



Faculdade de Ciências e Tecnologia da Universidade de Coimbra
Departamento de Engenharia Electotécnica e de Computadores
Mestrado Integrado em Engenharia Electrotécnica e de Computadores

Estimação da pose da cabeça e da direção do olhar de um utilizador

Miguel de Abreu Ferreira da Silva Brito

Júri:

Professor Doutor Helder de Jesus Araújo

Professor Doutor Paulo Jorge Carvalho Menezes

Professor Doutor Jorge Nuno de Almeida e Sousa Almada Lobo

Julho 2012

“Sou um técnico, mas tenho técnica só dentro da técnica.

Fora disso sou doido, com todo o direito a sê-lo.”

- Álvaro de Campos

Agradecimentos

À minha mãe, que me atura há mais tempo que o que eu me consigo recordar.

Ao meu pai, que me convenceu de que no futuro os carros serão cada vez mais dos engenheiros eletrotécnicos e cada vez menos dos mecânicos.

Às minha irmãs, por serem umas chatas e umas queridas quando é preciso.

Aos meus colegas de laboratório, Adriano, Braga, Bruno e Gustavo, pelo apoio e pelos momentos de descontração.

Ao Luís Almeida, pela ajuda e pelo apoio.

Ao Professor Paulo Menezes, pelo apoio, sapiência, e beleza gasta com infindáveis perguntas.

À minha Guidinha, pela amizade, amor, e companhia!

E a restantes amigos e familiares que tanto me apoiaram, o meu sincero obrigado!

Índice

Lista de Figuras	ix
Lista de tabelas	xi
Abreviaturas	xiii
Glossário.....	xv
Resumo.....	1
Abstract	3
1 - Introdução	5
1.1 - Estado de arte	9
1.2 – Motivação, objetivos e contribuição deste trabalho	15
1.3 – Análise do problema.....	16
1.4 - Estrutura da tese	16
2 – Formulação e solução do problema.....	17
2.1 - Dados biométricos	17
2.2 – Seguimento dos olhos.....	19
2.3 - Estimação da pose da cabeça	27
2.4 – Estimação da direção do olhar	35
2.5 – Estimação da direção do olhar com o capacete X-View HED.....	39
2.6 - Discussão	42
3 - Conclusão.....	45
3.1 - Trabalho Futuro	47
Referências	49

Lista de Figuras

Figura 1 - Lente de contato para seguimento do olhar.	9
Figura 2 - Eletrodos para seguimento do olhar.	9
Figura 3 – Seguidor de olhar remoto com "head mount"	10
Figura 4 – Seguidor de olhar móvel.	11
Figura 5 - Par de óculos para seguimento do olhar.	12
Figura 6 – Seguidor de olhar remoto portátil.	12
Figura 7 - Par de olhos, gaze.	18
Figura 8 – O olho humano em corte.	18
Figura 9 - Imagem capturada pela webcam.	19
Figura 10 - Cara encontrada da imagem.	20
Figura 11 - Janela de pesquisa para o olho direito.	21
Figura 12 - Exemplo da aplicação da transformada de Hough.	22
Figura 13 - Transformada de Hough aplicada a um olho.	22
Figura 14 - Olho e correspondente histograma.	23
Figura 15 - Olho, usando acumuladores por linhas e por colunas.	24
Figura 16 - Processamento da imagem do olho.	25
Figura 17 - Obtenção do centro da íris.	25
Figura 18 - Fluxograma do módulo de Seguimento dos olhos.	26
Figura 19 – Resultado final do Seguimento dos olhos.	26
Figura 20 - Ilustração do modelo de câmara <i>pinhole</i>	28
Figura 21 – Exemplo de um elipsoide.	30
Figura 22 - Obtenção de pontos SURF da cara.	31
Figura 23 - Projeção de pontos da cara no modelo da cabeça.	32
Figura 24 - Fluxograma do estabelecimento de referência para a estimação da pose da cabeça.	33
Figura 25 - Emparelhamento de pontos SURF.	33
Figura 26 - Verificação da projeção de pontos.	35
Figura 27 - Projeção das pupilas no modelo da cabeça.	36
Figura 28 - Correção da posição da pupila.	37
Figura 29 - Estimação do gaze.	38
Figura 30 - Fluxograma de Estimação da direção do olhar.	38

Figura 31 – Seguimento do olhar com o capacete.	39
Figura 32 - Imagem do olho, captada com recurso ao capacete.	40
Figura 33 - Marcação do centro da pupila.	41
Figura 34 - Exemplo de aplicação da estimação do gaze.	43
Figura 35 - Teste de sensibilidade.....	44
Figura 36 - Exemplo de um programa de comunicação assistida.....	45
Figura 37 – Hilário.....	48

Lista de tabelas

Tabela 1 - Comparação de alguns sistemas de eye tracking	14
--	----

Abreviaturas

APCC – Associação de Paralisia Cerebral de Coimbra

DEEC – Departamento de Engenharia Electrotécnica e de Computadores

ISR – Instituto de Sistemas e Robótica

Glossário

Frame – imagem capturada pela *webcam*.

OpenAR – Biblioteca de C/C++ para a realidade aumentada desenvolvida no ISR/DEEC

OpenCV – Biblioteca de C/C++ para a manipulação de imagens.

Pose – refere-se à posição de um objecto no espaço relativamente à origem do sistema de eixos de uma referência. A pose decompõe-se como sendo uma rotação e uma translação.

Sacada – movimento rápido dos olhos, normalmente associado a pesquisa espacial.

SURF – Speeded Up Robust Features.

Pontos SURF – é o nome dado para indicar um conjunto de pontos obtidos com o recurso ao algoritmo SURF (Bay, Tuytelaars, & Van Goo, 2006), dos quais assim se conhece as suas coordenadas na imagem e a sua assinatura.

Utilizador – sujeito que usa o programa.

Webcam - câmara de captura de vídeo.

Resumo

Neste trabalho pretende-se estudar a utilização da webcam de um computador portátil como dispositivo para a estimação da direção do olhar de um utilizador.

É apresentado o estado de arte de sistemas de estimação da direção do olhar e explica-se o porquê do uso da webcam como dispositivo escolhido.

Apresenta-se e analisa-se o problema de usar a webcam para a estimação da direção do olhar de um utilizador. São assim apresentados alguns dos pressupostos teóricos e dados biométricos tidos em conta na execução deste trabalho.

São apresentados os algoritmos necessários para obter a direção do olhar; um algoritmo para a detecção de íris, um para a estimação da pose da cabeça e um que combina os dois anteriores para a estimação da direção de olhar de um utilizador. É feita uma explicação de algumas soluções tomadas aquando da implementação dos algoritmos.

Conclui-se mostrando o resultado da aplicação dos algoritmos desenvolvidos em duas situações diferentes, o do uso da webcam, e o do uso de hardware específico para aplicações de seguimento de olhar.

Mostra-se ainda a contribuição deste trabalho noutros, enquanto trabalho futuro e mesmo em artigos de conferência entregues e aceites para apresentação em conferência.

Abstract

This work aims to study the usage of a laptop's webcam as a device for estimating the gaze direction of a user.

The state of the art of systems for the estimation of gaze direction is presented and clarifies the use of the webcam as device chosen.

It is presented the problem and analysis of using a webcam to estimate the gaze of a user. Are thus presented some of the theoretical and biometrics taken into account in carrying out this work.

The algorithms required for the direction of gaze are presented as: an algorithm for the detection of iris, one for estimating the head pose and one that combines the previous two to estimate the direction of a user's gaze. It is made an explanation of some of the solutions taken when implementing the algorithms.

It is shown the result of applying the algorithms developed in two different situations, the use of webcam, and the use of specific hardware for eye tracking.

The contribution of this work is showed as future work and even in conference articles delivered and accepted for presentation at conference.

1 - Introdução

A rápida disseminação dos computadores pessoais portáteis tem sido notória nos últimos tempos; a escolha destes comparativamente aos computadores pessoais fixos deve-se a vários factores, desde a óbvia portabilidade, ao baixo consumo, ou até mesmo à quantidade de *hardware* que trazem já incorporado. Podemos assim extrapolar que o comum utilizador de um computador, utiliza um computador portátil com webcam incorporada, e marcamos este como o nosso paradigma.

Pode-se afirmar que quando um utilizador está sentado ao computador tem sempre a “observá-lo” a *webcam*, e podemos verificar facilmente que em quase todas as situações na imagem captada pela webcam são claramente observáveis os olhos do utilizador. Ora, a partir dos olhos, pode-se estimar o olhar do utilizador; vamos assim tentar criar uma aplicação que usando apenas uma webcam sem nenhuma alteração, seja capaz de estimar a direcção do olhar de um utilizador.

Conhecer o comportamento do olhar de um utilizador ao computador é uma mais valia em muitas áreas, de que são exemplo a saúde, o marketing, a computação gráfica, a interação homem-máquina, entre outras.

Na saúde

Há atualmente vários estudos que comprovam uma correlação de padrões de olhar com condições físicas desde o cansaço à doença (Quinn, 1995). Reconhecer certos padrões de movimento do olhar poderia ser utilizado para “obrigar” o utilizador do computador a descansar por um certo período de tempo, ou, caso fosse determinado, lançar um aviso médico de uma doença. Por exemplo, atualmente há já vários jogos de computador que, após um certo período de tempo, obrigam o jogador a fazer um intervalo para descansar a vista; não têm, no entanto, nenhum dado fisiológico, apenas o tempo que já decorreu desde o início da utilização do jogo; com um sistema que reconhecesse um padrão de olhar cansado, estes intervalos seriam lançados de forma dinâmica ao invés da atual forma calendarizada.

No marketing

Sistemas capazes de reconhecer a direcção do olhar de um utilizador são atualmente usados para testar em laboratório, desde a disposição de ícones num programa de computador, à disposição da publicidade num *website*. Dotar um computador pessoal desta capacidade de análise traria uma população de teste muito maior, ou até mesmo personalizar a forma como recebemos informações e publicidade no nosso computador de forma a potenciar o seu efeito.

Na computação gráfica

A qualidade dos mundos virtuais produzidos é cada vez melhor e, de acordo com alguns, até demasiado perfeita. O sistema visual humano é limitado na região que do mundo que foca, sendo que o restante é-nos percebido desfocado. Imagine-se um cenário virtual a três dimensões, mostrado num ecrã com capacidade 3D e o utilizador com óculos específicos para visualizar o ecrã. Para quase todos os ecrãs atualmente disponíveis há uma distância e uma posição da cabeça do utilizador limitada para a correta visualização, em que fora dessa posição a percepção do 3D é deformada, porque a disparidade das imagens emitidas é fixa. Pode-se assim atuar em três frentes diferentes: sabendo a pose da cabeça do utilizador, a disparidade das imagens que formam o 3D; conhecendo a direção do olhar do utilizador pode-se desfocar o restante do mundo virtual, fora do foco de visualização; conhecendo a direção do olhar pode-se também reduzir os recursos usados para a produção da imagem do cenário que não é observado pelo utilizador, tornando assim o cenário percebido mais real. Um cenário virtual em três dimensões observado num ecrã 3D com óculos específicos fora da posição de correta visualização, por a percepção 3D ser deformada, ou por estar tudo nítido (ainda que a profundidades diferentes) e não se assemelhar à realidade, provoca nalgumas pessoas cefaleias (dores de cabeça) e até mesmo cinetose (enjoo causado pelo movimento) (Lambooi, IJsselsteijn, & Heynderick, 2007), pois o cérebro recebe dados contraditórios e tenta ajustar-se. Como tal, no sentido de evitar este desconforto, atuar na correção da percepção e na desfocagem do cenário não focado revela-se bastante importante.

