados a partir dos 99 % a nível de reconhecimento dos gestos com o nível de precisão entre 99% e 100%.

Com os resultados obtidos e com um trabalho contínuo, se o sistema estiver incorporado por exemplo em escolas, nos muitos locais de trabalho ou noutros lugares quaisquer, a comunicação entre os utilizadores surdos e os utilizadores comuns poderia ser mais simples e mais fácil.

Apesar do objetivo da tese ter sido de modelização de um tradutor de Língua Gestual Portuguesa Bidirecional, implementamos esse modelo, de forma a perceber as dificuldades que se carrega quando trabalhamos com os dispositivos que a partida tínhamos disponíveis.

No decorrer da tese, procedemos a uma apresentação deste projeto no âmbito de uma UC de Mestrado e também foram publicados dois artigos científicos sobre o projeto.

5.1 Trabalho Futuro

Embora a abordagem que tivemos satisfaz o objetivo do projeto, o futuro na modelização e implementação de tradutores de Língua Gestuais é claramente através de NUI utilizando sensores de profundidade como o Kinect. A partir do momento que for simplificada a deteção de outras características tais como a posição dos dedos a forma da mão, a expressão facial, se tornara mais simples o reconhecimento dos gestos cada vez mais complexos seja em movimento ou estático.

Visto que o tempo para fazer esta tese foi limitado, não permitiu que pudesse desenvolver um modelo que funcionasse com estas características mas considerando todas estas características como por exemplo a característica extremamente importante que é a expressão facial teríamos um modelo completo de tradução bidirecional.

O desenvolvimento de aplicações utilizando sensores de profundidade com mais resolução e precisão tal como o Kinect 2 (disponível ao público em geral a partir de Janeiro de 2014), terão um avanço significativo sendo mais fácil e mais rápido de em termos de Língua Gestual distinguir gestos muito parecidos. Para um sensor mais fraco em termos de resolução e precisão, não tendo a capacidade de extrair a diferença, não será possível detetar o gesto. Dito isto, a maior limitação, está na capacidade dos sensores. Portanto, quanto melhor e mais detalhe tivermos nos dados recebidos mais sucesso teremos no reconhecimento dos gestos.

A implementação utilizando dois dispositivos Kinect, um a frente e outro lateral já é possível através de controladores de dispositivo, "Kinect for Windows SDK 1.8", permitindo desta forma obter mais dados dos movimentos da língua gestual.

Da mesma forma, integrando outros tipos de sensores como o Leap Motion, que deteta movimentos de mãos na ordem de 1/100 mm permite criar um conjunto de sensores com informação bastante mais precisa de forma a ser possível detetar a diferença em movimentos ou gestos muito similares entre eles.

6 Bibliografia

[**3Gear**] 3Gear [Online] // 3Gear.

[**5DT**] 5DT [Online] // Series, 5DT Data Glove Ultra. - Janeiro de 2011. - <http://www.5dt.com/downloads/dataglove/ultra/5DT%20Data%20Glove%20Ultra%20Manual%20v1.3.pdf>.

[**ApSurdos**] Ap Surdos [Online] // Associação Portuguesa de Surdos. - http://www.apsurdos.org.pt/index.php?option=com_content&view=article&id=41&Itemid=8.

[**ArGis**] How Radial Basis Functions (RBF) work [Online] // ArGis Desktop Help. - Agosto de 2007. - [http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=How_Radial_Basis_Functions_\(RBF\)_work](http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=How_Radial_Basis_Functions_(RBF)_work).

[**Baltazar Ana Bela**] Funcionalidade de Língua Gestual Portuguesa [Online]. - Outubro de 2012.

[**Beckon Omek**] Omek Beckon [Online] // Omek Interactive. - <http://www.omekinteractive.com/products/beckon-3-for-pc/>.

[**Comunidade ApSurdos**] [Online] // Associação Portuguesa de Surdos. - http://www.apsurdos.org.pt/index.php?option=com_content&view=article&id=43&Itemid=57.