Na interação homem-máquina

A interação de uma pessoa com um robô não é semelhante à interação entre duas pessoas. Como não estamos acostumados a esse tipo de interação tendemos a comportarmo-nos de forma diferente. No entanto há certas reações que se mantêm, como o gesticular e o apontar, e até mesmo, quando descrevemos um objecto de uma sala normalmente tentamos manter o contacto visual com esse objecto. Esse tipo de reações, aquando de uma conversação com uma pessoa, permitem simplificar a comunicação, ou perceber o nível de atenção da outra pessoa. Assim, transpondo para a interação homem-máquina, perceber o que é que um utilizador está a gesticular ou apontar, ou saber para onde está a olhar é bastante interessante. Neste trabalho vamos apenas focarmo-nos na parte do olhar do utilizador. Perceber a concentração e envolvimento de uma pessoa na interação com um robô, é atualmente uma área em investigação. Analisar a direção e padrão do olhar de uma pessoa enquanto interage com um robô, seria uma ferramenta interessante para quantificar a sua concentração e envolvimento na interação, trazendo dados ao robô que o possibilitassem à mudança da forma de funcionamento. Por exemplo, quando descrevemos um objecto de uma sala, tendencialmente olhamos para este. O

robô conhecendo a zona para onde o utilizador está a olhar permite-lhe perceber que de muitos objetos numa sala, é um que se encontra naquela zona. Já na área da utilização de um computador pessoal, um sistema capaz de estimar a direção do olhar torna-se bastante interessante. Um sistema destes pode ser utilizado como uma forma de interface, substituindo por exemplo o movimento do rato do computador. Estes sistemas podem até mesmo ser usados para analisar se a disposição de ícones de um sistema operativo está na forma mais correta, verificando por exemplo se um utilizador passa muito tempo à procura de um ícone.

1.1 - Estado de arte

Os primeiros estudos efetuados usando o seguimento dos olhos datam do principio do século XIX. A esta altura era efectuada uma análise direta da direção e padrões do olhar de um sujeito de teste. Com estas primeiras experiências surgem estudos que relacionam a forma como lê, e como se adquire informação visual com movimentos específicos do olhar. Em meados do século XX começaram a usar-se sistemas menos intrusivos, recorriam a câmaras e raios de luz, registavam em filme a os movimentos dos olhos de um utilizador para posterior análise. Os sistemas atuais são assim evoluções deste ultimo, usam câmaras para captar o movimento dos olhos, mas ainda menos intrusivos para o sujeito de teste. Há no entanto outras formas de fazer o seguimento dos olhos de um utilizador. Há sistemas em que se coloca uma lente de contacto que tem uma espira (Figura 1). A deformação de um campo magnético que se conhece é percebida como movimento do olho.

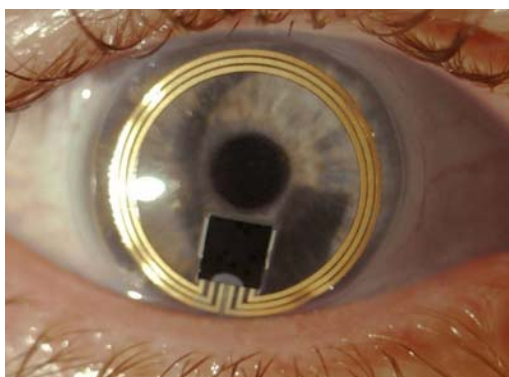


Figura 1 - Lente de contato para seguimento do olhar.

Outros sistemas recorrem ao uso de eletrodos colocados na zona do olho para capturar os movimentos do olho pelo diferencial de potencial elétrico.



Figura 2 - Eletrodos para seguimento do olhar.

Os sistemas que se tornaram mais comuns e com maior evolução são os que se baseiam na utilização de câmaras para a captação do movimento dos olhos. Estes vingaram por serem pouco ou nada intrusivos, de fácil instalação e utilização, entre outras características. Cedo passou-se da utilização da luz visível para identificar os olhos para a luz infravermelha. Esta é invisível ao ser humano tornando-se num método menos intrusivo para o seguimento dos olhos do utilizador. A luz infravermelha tem ainda vantagens acrescidas. Na imagem captada por uma câmara que está a filmar os olhos do utilizador, a luz infravermelha, aumenta bastante o contraste; a pupila é bastante definida e muito escura.

Neste trabalho vamos abordar em maior detalhe alguns destes sistemas que utilizam câmaras para o seguimento dos olhos e estimação da direção do olhar do utilizador.

Os sistemas de seguimento do olhar são vulgarmente conhecidos apenas por *eye trackers*, englobam-se em dois grandes grupos, os remotos e os móveis. Os remotos baseiam-se num sistema de câmaras fixas que observam o utilizador, os móveis têm câmaras presas à cabeça do utilizador que observam os olhos do próprio. Quer os remotos, quer os móveis podem ser do tipo binocular ou monocular, ou seja, para a estimação da direção do olhar utilizam os dois olhos ou apenas um.

Seguidamente, apresentam-se alguns dos sistemas mais comuns e as suas principais características.

Seguimento do olhar do tipo remoto

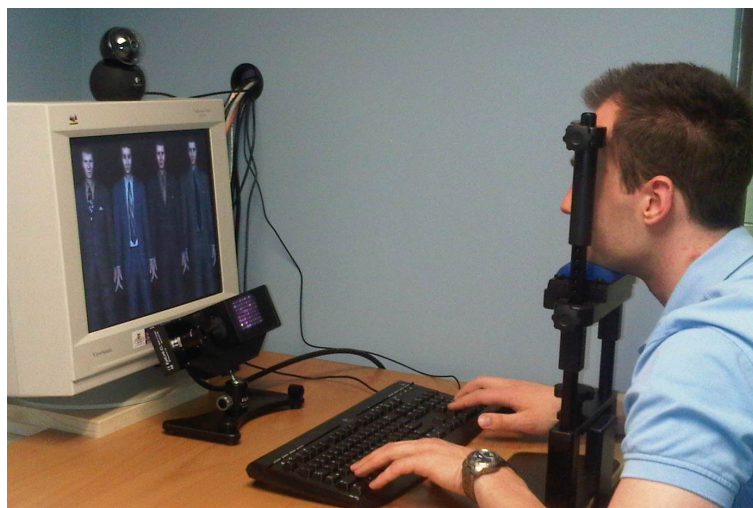


Figura 3 – Seguidor de olhar remoto com "head mount".

Na Figura 3 vemos um sistema de seguimento do olhar remoto, cuja principal característica é o conjunto de câmaras fixo (na figura, é a câmara em cima do tampo da mesa); este sistema usa LEDs de infravermelhos e uma câmara de alta resolução para captar os olhos do

utilizador. Como a maioria dos sistemas não compensam o movimento do utilizador em relação à câmara (do qual este é exemplo), este *setup* recorre a um *head mount* (uma armação que segura a cabeça do utilizador), garantindo uma boa eficiência e pontaria na estimação da direção do olhar.

Seguidor de olhar móvel



Figura 4 – Seguidor de olhar móvel.

Na Figura 4 vemos um seguidor de olhos do tipo móvel, um capacete de seguimento de olhar da SMI, o X-View HED. A principal característica deste é que permite completa liberdade de movimentos ao utilizador, uma vez que está agarrado à sua cabeça. Este sistema é composto por duas câmaras, uma que capta o olho do utilizador através de um semi-espelho, outra que capta o campo visual para o qual o utilizador olha. A estimação da direção do olhar é feita entre as duas imagens captadas, marcando-se um ponto na imagem do campo visual, para o qual o utilizador está a olhar.

Neste caso o sistema é do tipo monocular, que significa que apenas faz uso de um olho para a estimação da direção do olhar do utilizador.

A SMI, uma das empresas líder na produção de sistemas de seguimento dos olhos com estimação da direção do olhar apresenta evoluções dos sistemas anteriores.

O seguidor de olhos móvel evolui de um capacete monocular para um par de óculos binocular (Figura 5). Assim há logo duas grandes vantagens, uma direta para o utilizador, que é

o conforto de utilização, outra, a maior precisão da estimação da direção do olhar (por recorrer aos dois olhos).



Figura 5 - Par de óculos para seguimento do olhar.

Este sistema tem duas câmaras para captação dos olhos do utilizador (uma por olho) e uma câmara para captar o campo visual – no modelo antigo era uma *webcam* normal, neste é uma câmara de alta resolução e alto contraste. Para a captação dos olhos, este sistema, como todos os anteriores, utiliza LEDs infravermelhos para melhorar o contraste dos olhos.

O seguidor de olhos remoto (Figura 6), tem como grande inovação, o facto de ser portátil (pelas dimensões e peso e só precisar da conexão USB para alimentação). Este sistema, como se liga ao computador e não usa nenhuma câmara para captar o campo de visão do utilizador, apenas devolve a direção do olhar em relação ao monitor do computador. É um sistema binocular, com uma câmara por olho.



Figura 6 – Seguidor de olhar remoto portátil.

Também este sistema usa iluminação infravermelha para melhorar a imagem captada dos olhos do utilizador. Este sistema, baseando-se nos olhos, consegue fazer a compensação do movimento

do utilizador; deste modo, não é restritivo ao movimento, mas ainda assim, a compensação do movimento é limitada a uma caixa de operação (virtual) de 30x20x30cm, fora da qual não há estimação da direção do olhar.

De notar que para todos os sistemas de seguimento do olhar até agora apresentados, é necessária uma calibração inicial antes de cada utilização, de modo a fazer-se o mapeamento do movimento dos olhos para o movimento de um ponto na imagem captada no mundo real, ou no caso dos sistemas seguidores de olhos remotos, entre o movimento dos olhos e a imagem do monitor. Esta calibração vai desde um ponto, em sistemas binoculares, e entre cinco a quinze pontos, em sistemas monoculares. A calibração consiste em fazer o utilizador olhar para um determinado ponto pré definido, de forma a que se registre a posição da pupila; fica-se assim com um mapeamento entre o movimento da(s) pupila(s) com a estimação da direção do olhar do utilizador.

Estes sistemas tem outras funcionalidades para além do seguimento e estimação da direção do olhar: pela resolução da imagem obtida, são também capazes de devolver informações como fixações (baseando-se no tamanho das pupilas); com frequências acima dos 30 Hz conseguem devolver informações como o piscar de olhos, e com frequências acima dos 60 Hz devolvem já informações como as sacadas (movimentos dos olhos rápidos de pesquisa).

Nestes sistemas a medida de erro mais usada é o erro angular e frequentemente ronda 1 grau de erro, sendo que alguns apresentam valores mais altos, rondando os 2 graus. Note-se que 1 grau de erro, significa que a uma distância de 60cm ao alvo, o sítio real para onde o utilizador olha, está dentro de uma circunferência com sensivelmente 1cm de raio.

Apresentam-se em seguida na Tabela 1 alguns sistemas de seguimento do olhar, dados retirados dos respectivos *sites*, sendo que os preços variam entre os 50€ (muito baixo) e os 40 000€ (muito alto).

	Tipo	Portabilidade	Movimento do utilizador	Recursos	Erro	Preço
Remoto com head mount	Remoto	Limitada	Não	Iluminação infravermelhos; Câmara de alta resolução; Mecanismo para fixar cabeça;	0.2°	Alto
Capacete SMI	Móvel	Sim	Sim	Iluminação infravermelhos; Webcam modificada;	1.0°	Médio
Óculos SMI	Móvel	Sim	Sim	Iluminação infravermelhos; Câmaras de alta resolução e alto contraste;	0.5°	Muito Alto
SMI RED M	Remoto	Sim	Limitado	Iluminação infravermelhos; Câmaras de alta resolução e alto contraste;	0.5°	Médio
ITU gaze tracker	Remoto	Sim	Não	Iluminação infravermelhos; Webcam de visão noturna;	2°	Muito Baixo

Tabela 1 - Comparação de alguns sistemas de seguimento de olhar.