[**Correia Isabel Sofia, 2009**] PorSinal [Online] // PorSinal. - 2009 - <http://www.porsinal.pt/index.php?ps=artigos&idt=artc&cat=9&idart=234>.

[**CyberGlove**] Cyber Glove [Online] // cyberglovesystems. - <http://cyberglovesystems.com/?q=products/cyberglove-ii/overview>.

[**DTREG**] SVM [Online] // DTREG. - <http://www.dtreg.com/svm.htm>.

[**Frاند Jason, 1996**] Data Mining [Online] // Data Mining. - 1996. - <http://www.anderson.ucla.edu/faculty/jason.frاند/teacher/technologies/palace/datamining.htm>.

[**Historia dos Surdos**] Historia dos Surdos [Online].

[**Karatzoglou Alexandros, Meyer David e Hornik Kurt , 2006**] Support Vector Machines in R [Online] // jstatsoft.org. - Abril de 2006. - <http://www.jstatsoft.org/v15/i09/paper>.

[**MacCormick John, 2010**] How does the Kinect work? [Online]. - 2010. - <http://users.dickinson.edu/~jmac/selected-talks/kinect.pdf>.

[Microsoft's Digits] Microsoft's Digits hand-gesture sensor bracelet detailed [Online] // BBC. - <http://www.bbc.co.uk/news/technology-19884218>.

[PorSinal] PorSinal [Online] // PorSinal. - <http://www.porsinal.pt/index.php?ps=historia>.

[SigmaNil-Framework] Sigmanil-framework [Online] // OpenNi. - <http://www.openni.org/files/sigmanil-framework/>.

[SoftKinetic] SoftKinetic [Online] // Soft Kinetic. - <http://www.softkinetic.com/Portals/0/Download/iisu%E2%84%A2%20Datasheet%20V0.91.pdf>.

[Microsoft's Kinect SDK 1.8] <http://blogs.msdn.com/b/kinectforwindows/archive/2013/09/16/updated-sdk-with-html5-kinect-fusion-improvements-and-more.aspx>

7 Anexos

Artigo 1 - Virtual Sign Translator

Paula Escudeiro¹, Nuno Escudeiro², Rosa Reis³, Maciel Barbosa⁴ José Bidarra⁵, Ana Bela Baltazar⁶, Bruno Gouveia⁷ ^{1,2,3,6,7}

Departamento de Engenharia Informática,
Instituto Superior de Engenharia do Porto,
Porto, Portugal

Departamento de Engenharia Electrotécnica,
Faculdade de Engenharia da Universidade do Porto,
Porto, Portugal

Departamento de Ciências Exactas e Tecnologia,
Universidade Aberta,
Lisboa, Portugal

{pmo & nfe & rmr}@isep.ipp.pt; fmb@fe.up.pt; bidarra@univ.ab.pt; altazaranabela@sapo.pt; brunogouveiarodrigues@gmail.com

Abstract - In this paper we present the overall study that includes the model developed (VS Model) and the experiences performed, with an automatic bidirectional sign language translator, between written and sign language, which is being supervised by the research group GILT (Graphics, interaction & learning technologies) under the frame of a national project called Virtual Sign (VS project). This project aims to develop and evaluate a model that facilitates access for the deaf and hearing impaired to digital content - in particular the educational content and learning objects - creating the conditions for greater social inclusion of deaf and hearing impaired people. Access to digital content will be supported by an automatic translator between Portuguese Writing (LEP) and Portuguese Sign Language (LGP) supported by an interaction model.