Analisando os sistemas de seguimento do olhar, conclui-se que: a resolução da câmara para captação da imagem dos olhos diminui com a distância do utilizador ao sistema de seguimento do olhar; todos os sistemas supracitados recorrem a iluminação infravermelha para evidenciar melhor a pupila do olho; com o aumento da frequência de captação de imagens é possível recolher mais informações acerca dos movimentos rápidos dos olhos, como as sacadas; o erro menor é conseguido em sistemas remotos em que se garante a imobilidade do utilizador; o preço dos equipamentos está relacionado diretamente com a quantidade e performance do hardware.

1.2 – Motivação, objectivos e contribuição deste trabalho

Atualmente grande parte dos computadores, quer portáteis quer de secretaria trazem já incorporados uma webcam, assim este trabalho tem como objectivo testar a utilização de uma webcam de um computador para a função de seguidor de olhar de um utilizador. A escolha da webcam ao invés de um outro sistema de seguimento dos olhos deve-se à disponibilidade da primeira. Conhecer a direcção do olhar de um utilizador é bastante interessante em muitas áreas como já vimos anteriormente.

Parece evidente que a partir do momento em que há uma *webcam* que nos observa, e consegue “ver” os olhos de um utilizador, só falta uma aplicação capaz de estimar a direcção do olhar utilizador para que tenhamos à disposição uma ferramenta poderosa.

Neste trabalho tenta-se, assim, desenvolver um seguidor de olhar usando apenas uma *webcam* normal sem qualquer alteração; o principal motivo desta escolha é oferecer uma ferramenta que qualquer pessoa (que tenha um computador pessoal) possa usufruir sem recorrer a mais nenhum equipamento, ou sem necessitar de alteração de equipamentos existentes.

Durante o decorrer deste trabalho surgiu a oportunidade de colaborar com outro projeto em desenvolvimento e que tinha como objectivo usar um sistema de seguimento do olhar para criar um mecanismo de comunicação para pessoas com certos graus de paralisia cerebral. O objectivo aqui foi de testar até que ponto seria possível substituir o software que acompanha um sistema comercial pelos algoritmos desenvolvidos para a webcam dos computadores portáteis. Aplicaram-se assim os algoritmos desenvolvidos neste trabalho, no capacete de seguimento de olhar SMI X-View HED.

Neste trabalho desenvolveram-se algoritmos para a detecção e seguimento do olhar de um utilizador, e algoritmos para a determinação da pose da cabeça de um utilizador.

Houve a colaboração deste trabalho com os algoritmos desenvolvidos, num artigo científico com o título “Context-based understanding of interaction intentions” para apresentação na conferência IEEE RO-MAN 2012. Neste artigo o olhar de um utilizador é usado para facilitar a interação e a comunicação com um robô. A estimação da direcção do olhar é usado para limitar ou indicar ao robô para que zona ou objecto um utilizador está a olhar, diminuindo a computação necessária para encontrar esse objecto ou zona.

1.3 – Análise do problema

Este trabalho pode dividir-se na solução de três problemas diferentes: estimar a direção do olhar de um utilizador quando comunica com um robô, estimar a direção do olhar quando um utilizador está sentado em frente a um computador portátil, e estimar a direção do olhar quando o utilizador utiliza um sistema de seguimento de olhar como o capacete da SMI.

Estimar a direção do olhar do utilizador quando na interação com um robô tem limitações específicas. Como a câmara que observa o utilizador continua a ser uma webcam, à distância normal de interação do utilizador com o robô (>60cm) não é possível estimar a direção do olhar, porque não há resolução de imagem suficiente para cumprir esse objectivo. A estimação do olhar é assim aproximada ou por um vector perpendicular à cara, ou apenas é feita se o robô se aproximar do utilizador.

Quando o utilizador se encontra em frente a um computador portátil a estimação do olhar é feita usando as imagens captadas pela webcam. Vamos apresentar algumas das limitações que foram tidas em conta para a solução final. As imagens da webcam são imagens de baixa resolução a uma cadência não superior a 24 imagens por segundo. Como o utilizador comum não tem equipamentos para iluminação infravermelha, que iria melhorar e facilitar a obtenção dos centros dos olhos, também aqui não se vão usar estes equipamentos.

Já quando o utilizador usa equipamentos como o capacete da SMI a estimação da direção do olhar é feita de forma diferente. Primeiro temos uma câmara que observa o olho do utilizador a uma distancia muito próxima, tendo grande qualidade de imagem e grande resolução da zona do olho. Segundo, há iluminação infravermelha, que melhora bastante o contraste dessa zona, facilitando a segmentação de imagem a ser feita para a obtenção do olhar do utilizador.

1.4 - Estrutura da tese

No segundo capítulo são apresentados alguns dados biométricos tidos em conta neste trabalho. Neste capítulo cada um dos módulos é apresentado enquanto problema isolado; são referidas algumas das soluções tomadas e são mostrados os resultados de cada módulo. Termina com uma discussão geral sobre os módulos. No terceiro capítulo apresentam-se conclusões, são sugeridas melhorias ao sistema, e apresenta-se trabalho futuro.

2 – Formulação e solução do problema

Neste capítulo apresentam-se alguns dados biométricos tidos em conta para a realização deste trabalho. O problema de estimar a direção do olhar do utilizador é partido em módulos. Um primeiro módulo de seguimento dos olhos, que devolve as coordenadas das íris do utilizador. Segundo módulo que devolve a estimação da pose da cabeça. Terceiro módulo utiliza os dois anteriores para estimar a direção do olhar do utilizador. É depois apresentada aplicação do módulo de seguimento dos olhos a um hardware específico para seguimento dos olhos. Este capítulo termina com uma discussão sobre os módulos e aplicações.

2.1 - Dados biométricos

Aqui apresentam-se alguns dados biométricos tidos em conta para a realização do trabalho; neste trabalho, que incluía desenvolver um sistema de seguimento do olhar, numa necessidade de coerência da linguagem, e até mesmo da simplificação da programação, conhecer alguns dados biométricos é bastante importante.

A direção do olhar de uma pessoa, pode ser determinada se conhecermos precisamente as dimensões e posições dos olhos, e as posições das pupilas de uma pessoa. Se projetarmos raios que passam pelos centros dos olhos pelas respectivas pupilas, estes intersectar-se-ão num ponto no espaço, que será o ponto para o qual a pessoa está a olhar, tal como o representado na Figura 7.

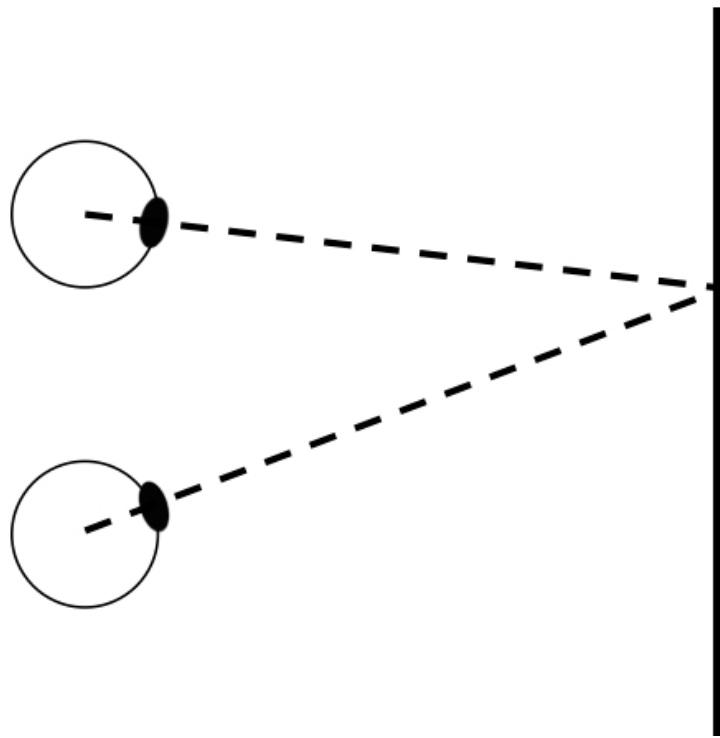


Figura 7 - Par de olhos, gaze.

Podemos simplificar a forma do olho humano (Figura 8) como sendo um corpo aproximadamente esférico, e de acordo com muitos autores, estipula-se que o diâmetro do olho humano é de aproximadamente 24.5mm (Press, Salem, 1998, p. 608).

Relativamente ao par de olhos, sabe-se que a distância entre centros dos globos oculares varia entre os 60mm e os 70mm e a média é de aproximadamente 65mm para 90% das pessoas (Dogson, 2004).

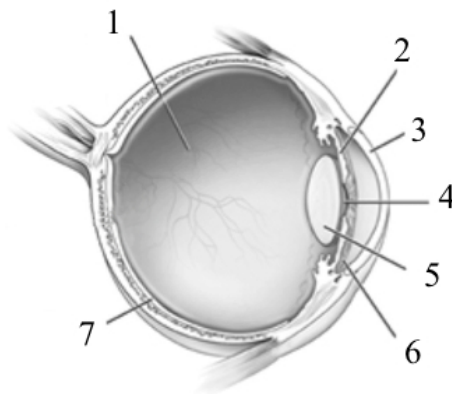


Figura 8 – O olho humano em corte.

Legenda da Figura 8:

- 1 – Humor vítreo
- 2 – Íris
- 3 – Córnea
- 4 – Pupila

5 – Cristalino

6 – Humor aquoso

7 – Retina

Tendo isto, conhecemos já a forma e as medidas do olho humano e, também, a distância média entre centros dos globos oculares, futuramente, e mesmo em situação do desenvolvimento da aplicação considere-se o olho como sendo perfeitamente esférico, com 24.5mm de diâmetro e a distância entre centros de globos oculares 65mm.

2.2 – Seguimento dos olhos

Este trabalho tem como objectivo final estimar a direcção do olhar de um utilizador. Vimos anteriormente que a direcção do olhar é dada por dois vectores, que começam no centro dos olhos, intersectam o centro das íris (as pupilas), e encontram-se no alvo para onde o utilizador olha. Assim neste módulo começamos por identificar e marcar o centro das íris de um utilizador. A finalidade nesta parte é ter um módulo que funcione em tempo real, capaz de devolver as coordenadas dos centros das íris de um utilizador em frente ao computador. A abordagem tomada para a detecção um utilizador em frente de um computador é com um classificador HAAR para caras (Viola & Jones, 2001), (Menezes, Barreto, & Dias, 2004). Temos assim já na imagem captada pela câmara a região que corresponde à cara do utilizador.



Figura 9 - Imagem capturada pela webcam.

Recorrendo a um classificador HAAR para a identificação de caras, já implementado no *OpenAR*, obtém-se numa imagem as regiões de interesse em que cada região contém uma cara. Uma vez que só se pretende identificar um utilizador, e o classificador pode, caso existam várias caras na imagem, devolver várias regiões de interesse, há que limitar à região que mais interessa. Numa imagem captada pela webcam de um portátil, à partida, a cara que tem mais interesse será a que está mais perto, e sendo a que está mais perto, na imagem, é a maior.

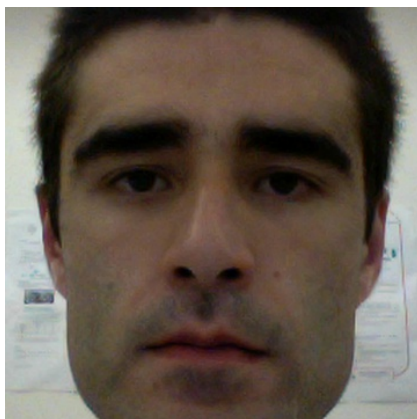


Figura 10 - Cara encontrada da imagem.

Limita-se assim de várias possíveis caras numa imagem, a maior, a que está mais perto. Quanto maior for a área de pesquisa, ou quanto mais caras existam na imagem, maior será o atraso gerado entre captura e apresentação da imagem; assim, após encontrar a cara de interesse, marca-se uma janela de pesquisa que será usada no *frame* seguinte (Figura 19, retângulo a azul em torno da cara); caso aconteça não ser encontrada a cara dentro dessa janela de pesquisa, volta-se ao início, ou seja, à área total de pesquisa. Fazendo isto minimizou-se nitidamente os atrasos que este método pode introduzir.