Index Term - Access for deaf to digital content, Bidirectional translator, Learning objects, Interaction model, Sign Language

1. Introduction Disabled people face non-ending difficulties when they want to deal with the new technologies: the use of a computer, the access to Information, editing and printing a text, etc. Reading a document can be an extremely complex task, despite their simplicity for normal user [1]. Today, the extraordinary progress of the new technologies, bound to the data processing and the Internet, offers remarkable opportunities to bring a better quality of life to those who endure handicap and disabilities [2]. The target community addressed by the project has its own language, known as Sign language [3]. A sign language is a language that uses manual communication instead of sound to convey meaning - simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expressions to fluidly express a speaker's thoughts. The sign language remains nevertheless a fully-fledged language, with its own constructional method of the sentences. The Portuguese sign language (LGP) faces a low diffusion level among the deaf as well as the hearing communities. To be effective, the communication of a deaf person requires the knowledge of LGP, not only on the speaker's side, but also on the listener's. It is also essential that both speak the same sign language. Due to the shortage of information or lack of availability of the sign language, it is very important to improve it, allowing this group to have access to information in their first language, i.e., the sign language.

2. Principles of the Sign Language

Languages can be oral-auditives when written representations are not used but, instead, all communication is oral. As it happens with the Portuguese Language and all the oral languages. They can also be visual-spatial, with a realization almost natural of signs and visual reception. In this last case we could refer several sign languages used by many different deaf communities spread around the world.

These, such as oral languages, have their own grammar that allows distinguishing between sign languages and oral languages. The interest in the Portuguese Sign Language has been showing a remarkable growth, not only by the deaf community that today represents near 150.000 persons in Portugal but also by all the involving community, like family, educators, professors, etc. However, in Portugal there is not much done yet in this field to assist the deaf in their daily lives. The sign language is performed in a three-dimensional space, therefore the support from the new technologies is indispensable because it is possible to reach to the total perception of the sign, including the movement, the hands location and orientation, the configuration and also facial and body expressions. The richness of this tridimensional language is not limited itself on the simple technical realization of the sign. It also involves a total dynamic of the communication that is natural to the human being. The hands configuration, location and orientation are as important as the facial and body expressions that follow the realization of the signs. These two aspects could be crucial to distinguish very similar signs. Similarly to the oral language, sign language has a lexical, a “phonetic” (instead of articulated sounds it has articulated signs), a “phonology” (instead of phonemes, it has elements from different natures that accomplish the same differential function from the words visual form), a syntax, a semantic and a pragmatic of its own. Being characteristic from each country and culture, and not universal, allows describing all the reality that involves us, what we see, feel or think.

Considering all the basic principles of the Sign Language expressed above and the deaf community needs regarding the access to written sign language interpretation, we realize all the interest in the development of the Virtual Sign project. Its outcomes will assist the deaf community in accessing the written information through Portuguese Sign Language and vice versa thus, contributing to reduce info-exclusion of disabled persons.

3. Automatic Sign Language recognition

Sign Language communication uses two-handed gestures and non-manual signals. One of the biggest difficulties in recognizing sign language is that there are signs which involve global body movement while others involve only the configuration and orientation of the fingers of one hand. The sources of information that need to be acquired to decode sign language depend on the sign itself and are not always the same. The level of detail is also heterogeneous ranging from wide arm movements to confined finger configurations. This imposes conflicting requirements on the field of view; it must be large enough to capture the global motion, but at the same time, small local movements must not be lost. Moreover, both hands often touch or occlude each other when observed from a single viewpoint and, in some signs, the hands partially occlude the face. Occlusion handling is also an important consideration. Sign Language recognition includes tracking of the hands, face and body parts, feature extraction, modeling and recognition of time-varying signals, multimodal integration of information, etc. The use of various sensors for the wide range of features seems therefore inevitable.

A. The use of depth sensors

The development of a motion-sensing input device by Microsoft, the Kinect depth sensor, was a revolution in technology akin to those that shaped the most fundamental breakthroughs of the 20th Century. While this development may seem wide-ranging and diverse, it can be summarized simply: for the first time, computers can see [2].