Neste momento já só estamos a lidar com a imagem da cara, vamos tentar agora encontrar os olhos. Tentou-se algumas abordagens baseadas apenas em heurística, mas era difícil garantir a região dos olhos nalguns casos de rotação da cabeça. Usou-se novamente um classificador HAAR, desta vez para os olhos. O classificador é por olho, assim temos um só para o olho esquerdo, outro para o olho direito. Como já se estava consciente do atraso gerado pela pesquisa destes classificadores e para não se obterem falsos positivos, começou-se por limitar as regiões de pesquisa por heurística. Por exemplo, para o olho direito, este está na metade superior da cara, por sua vez na metade direita da cara, nunca está mesmo no topo da cara, e nunca está mesmo na parte mais exterior da cara, assim sendo, só nesta janela pequena é que se pesquisa o olho usando o HAAR.



Figura 11 - Janela de pesquisa para o olho direito.

Segue-se a identificação das íris do utilizador. Mas primeiro é necessário saber em que região da cara estão as íris. Por heurística limita-se a zona de pesquisa dos olhos, e por uma necessidade de simplificação, volta-se a usar classificadores HAAR, desta vez para os olhos. Para a detecção das íris do utilizador foram testadas várias técnicas, entre as quais uma aproximação destas a círculos, usando assim a transformada de Hough para circunferências. A transformada de Hough é um algoritmo que tenta enquadrar pontos em determinadas estruturas geométricas. É um algoritmo baseado em acumuladores; quando um determinado conjunto de pontos acumulados satisfaz a condição de pesquisa é marcada aí a estrutura geométrica que se está a tentar encontrar. Para a detecção de circunferências fazem-se coincidir estas com pontos encontrados numa imagem; normalmente tem associado um limiar mínimo de pontos encontrados para a identificação de um conjunto de pontos como fazendo parte de uma circunferência. Tal como se pode ver na Figura 12, a transformada de Hough começa por traçar circunferências com um determinado raio de pesquisa em pontos encontrados, as circunferências vão-se intersectando num mesmo ponto e esse ponto será o centro da circunferência que se pretende encontrar. O resultado da aplicação da transformada de Hough para circunferências são os centros das circunferências e os seus raios. Para a aplicação em questão, o centro das circunferências encontradas corresponderia aproximadamente aos centros das pupilas.

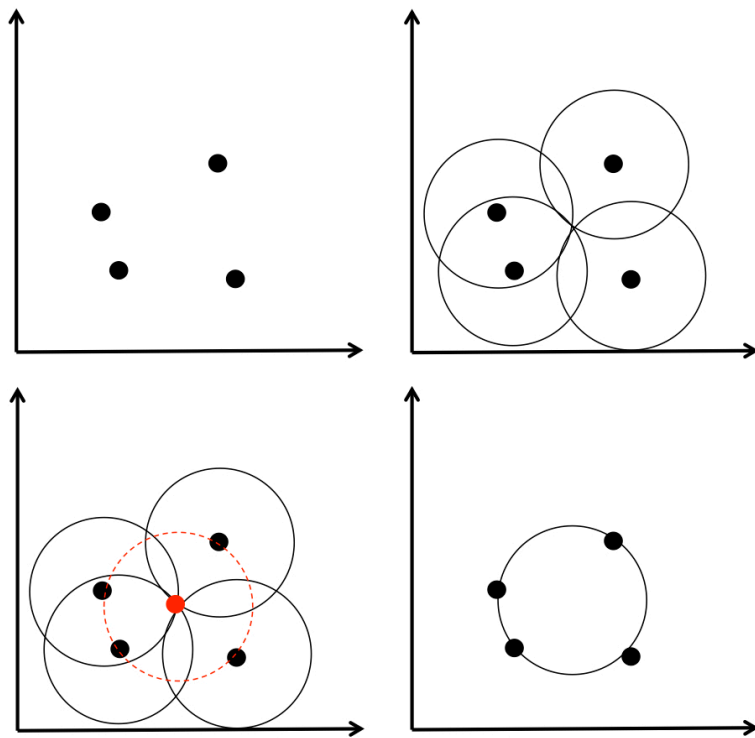


Figura 12 - Exemplo da aplicação da transformada de Hough.

Sendo a íris um corpo que na imagem é quase circular, seria óbvio o uso desta transformada para a detecção das íris. Na transformada de *Hough* temos alguns parâmetros que podemos afinar para a obtenção de circunferências, o raio mínimo e máximo de pesquisa, a distância mínima a considerar entre circunferências e o limiar de pontos a partir do qual se pode considerar uma circunferência. Com a transformada de *Hough* não se conseguiu garantir que o círculo encontrado era o da íris, tal como se pode ver pela Figura 13, mesmo limitando os raios de pesquisa mínimo e máximo, limitando os limiares de aceitação, e mesmo a distância entre círculos. A sequência que se observa na Figura 13 é captada com os mesmos parâmetros de entrada para a transformada de Hough para circunferências. Repare-se que tanto são encontradas falsas íris, como encontradas certas.



Figura 13 - Transformada de Hough aplicada a um olho.

Apesar de muitas vezes existir um círculo encontrado em torno da íris, este não era o único encontrado, e não havia método que garantisse sempre qual o correto. Muitas vezes a transformada de *Hough* era um processo lento para encontrar circunferências. Por estas razões também a transformada de *Hough* para circunferências foi descartada para uso nesta aplicação.

Para a detecção das íris do utilizador testou-se também outra abordagem, a utilização de histogramas da imagem da zona do olho. Pretendia-se através do histograma da imagem verificar uma transição notória na imagem, que marcasse a íris do utilizador. Verifica-se na Figura 14, que a zona da íris corresponde a uma mancha escura de pixéis que ocupa sensivelmente um sexto da imagem do olho. Esperava-se assim ver no histograma uma elevação maior que corresponderia a este grupo de pixéis, e que houvesse uma separação nítida deste grupo de pixéis.

Na verdade no histograma da imagem do olho não é visível uma variação nítida que marque uma íris (Figura 14).

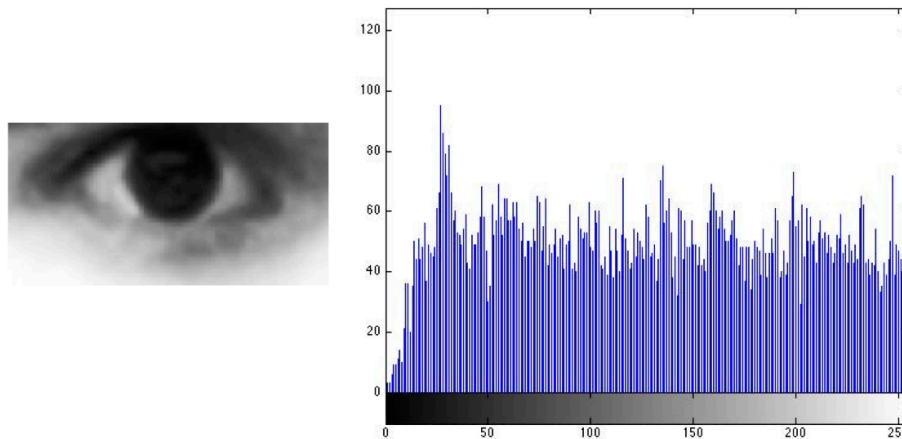


Figura 14 - Olho e correspondente histograma.

Tentou-se ainda fazer uma alteração ao histograma de forma a obter um acumulador de tons de pixéis por linhas e por colunas. Esperava-se que daí resultasse uma curva com um mínimo global que seria o centro da íris; no centro da curva originada por acumular os tons por coluna ficaríamos com a coordenada “ u ”, e no centro da curva originada por acumular os tons por linha, ficaríamos com a coordenada “ v ” do centro da íris (u,v).

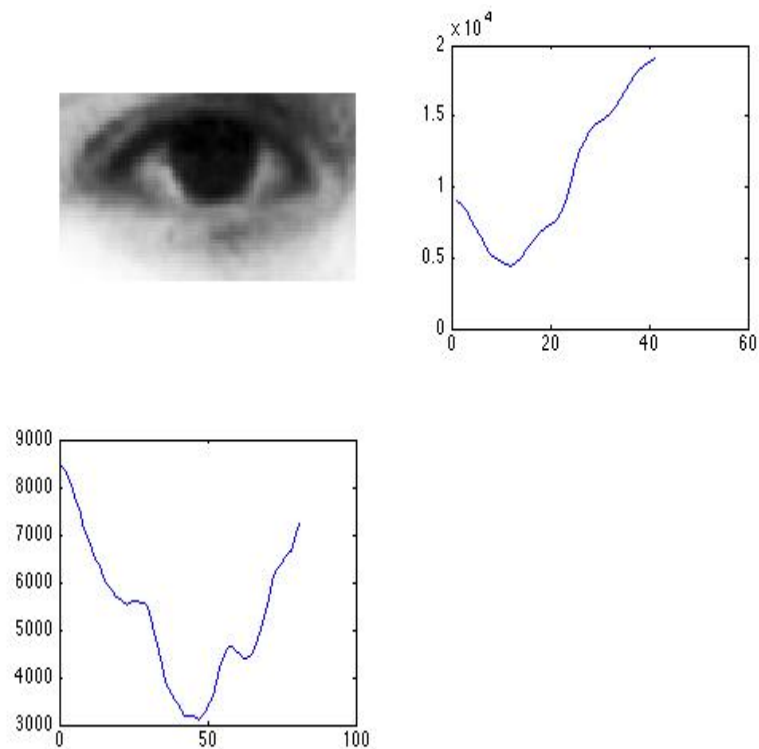


Figura 15 - Olho, usando acumuladores por linhas e por colunas.

Apesar de na Figura 15 ser possível verificar que há uma relação entre o centro da íris e os mínimos das curvas formadas pelos acumuladores por linhas (gráfico ao lado direito da imagem do olho) e por colunas (gráfico abaixo da imagem do olho), não é possível garantir que seja sempre evidente; por vezes verificava-se que as sombras interferiam bastante neste processo, e nalgumas pessoas, por causa das pestanas a distribuição de um tom escuro era mais uniforme ao longo da imagem, o que impossibilitava garantir que os mínimos encontrados nos gráficos correspondiam ao centro da íris, e consequentemente ao centro da pupila.

Um outro teste que se efetuou para a identificação das íris é o uso da binarização para o cálculo do centro de massa. Numa imagem binarizada, o cálculo do centro de massa de uma zona branca é o seu centro de massa; ou seja, se tivermos um círculo branco em fundo preto, o resultado do cálculo do centro de massa desta imagem corresponde ao centro do círculo branco.

Pretende-se aproximar o centro das íris do utilizador ao centro de massa de uma imagem binarizada do olho em que se evidencia a íris. Assim, para binarizar a imagem e evidenciar as íris do utilizador requer-se primeiro algum processamento da imagem.

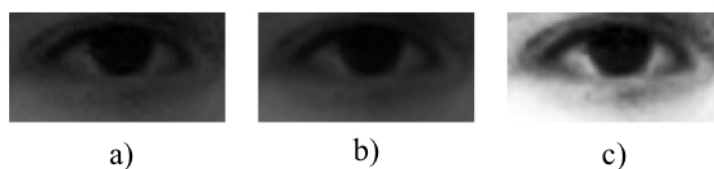


Figura 16 - Processamento da imagem do olho.

Para a segmentação das íris começamos por utilizar um filtro gaussiano (Figura 16, b) à imagem captada do olho (Figura 16, a) de forma a atenuar ruído da imagem. Em seguida faz-se uma equalização do histograma da imagem (Figura 16, c), que significa espalhar os tons da imagem em todo o seu espectro, por exemplo, uma imagem em cinzento de 8 bits em que só tenha tons até ao 128, a equalização do histograma força a espaçar os tons de forma a que o anterior máximo de 128 passa a ser de 255.

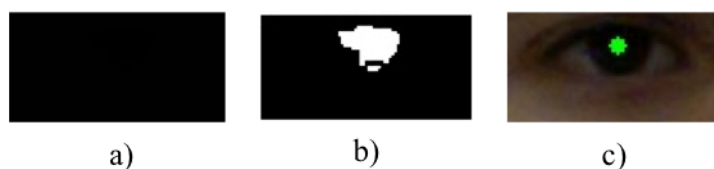


Figura 17 - Obtenção do centro da íris.

Após a equalização do histograma passa-se a uma binarização baseada na média da imagem (Figura 17, a). Apesar de não ser visível, há efetivamente pixéis com valores diferentes - alguns a nível “0” e outros a nível “1” - não sendo perceptível essa diferença pelo olho humano; o interessante de ser baseada na média da imagem é que permite alterações de brilho mantendo quase sempre o mesmo resultado final. Após a binarização (Figura 17, b) já temos a imagem num formato adequado para o cálculo do centro de massa. Para visualização do centro de íris encontrada, esta é marcada na imagem (Figura 17, c). Sendo que o *OpenCV* está bastante agilizado para este tipo de funções de processamento de imagem, este método tem como principal vantagem a rapidez de resultado, não transportando atraso significativo.

Este último método foi o escolhido para a aplicação de Seguimento dos olhos. Pode-se assim resumir a aplicação com o fluxograma que se apresenta na Figura 18.

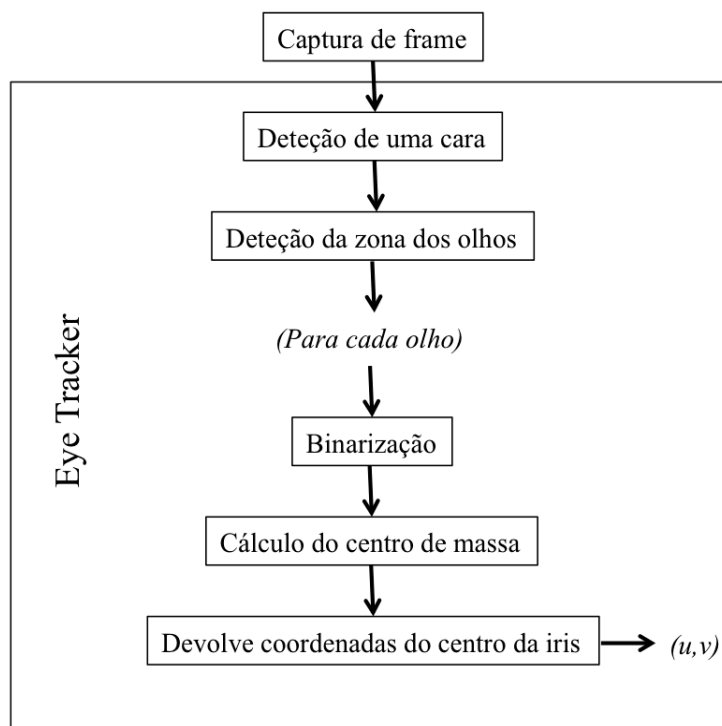


Figura 18 - Fluxograma do módulo de Seguimento dos olhos.

Na Figura 19 vemos o resultado final do módulo de *Seguimento dos olhos*. Mostrar esta imagem final serve apenas como output visual, pois este módulo só necessita de devolver as coordenadas dos centros de cada íris.

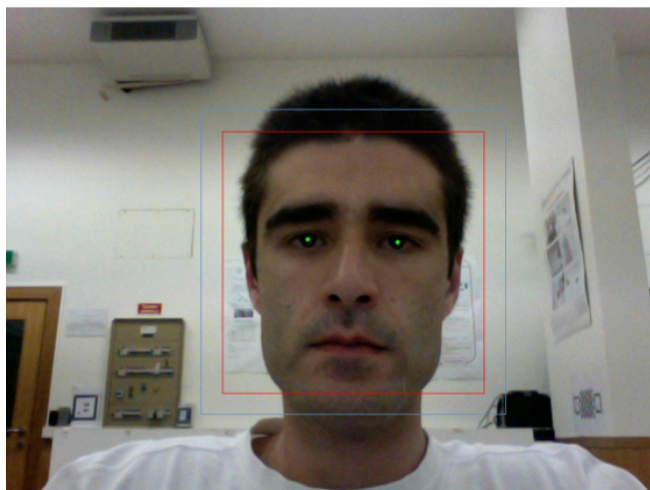


Figura 19 – Resultado final do Seguimento dos olhos

Este módulo foi testado por várias pessoas sendo robusto nos resultados obtidos. Enquanto a pessoa está de frente para a câmara, e numa distância normal de interação com o computador (<50cm), desde que a cara e os olhos sejam visíveis, as pupilas são sempre encontradas. As falhas detectadas verificaram-se em utilizadores em que de alguma forma a

imagem captada do olho é afetada, quer pelo uso de óculos ou maquilhagem. A iluminação também é um factor preponderante no funcionamento do módulo de seguimento dos olhos, iluminação indireta é preferenciada.

Analisando as imagens obtidas dos olhos, e as coordenadas dos centros das íris, é possível calcular a sua excursão máxima em pixéis. Como o tamanho da imagem captada do olho varia com a distância, vamos apenas considerar a distância normal de utilização de um computador portátil, aproximadamente 50cm, e a imagem captada pela webcam de 900 por 680 pixéis. A esta distância, usando o sistema de seguimento dos olhos obtém-se uma excursão máxima da íris de aproximadamente 25 pixéis, traduzindo-se numa resolução angular de aproximadamente 3.5 graus por pixel, em movimentos horizontais. Já em movimentos verticais, verificaram-se 13 pixéis de movimento, sendo que a resolução diminui para os 6 graus por pixel.

No final deste módulo temos já as coordenadas dos centros das íris do utilizador, mas isto não é suficiente para a Estimação da direção do olhar do utilizador, como se pode verificar pela Figura 7. Para a estimação da direção do olhar de um utilizador é necessário conhecer as coordenadas 3D das íris e também dos centros dos globos oculares. Só assim se podem traçar dois vectores que se hão de intersectar num ponto, o ponto para onde o utilizador olha. Como os olhos do utilizador estão dentro da cabeça deste, vamos a seguir estimar a pose da cabeça do utilizador.

2.3 - Estimação da pose da cabeça

Com este módulo pretende-se ter uma aplicação que seja capaz de estimar a *pose* da cabeça do utilizador em relação à webcam. A pose da cabeça é necessária para conhecer a posição 3D do centro dos olhos e a posição 3D das íris do utilizador. Para a estimação da *pose* da cabeça será necessário conhecer a câmara, e um modelo físico da cabeça.

Considere-se uma câmara calibrada, ou seja que se conhecem os seus parâmetros intrínsecos; este processo é efectuado recorrendo a um método que já se encontra implementado no OpenAR (ISR/DEEC). Considere-se assim que a câmara está calibrada, conhecem-se assim os parâmetros intrínsecos e os coeficientes de distorção. Os parâmetros intrínsecos de uma câmara estão relacionados com as características físicas da câmara e por isso são imutáveis ao longo do tempo (desde que o comprimento focal não se altere, o que não acontece numa câmara com zoom óptico). Há uma relação matemática que relaciona o mundo observado por uma câmara com a imagem que esta captura; consideraremos para tal o modelo de câmara *pinhole*. O

modelo de câmara *pinhole* é uma boa aproximação de uma câmara convencional; consiste numa caixa cúbica, fechada, com um pequeno orifício numa das faces, por onde os raios de projeção vão passar e projetar-se na face oposta à do orifício (Figura 20).

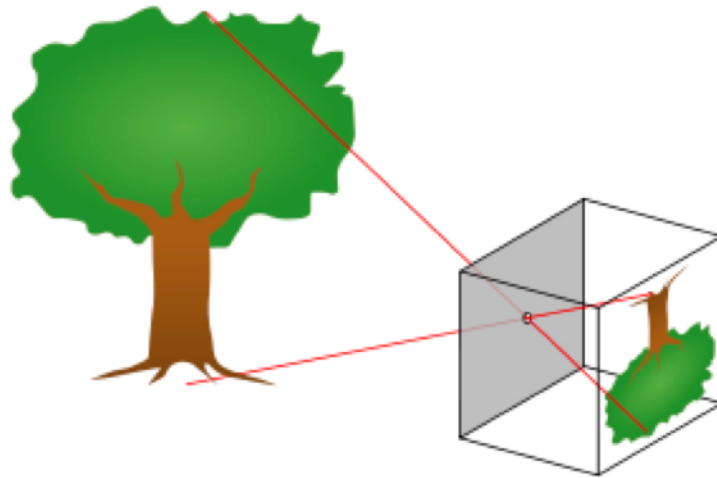


Figura 20 - Ilustração do modelo de câmara *pinhole*

A cena é registada através da projeção de pontos 3D do mundo, no plano de imagem, através de uma transformação de perspectiva. O modelo matemático (K) que descreve este tipo de câmaras representa essencialmente a distância focal (f_x, f_y) e as coordenadas do ponto principal (c_x, c_y)

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

em que as coordenadas do ponto principal representam o ponto em que há interseção do eixo óptico da imagem com o plano da imagem; o eixo óptico é o único raio perpendicular ao plano de imagem. Através deste modelo de câmara podem-se assim relacionar pontos do mundo 3D para pontos da imagem capturada.

$$s p' = K [R|t] P'$$

ou de outra forma

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Onde (X, Y, Z) são coordenadas de um ponto 3D nas coordenadas do mundo, (u, v) são coordenadas de um ponto de projeção em pixéis. A matriz K , tal como já vimos acima, é a matriz

de parâmetros intrínsecos, a matriz $[R|t]$ é uma matriz expandida de rotação e translação, também muitas vezes referida como matriz de parâmetros extrínsecos, o resultado é um ponto da imagem com coordenadas (u, v) com um factor de escala “s”. A matriz de parâmetros extrínsecos serve para descrever um movimento rígido da câmara em relação a uma cena, ou da cena em relação à câmara. Portanto a matriz de parâmetros extrínsecos traduz as coordenadas de um ponto (X, Y, Z) a algum sistema de coordenadas fixo em relação à câmara. Para um valor $Z \neq 0$, a relação pode ser descrita da seguinte forma:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t$$

$$x' = \frac{x}{z}$$

$$y' = \frac{y}{z}$$

$$u = fx * x' + cx$$

$$v = fy * y' + cy$$

Mas como as lentes reais têm distorção, principalmente radial e tangencial, o modelo acima estende-se:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t$$

$$x' = \frac{x}{z}$$

$$y' = \frac{y}{z}$$

$$x'' = x'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 x' y' + p_2 (r^2 + 2x'^2)$$

$$y'' = y'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1 (r^2 + 2y'^2) + 2p_2 x' y'$$

onde: $r^2 = x'^2 + y'^2$

$$u = fx * x'' + cx$$

$$v = fy * y'' + cy$$

onde k_1, k_2, k_3 são os coeficientes de distorção radial e p_1, p_2 são os coeficientes de distorção tangencial. Os coeficientes de ordem superior não são considerados no *OpenCV*. Estes coeficientes de distorção fazem também parte dos parâmetros físicos da câmara, e são obtidos pelo mesmo módulo já implementado para a calibração de câmara no OpenAR, e este módulo devolve apenas k_1, k_2, p_1, p_2 , sendo assim $k_3 = 0$.

Para o modelo físico da cabeça escolheu-se um elipsoide (Figura 21), sendo este modelo o mais semelhante.