The depth image outputted from a Kinect sensor is so important because it is much easier for a computer to interpret than a conventional color image. When processing a flat 2D image, pixels with similar colors that are near to each other might erroneously be assumed to belong to the same object. If you have 3D information then pixels that correspond to locations physically near to each other tend to belong to the same object, irrespective of their color. It has often been said that pattern recognition has been made artificially difficult because most systems rely on 2D data[3][4]. The use of depth sensors has been proven to be a more robust tracking method than vision based methods. Another big advantage of the Microsoft Kinect SDK is that it returns the position and orientations of the hand easily with a good accuracy.

B. Depth-based hand localization and tracking

To acquire hand gesture data we are using direct-measures outputted from a data glove with 14 sensors, (5DT Data Glove 5 Ultra). [finger abduction sensor have proven to be not accurate enough for joint angle measurement]]] The 5DT Data Glove SDK already provides basic gesture recognition of 15 hand gestures. [5].

C. Future of gesture recognition

The future of data acquisition of hand gestures is pointing to NUI (Natural User Interfaces) namely using depth sensors such as Kinect. The recently introduced Leap Motion Sensor should be an option for further development. For the first part of the project, we focused on the letters of the alphabet, in which the amplitude of hand motion is very small. Alphabet recognition depends mainly on fingers configuration and orientation of the hand in a static position. For vision-based approaches, the hand is generally restricted to palm facing the camera, against a uniform background.

D. Building the avatar hand coding the gestures

We are aiming to develop a bi-directional translator for Portuguese Sign Language. This requires, besides the recognition of signs based on body motion and their translation to written Portuguese, a way to simulate body motion representing written Portuguese. This direction, from written Portuguese to Portuguese Sign Language, will be assured by an avatar that will simulate a user using sign language. This avatar may be built on the common approach that hand-codes the categories of hand shape, hand orientation, hand location and movement type that make up each sign in the vocabulary, forming a lexicon of sign definitions. Classifying the sign label from component-level results is then performed by comparing the ideal lexicon categories with the corresponding recognized components.

E. Analysis of Non-manual Signals

In recent software like FaceReader [14], facial recognition is limited to the six basic emotions [14]: happiness, sadness, surprise, fear, anger, disgust and the neutral expression which is considered to be constrained. The Faceshift software [14] uses a depth sensor and is able to map with great precision a human face which allows for accurate feature extraction and the immediate animation of an avatar. Other body movements and postures involved in non-manual signals generally consist of torso motion, such as, body leaning forwards/backwards or turning to the sides. This features can also be easily recognized with a depth sensor like Kinect.

New devices for gesture recognition are showing up in the market, namely Omek Gasp. Omek Gasp is a software able to track the hand gesture which includes responsive and accurate tracking of both hands and fingers, even with occlusions. A full 3D model of the hands and, most importantly, a Gesture Authoring Tool (GAT) allow us to quickly and easily generate custom gestures [8]. Another important development is the Leap Motion sensor. Leap Motion is a depth sensor specialized in tracking every feature of the hands [9].

4. VS Model

The VS project addresses the lack of adaptability of most educational organizations to deaf and hard of hearing people. This lack of adaptability generates adverse conditions to deaf and hard of hearing people, preventing them to have the same opportunities in education as the other citizens. The main results of this project are:

- a model, that allows the deaf and hard of hearing people to improve their integration into mainstream education [5], as shown in Fig. 1.
- an animated virtual character, to be integrated in educational software, that translates text to sign language.
- a virtual reality environment to translate Portuguese sign language to Portuguese text.

As an engineering school, we intend to take the first steps for the integration of these individuals by developing a model to assist the translation of the educational content of the different course units of the undergraduate computer science degrees to the Portuguese Sign language.

This model may be applied in different fields, such as virtual museums, web pages, services, so on [4].

Fig. 1 Phases in the Creation of a Gesture.