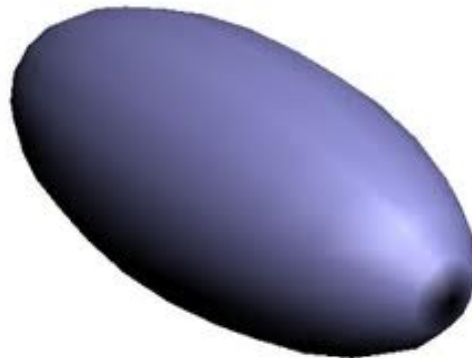


Figura 21 – Exemplo de um elipsoide.

Um elipsoide é representado pela superfície quádrlica:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

Um elipsoide pode também ter a representação matricial:

$$Q = \begin{bmatrix} 1/a^2 & 0 & 0 & 0 \\ 0 & 1/b^2 & 0 & 0 \\ 0 & 0 & 1/c^2 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

sendo este o formato (matricial) que será usado para a representação e operação.

Os valores a , b e c são em mm em que $a = metade da largura da cabeça$, $b = metade da laltura da cabeça$ e $c = metade da profundidade da cabeça$. As medidas são estáticas de tal forma que sejam aceitáveis para a grande maioria das pessoas.

Para o cálculo da *pose* precisamos de primeiro relacionar o modelo da cabeça com as imagens obtidas pela câmara. Fazemos isso recorrendo aos *pontos SURF*; estes pontos na imagem são pontos dos quais se conhece as suas coordenadas na imagem (u,v) e um descritor (a sua assinatura).

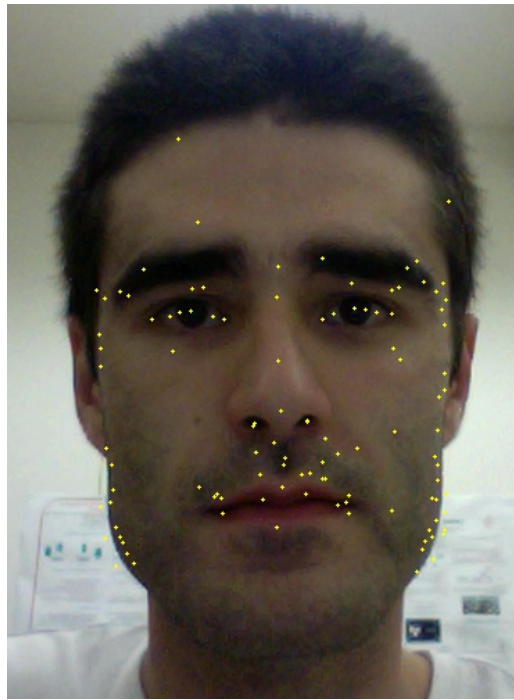


Figura 22 - Obtenção de pontos SURF da cara

Com os pontos *SURF* consegue-se fazer *tracking* destes de *frame* para *frame*. Numa fase de estabelecer a referência, pegamos na imagem obtida, identifica-se uma cara; o tamanho da cara é usado para estimar de uma forma aproximada a distância do utilizador, com o objectivo de saber onde colocar o modelo da cabeça (Q) nas coordenadas do mundo. Da imagem da cara obtém-se pontos SURF e conhecendo já as relações entre coordenadas da imagem e coordenadas do mundo faz-se a reprojeção dos pontos da imagem no modelo da cabeça. Faz-se esta reprojeção resolvendo a equação

$$P^T Q P = 0$$

$$P = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

sendo que P é um ponto que pertence à superfície do elipsoide. O ponto P é, no entanto, também um ponto pertencente a um vector que começa no centro da câmara, passa pela imagem obtida e há de intersectar o modelo da cabeça

$$P = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + \lambda \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

sendo (x,y,z) obtidos pela relação já apresentada mais acima aquando da apresentação do modelo de câmara. Resolvendo agora $P^TQP = 0$, estamos a resolver um sistema em relação a uma incógnita, “ λ ”, que sendo possível resolver (é impossível resolver se o vector não intersectar o modelo), dá dois valores, porque o vector intersecta dois pontos do modelo; estamos interessados apenas no valor que está mais próximo da câmara, o que corresponderá à “cara” do modelo. Repare-se que neste momento passamos a ter pontos 3D que estão relacionados com pontos 2D numa imagem de referência, destes pontos como são *SURF points*, conhecemos também a sua assinatura.

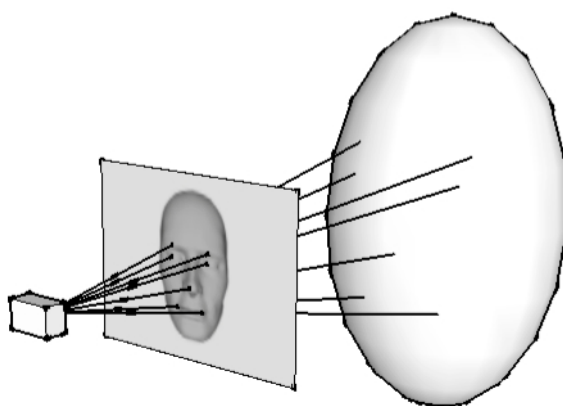


Figura 23 - Projeção de pontos da cara no modelo da cabeça.

Pela janela devolvida pelo HAAR para identificação de caras é possível obter uma boa aproximação da posição do modelo do elipsoide. Estabelece-se uma medida para a janela que envolve a cara caso aquela fosse real. Como conhecemos as medidas da janela que envolve a cara no mundo 3D e conhecemos as medidas (em pixéis) na imagem captada, podemos estimar a translação, ou seja, os valores T_x , T_y e T_z da janela do HAAR em relação à câmara. Coloca-se assim o modelo físico da cabeça (elipsoide) nessa posição.

A pose é estimada de *frame* para *frame* entre um modelo no qual se conhecem alguns pontos, e pontos obtidos numa imagem. A relação entre os pontos do modelo e os pontos da imagem é a da vista em perspectiva. Assim, os pontos do modelo são pontos da cara projetados num elipsoide. Estes pontos são colocados numa altura em que se supõe que se conhece a *pose* da cabeça do utilizador. Garante-se isso pedindo ao utilizador que se coloque perfeitamente frontal com a câmara, para podermos colocar nessa posição o modelo físico da cabeça, não tendo que nos preocupar com nenhum ângulo da cabeça; isto é conseguido logo na primeira *frame* obtida e ficam os pontos deste modelo gravados como referência para os *frames* seguintes. É

apresentado na Figura 24 sobre a forma de fluxograma o estabelecimento da referência para a estimação da pose da cabeça do utilizador.

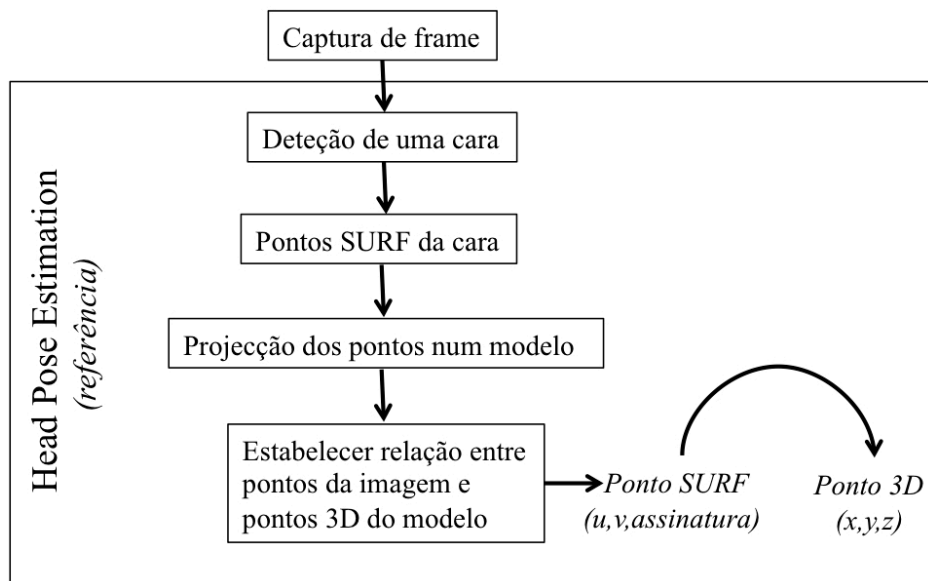


Figura 24 - Fluxograma do estabelecimento de referência para a estimação da pose da cabeça.

Nos *frames* seguintes é que começa a estimação da pose da cabeça. A cada *frame* é feita a extração dos pontos SURF da cara, é feito um emparelhamento destes pontos com os de referência (que têm uma relação com pontos 3D do modelo). Na Figura 25 é possível observar o emparelhamento de pontos SURF encontrados num *frame*, com os pontos SURF marcados no *frame* de referência.



Figura 25 - Emparelhamento de pontos SURF.

Assim são passados a uma função os pontos SURF desta imagem, os pontos 3D do modelo (que têm uma relação de emparelhamento com estes pontos), os parâmetros intrínsecos da câmara e os coeficientes de distorção; a função resolve a rotação e a translação que os pontos do modelo (3D) têm que ter de modo a satisfazer a projeção (pontos da imagem).

A estimação da *pose* da cabeça é feita pelo cálculo da matriz de parâmetros extrínsecos que satisfaz a projeção de pontos de um modelo que se conhece. Novamente recorrendo à relação apresentada no modelo de câmara, que relaciona pontos do mundo 3D com as coordenadas dos pontos na imagem projetada, é possível fazer a estimação da *pose*. Para tal recorreu-se a dois algoritmos diferentes do *OpenCV*, que através das coordenadas de pontos numa imagem, pontos de um modelo físico, e os parâmetros intrínsecos da câmara com os coeficientes de distorção, devolvem a matriz dos parâmetros extrínsecos. Usámos o POSIT (DeMenthon & Davis, 1995), e um algoritmo que resolve o PNP (Rosenhahn) para a estimação de *pose*.

Verificou-se que o POSIT para estimar a *pose* da cabeça era por vezes muito lento na devolução da *pose*, porque se trata de um algoritmo iterativo. Assim este algoritmo foi descartado para esta aplicação.

Para o outro algoritmo, o PNP, usou-se a função do *OpenCV* “*cvFindExtrinsicCameraParams2*”, os parâmetros de entrada são os parâmetros intrínsecos da câmara, parâmetros de distorção da câmara, pontos 3D do objecto e pontos da projeção do objecto (pontos imagem), e devolve uma matriz de translação e outra de rotação, que definem uma *pose*. Para verificar o correto emparelhamento dos pontos, usa-se a *pose* calculada e aplicamos aos pontos do modelo 3D. Fazemos a projeção destes na imagem e calcula-se a diferença entre eles; se existir um erro grande nalgum par de pontos (pontos SURF desta *frame* e pontos do modelo projetados), este par é retirado do cálculo da *pose*, e apenas os pontos bem emparelhados seguem para a estimação final da *pose*.

Volta-se assim a fazer a estimação da *pose*, e esta será agora correta (ou melhor que a anterior calculada). Na Figura 26 é possível verificar visivelmente a projeção de pontos do modelo (na figura, são os pontos a vermelho), com os pontos SURF extraídos neste *frame* para o cálculo da *pose* (na figura a amarelo). O facto de estarem quase coincidentes demonstra a correta estimação da *pose*. No entanto, verificou-se experimentalmente que, por vezes, a *pose* era mal calculada (quando por exemplo acontecia maus emparelhamentos no SURF), e verificava-se no Tz (translação no eixo dos zz) variações superiores a 100mm (entre um *frame* bem estimado, e o seguinte mal estimado). É fisicamente pouco provável que de um *frame* para o seguinte, uma

pessoa mude de posição em 100mm. Conscientes desse facto, se uma estimação de *pose* implicar uma variação destas, é ignorada e atribuída a *pose* anterior, considerada correta.

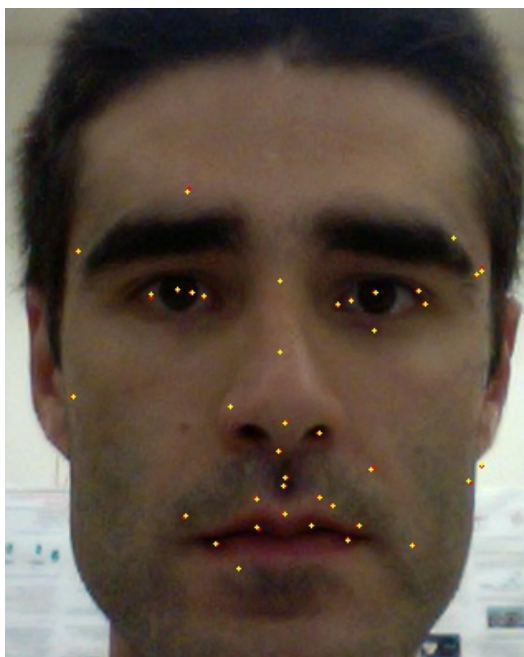


Figura 26 - Verificação da projeção de pontos.

A aplicação de estimação da pose da cabeça do utilizador termina devolvendo esse mesmo dado. Assim para a tarefa geral de estimar a direção do olhar de um utilizador é ainda necessário primeiro relacionar o modelo da cabeça com os modelos físicos dos olhos do utilizador. No próximo módulo, junta-se o módulo de Seguimento dos olhos ao de Estimação da pose da cabeça para que seja possível a Estimação da direção do olhar do utilizador.

2.4 – Estimação da direção do olhar

Nesta secção conhecemos a *pose* da cabeça e a posição das pupilas na imagem captada pela câmara, dados devolvidos pelos dois módulos anteriores. Para traçar os vectores de olhar de cada olho precisamos das coordenadas 3D dos centros dos globos oculares e das coordenadas 3D dos centros das pupilas. Como conhecemos a *pose* da cabeça, conhecemos também a *pose* de cada uma das esferas que representam os globos oculares.

Usando o módulo de *Estimação da pose da cabeça*, na fase de estabelecimento de referência, precisamos de marcar os centros 3D dos olhos no modelo do elipsoide. Isso é feito da mesma forma que com os pontos SURF, mas nesta secção é feito com as pupilas encontradas pelo algoritmo de *Seguimento dos olhos*.

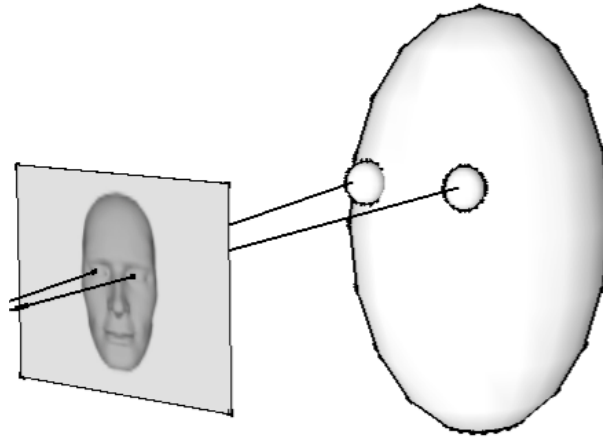


Figura 27 - Projeção das pupilas no modelo da cabeça.

Na Figura 27 mostra-se graficamente o que é feito para encontrar os centros dos olhos. Recorre-se ao mesmo modelo que o usado já na estimação da pose da cabeça, com a inclusão das duas esferas circulares (de diâmetro 24.5mm e distam entre si de 65mm). No modelo da cabeça, Este conjunto das duas esferas é colocado, nos pontos de intersecção dos raios que passam pelas pupilas.

Podendo assim estimar de *frame a frame* e usar a pose da cabeça para conhecer a posição 3D exata dos centros dos olhos.

Considere-se a este ponto o seguinte referencial de eixos: o eixo dos *zz* cresce à medida que nos afastamos da câmara, o eixo dos *yy* cresce para cima, e a origem é no centro da câmara.

Para marcar as pupilas, sabe-se que estão projetadas num plano mais próximo da câmara que dista 12.25mm do centro do olho. Vejamos então como exemplo, num determinado instante, que a estimação do centro do olho seria no ponto $(x,y,z)=(30,40,330)$ mm. Sabendo que a projeção da pupila é feita num plano $z=317.75$ mm e usando as relações de perspectiva já apresentadas, calculamos as componentes “x” e “y” do ponto 3D da pupila (conhecendo a coordenada “z”). Este processo é como traçar um raio de projeção da imagem captada pela câmara de uma pupila até ao plano distante que lhe corresponde.

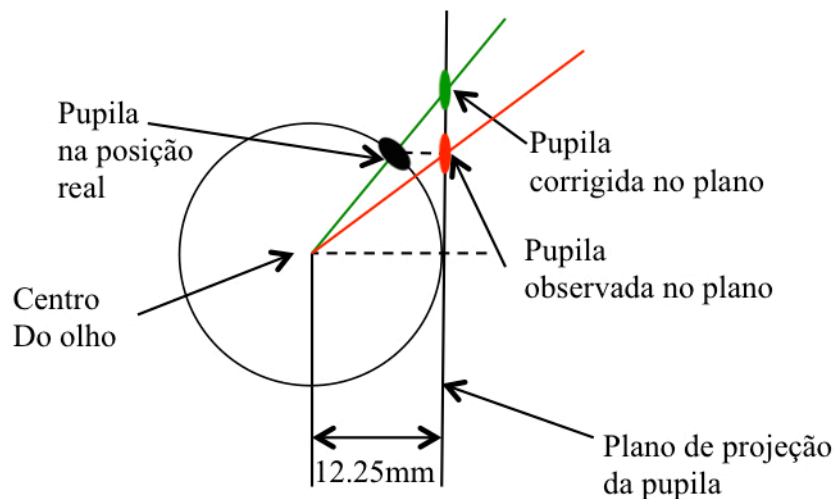


Figura 28 - Correção da posição da pupila.

Conforme se pode verificar na Figura 28 o traçar do raio até um plano que é paralelo ao plano da câmara faz com que a estimação do *gaze* (linha vermelha na figura) seja errada. Mas como se conhecem todas as dimensões, corrige-se a posição da pupila, sendo um problema trigonométrico. Este problema é resolvido conhecendo a altura entre a pupila observada e o centro do olho, este valor é igual ao seno do ângulo que a linha verde faz com a perpendicular entre o plano e o centro do olho, fazendo a tangente desse ângulo fica-se com a posição corrigida da pupila. Esta operação é feita nas duas coordenadas x e y , sendo que na figura só está representada a correção em uma. Verifica-se que a linha verde (Figura 28), é a que tem a estimação correta do *gaze*, sendo um vector que começa no centro do olho, passando pela pupila e segue até interseção num plano.

Nesta etapa temos já dois vectores, um por cada olho; conhecemos os pontos 3D de cada centro e conhecemos os pontos 3D de cada pupila. Sabemos que o monitor está no mesmo plano que o da câmara (plano $z=0$). Assim estendemos os vectores de cada olho até ao plano $z=0$ determinando as coordenadas (x,y) em que cada um intersecta. A estimação do *gaze* é feita pelo ponto médio dos dois pontos obtidos. Seria de esperar que um utilizador ao olhar para um ponto do plano, a interseção de cada um dos vectores dos olhos fosse no mesmo ponto (Figura 7). Observa-se experimentalmente que a interseção dos dois vectores com um plano é feita em dois pontos diferentes muito próximos, assim, o ponto marcado como *gaze* é o ponto médio dos dois pontos obtidos pela interseção dos vectores no plano $z=0$.

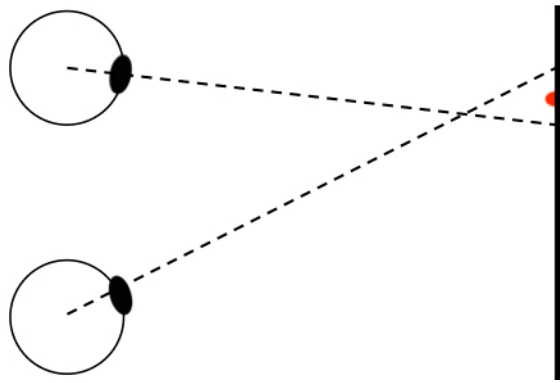


Figura 29 - Estimação do gaze.

Na Figura 29 está representado o fenómeno descrito anteriormente. O ponto a vermelho na figura é o ponto estimado como o *gaze* do utilizador no plano da câmara.

Na Figura 30 apresenta-se sobre forma de fluxograma o algoritmo aqui desenvolvido para a estimação da direção do olhar.

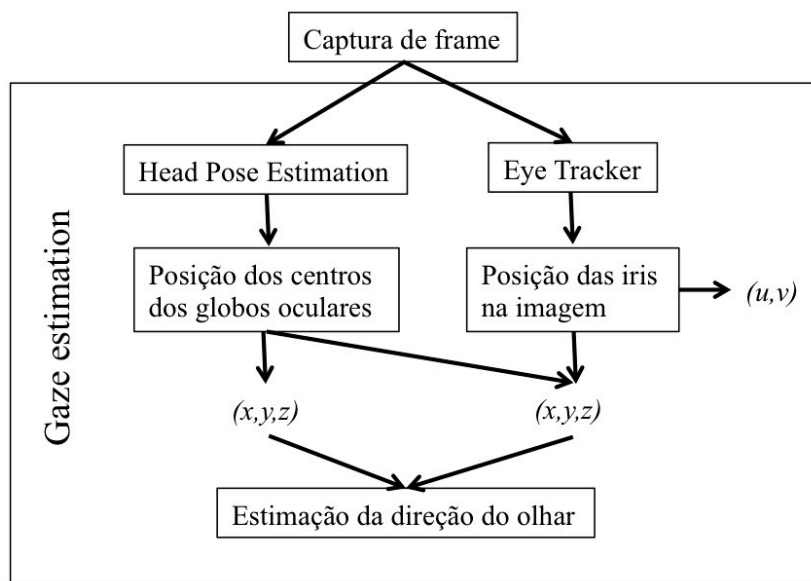


Figura 30 - Fluxograma de Estimação da direção do olhar.

O resultado final desta aplicação de estimação de direção do olhar é um ponto pertencente a um plano.

Neste trabalho temos como objectivo obter a direção do olhar do utilizador em frente de um computador, assim, o ponto vem referenciado ao ecrã do computador, e em coordenadas do monitor, ou seja, em pixéis.

2.5 – Estimação da direção do olhar com o capacete X-View HED

No ISR/DEEC temos disponível um capacete da “SMI” para *Seguimento do olhar*, que já foi referido no “1.1 - Estado de arte”. A principal desvantagem deste é o seu preço, que torna o uso deste incomportável para a grande maioria das pessoas. No entanto, pela simplicidade do capacete, facilmente se verifica que o grande custo associado a este é de *software*, e não do *hardware*. Na perspectiva do *hardware*, o custo será aproximadamente o da compra de duas *webcams*. É interessante produzir *software* capaz de fazer o *Seguimento do olhar*, utilizando este *hardware*, diminuindo convenientemente o custo da aquisição de equipamentos como o capacete.

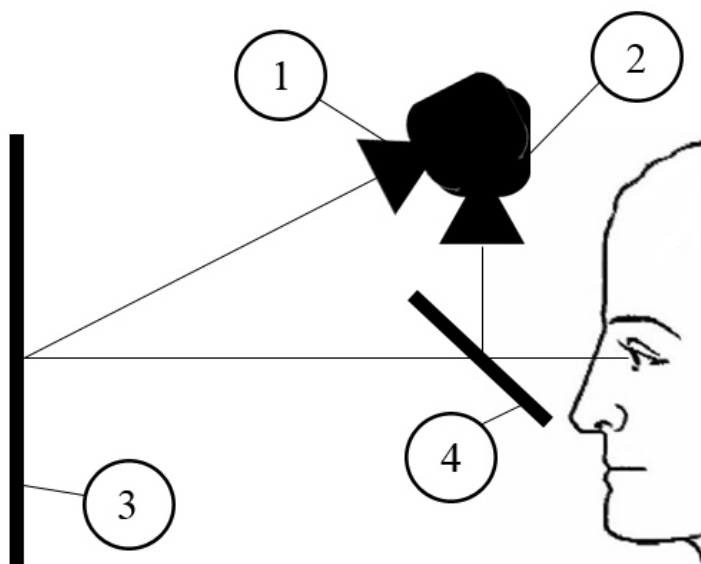


Figura 31 – Seguimento do olhar com o capacete.