5. Evaluation

To evaluate the performance of gesture recognition based on the input provided by both the Data Glove and Kinect we have collected data from seven different users. Each user, wearing the Data Glove and being tracked by Kinect skeletal tracking performed the Portuguese sign language symbols representing the alphabet and the ordinals from 1 to 9 in a total of 35 symbols. Each symbol was repeated 10 times. In total we have captured 2450 snapshots (seven users times 35 symbols times 10 repetitions). Since these symbols are represented only by the right hand and the right hand configurations we have extracted the 26 relevant features from all the sensors available through the Data Glove and Kinect. These features correspond to the 14 glove sensors plus 12 spatial coordinates provided by Kinect with regards to the right shoulder, right elbow, right wrist and right hand (x, y and z coordinates for each). This dataset was then used to evaluate the accuracy of an SVM classifier [13] using radial basis function (RBF) kernels. Error rates were estimated using 10-fold cross validation.

The performance of the classifier was evaluated on two distinct feature sets. We have started to evaluate the accuracy of the classifier based only on the input from the Data Glove. The estimated error in such a setting is 0.02. Then we have used all the data, including the input from both the Data Glove and the Kinect skeletal tracking (right arm). The error rate dropped to 0.01. The difference of the mean error observed in these two situations is statistically significant. The t-test for equal means that was performed yielded a p-value of 0.01344.

6. Conclusions

The selection of this target population arises due to the growing number of students with special needs who complete the elementary and high school and come to higher education. This situation demands for new means that allow these individuals to have easy access to educational content. Higher education institutions must host and provide appropriate conditions for students to get their degrees despite their disabilities. These institutions play an important role in the inclusive education, i.e., in the implementation of the educational system that includes everyone truly, (Brazil / MEC / SEESP, 1998). As university teachers, we want to develop a methodology that allows the translation of textual educational content to sign language, and the other way round. This serves as a launching pad for the development of applications targeted for individuals with special needs. The first developments, focused on the use of a data glove and Kinect to recognize the Portuguese alphabet in a static setting using automatic classification are promising. We are now moving to the next step which is to recognize the signs in a dynamic setting.

7. Acknowledgment

The author would like to thank the Engineering Institute of Oporto and GILT (Graphics, Interaction and Learning Technologies) and FCT (Fundação para a ciência e Tecnologia) for making possible this work

8. References

- [1] Reyes, I. 2005 "Comunicar a través del silencio :las posibilidades de la lengua de signos
- [2] Herrero-Blanco, A., Salazar-Garcia, V. 2005 "Non-verbal predicability and copula support rule in Spanish Sign Language". In: de Groot, Casper / Hengeveld, Kees (eds): Morphosyntactic.
- [3] Annelies Braffort, 2008 " TRAITEMENT AUTOMATIQUE DE LA LANGUE DES SIGNES.
- [4] VSigns v0.1a running on dbX v1.00 (c) 2004, Informatics & Telematics Institute HANIM models (c) Christian Babski (LIG/EPFL).
- [5] SignWriting: um sistema de escrita para linguas de sinais. EFETA – Publicação Pastoral do Surdo, ano 5, n 32. Set/out. 1998
- [6] Amaral, M. A., Coutinho, A., & Martins, M. R. (1994). Para Uma Gramática da Língua Gestual Portuguesa. Lisboa: Caminho.
- [7] Baltazar, A.B. (2010). Dicionário da Língua Gestual Portuguesa. Porto: Porto Editora.
- [8] Sylvie C.W. Ong and Surendra Ranganath, "Automatic Sign Language Analysis: A Survey and the Future beyond Lexical Meaning", IEEE Transaction Pattern Analysis and Machine Intelligence, 27(6): 873-891, June 2005.
- [9] Primesense, PrimeSense™ NITE Algorithms, Release 1.5. PrimeSense, 2011
- [10] S. Kean; J. Hall; P. Kerry, Meet the Kinect: An Introduction to Programming Natural User Interfaces, Technology In Action, 2011.
- [11] D. Catue, Programming with the Kinect for Windows: Software Development Kit, Microsoft, 2012
- [12] 5DT, 5DT Data Glove Ultra Series User's Manual, January 2011
- [13] Campbell, C., Ying, Y., Learning with Support Vector Machines, Morgan & Claypool Publishers, 2011
- [14] Noldus Information Technology bv Wageningen, The Netherlands