Legenda da Figura 31:

- 1 – câmara que tem como funcionalidade observar o mundo;
- 2 - câmara que observa o olho do utilizador;
- 3 – alvo observado pelo utilizador;
- 4 - semi-espelho (para que a câmara 2 não esteja frontal ao utilizador, e possa estar acima, fora do campo visual do utilizador).

Utilizando no código já desenvolvido nesta tese para encontrar os centros das íris seria suposto conseguir-se bons resultados tendo agora como base o capacete para *Seguimento do olhar*. Encontrar os centros das íris com o capacete apresenta grandes diferenças,

comparativamente ao já implementado para a webcam. Como o capacete está fixado à cabeça do utilizador, conhece-se sempre a posição do olho na imagem captada pela câmara de infravermelho do capacete (Figura 31, nr.º 2); assim torna-se completamente desnecessário os classificadores HAAR para restringir as zonas de pesquisa. A própria câmara de infravermelhos, combinada com a iluminação de infravermelhos aumenta significativamente o contraste da imagem, e em última análise, porque é uma imagem captada mais próxima do alvo (o olho), a imagem é maior e de melhor qualidade, conforme se pode verificar pela Figura 32.



Figura 32 - Imagem do olho, captada com recurso ao capacete.

Usando a webcam do portátil para captar a imagem do olho, conseguia-se confinar este a uma janela de 115 por 80 pixéis (para uma distância normal de interação). Usando-se o capacete esta janela aumenta para os 380 por 190 pixéis. A imagem captada do olho, é obtida em tons de cinzento, com um contraste adequado, assim o processamento desta imagem para a segmentação da íris é menor, pelo que se diminui tempo e recursos. Como a imagem do olho captada pela câmara do capacete tem uma resolução desta zona tão alta, e com qualidade de imagem, pode-se mesmo determinar o centro da pupila. Assim modificando-se ligeiramente o código já implementado, mas desta vez ao invés de devolver o centro de uma zona escura que é a íris, aqui calcula-se o centro de uma zona escura que é a pupila.

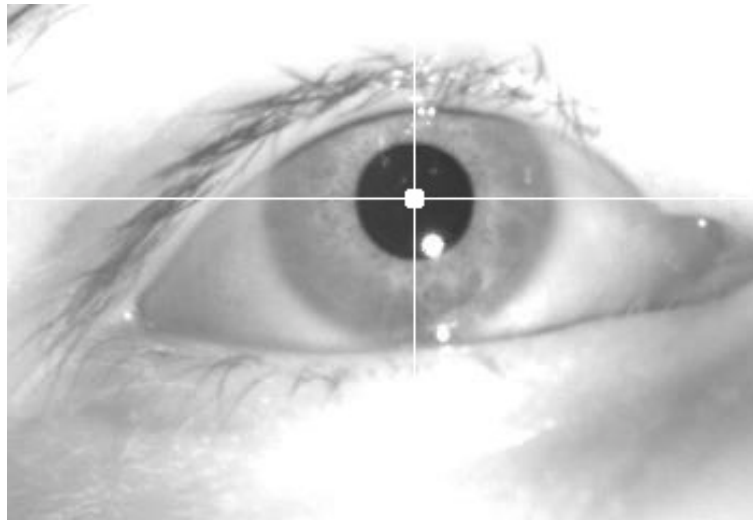


Figura 33 - Marcação do centro da pupila.

Observa-se que na Figura 33 está identificado o centro da pupila. A marcação é feita após uma simples segmentação da imagem, sendo este processo bastante rápido, não acrescentando assim qualquer atraso ao sistema.

Aplicar o código desenvolvido para a detecção de íris no capacete foi bastante importante. Primeiro demonstrou a portabilidade e a universalidade do código desenvolvido, ou seja, até mesmo numa imagem evidentemente diferente; segundo, demonstra que grande parte do erro que se obtém usando a webcam do portátil como *Estimador da direção do olhar* se deve à baixa resolução da imagem captada. Assim, obviamente, espera-se que com uma imagem de maior resolução e iluminação adequada seja possível fazer *Estimação da direção do olhar* com menor erro no contexto do uso de uma câmara de um computador portátil.

Com esta implementação podemos deixar de utilizar o software da SMI e até mesmo ponderar a realização no ISR/DEEC um sistema completo de *Estimação da direção do olhar* semelhante ao *X-View HED*. Está também já implementado o código que estabelece o mapeamento do movimento do olho do utilizador com o movimento do ponto de gaze na imagem captada pela câmara do mundo do capacete (Figura 31, nr.º 1). Relembrar que este sistema de *Estimação da direção do olhar* baseia-se num mapeamento entre a imagem captada do olho e a imagem captada do mundo. Assim os movimentos do olho são mapeados para a imagem captada do mundo. Sendo que é um mapeamento de uma imagem para outra o mapeamento é uma homografia. Para o cálculo desta homografia usam-se cinco pontos. Assim bastará fazer o registo de cinco pontos para onde o utilizador olhou sendo estes pontos correspondentes de movimento do olho. Após este procedimento de calibração temos já na imagem do mundo, um ponto para onde o utilizador está a olhar.

Usando o sistema de seguimento dos olhos com o capacete da SMI verifica-se uma resolução angular horizontal de 0.5 graus e uma resolução vertical de 1 grau.

2.6 - Discussão

Para o desenvolvimento das aplicações usaram-se linguagens de programação como o C/C++, em ambiente Unix, em algumas situações de teste usou-se também o *MatLab*. As aplicações foram desenvolvidas recorrendo a bibliotecas como o *OpenAR* e o *OpenCV* (Kaehler & Bradski, 2008).

No caso de um computador portátil, o plano da câmara é também o plano do monitor. A estimação do gaze do utilizador é assim um ponto que pertence ao mesmo plano que o de origem de coordenadas da câmara do computador portátil, o plano $z=0$, com coordenadas (x,y) em milímetros. Tendo isto, e como conhecemos a posição do monitor em mm em relação à câmara, podemos transformar as coordenadas (x,y) em mm para coordenadas (u,v) em pixéis do monitor.

Como o resultado da aplicação de *Estimação da direção do olhar* é um pixel do monitor, podemos usar esta aplicação, por exemplo, para analisar onde o utilizador olha numa imagem. Criou-se assim uma aplicação que com os resultados obtidos pelo *Estimação da direção do olhar* marca numa imagem o sitio para onde o utilizador está a olhar. Para análise posterior é também possível gravar os pontos para onde o utilizador olhou, sendo que estes pontos podem ser gravados sobre uma imagem.



Figura 34 - Exemplo de aplicação da estimação do gaze.

Na Figura 34 temos um exemplo de uma imagem para onde um utilizador olhou. Os pontos a vermelho são sítios onde o utilizador olhou ligados com linhas, de forma a que se perceba o padrão de olhar do utilizador.

Supondo que o utilizador está a olhar apenas para um ponto, é possível estimar o erro associado a este sistema de *Estimação da direção do olhar*. Assim, verificou-se que quando o utilizador a uma distância de 50cm olhava para um determinado ponto específico no monitor, havia um variação que no ecrã podia chegar aos 3cm, significando assim um erro de 3.5 graus e dessa maneira superior a qualquer equipamento comercial.

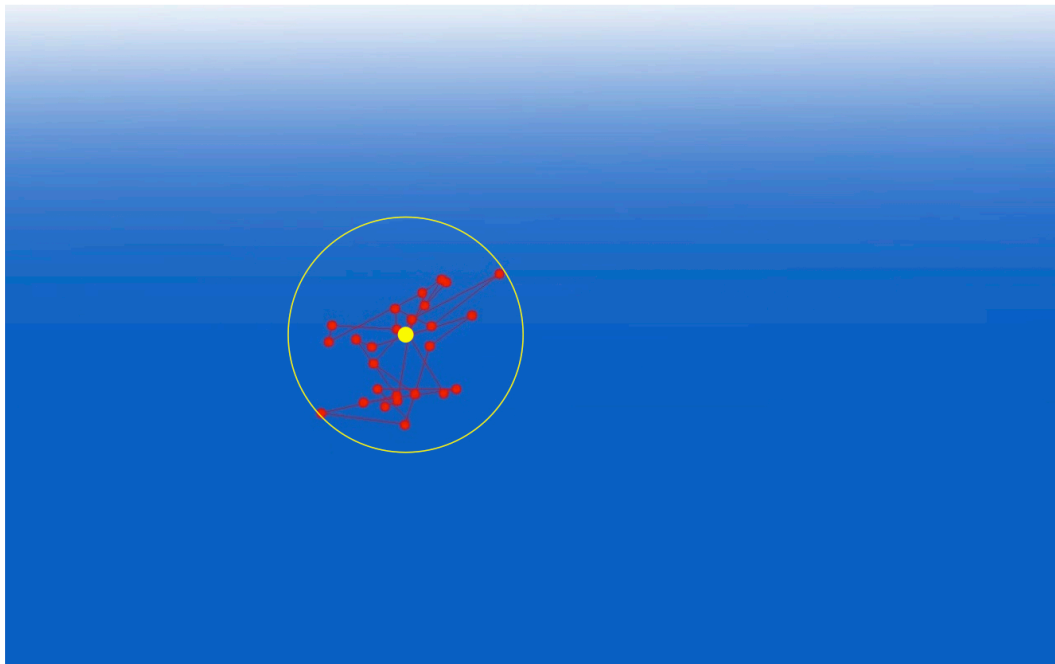


Figura 35 - Teste de sensibilidade.

Na Figura 35 apresenta-se o resultado quando um utilizador olha apenas para um ponto conhecido (a amarelo cheio). A circunferência a amarelo, tem o tamanho real de 6 cm de diâmetro; como o utilizador se encontrava a 50 cm do monitor, comprova-se um erro máximo de 3.5 graus. Ainda assim, a maior concentração de pontos consegue-se circunscrever com uma circunferência de apenas 4 cm de diâmetro, o que significa um erro menor, de aproximadamente 2.5 graus. Ora, se permitirmos algum erro, pode-se dividir o ecrã em quadrados de 4 cm de lado e utilizá-los como botões de grandes dimensões para, por exemplo, uma aplicação de comunicação assistida (Galante & Menezes, 2012).

3 - Conclusão

Esta tese teve como objectivo testar a possibilidade de desenvolver um sistema de *Estimação da direcção do olhar* usando apenas a webcam, sendo que este objectivo foi cumprido. Com o sistema de *Estimação da direcção do olhar* que utiliza apenas a webcam, sem recurso a qualquer outro tipo de equipamento para melhorar a obtenção das pupilas, obtiveram-se resultados que superaram as expectativas. Este sistema de estimação de direcção do olhar de um utilizador tem um erro estimado de aproximadamente 3.5 graus, superior a qualquer equipamento comercial, mas para um sistema, *open-source*, que não necessita de alterações ao equipamento adquirido na compra de um computador portátil, considera-se bastante comportável. Com este tipo de erro, na ordem dos 3cm de variação no ecrã (a uma distância normal de interacção com o computador) torna-se impossível comandar o cursor do rato com os dados obtidos com a estimação da direcção do olhar, num mesmo ambiente de trabalho de um sistema operativo normal; há botões e menus demasiado pequenos para serem controlados com o olhar. Contudo, para uma aplicação de comunicação assistida em que se use botões grandes de aproximadamente 4x4cm, é possível a utilização do sistema de *Estimação da direcção do olhar* implementado nesta tese.

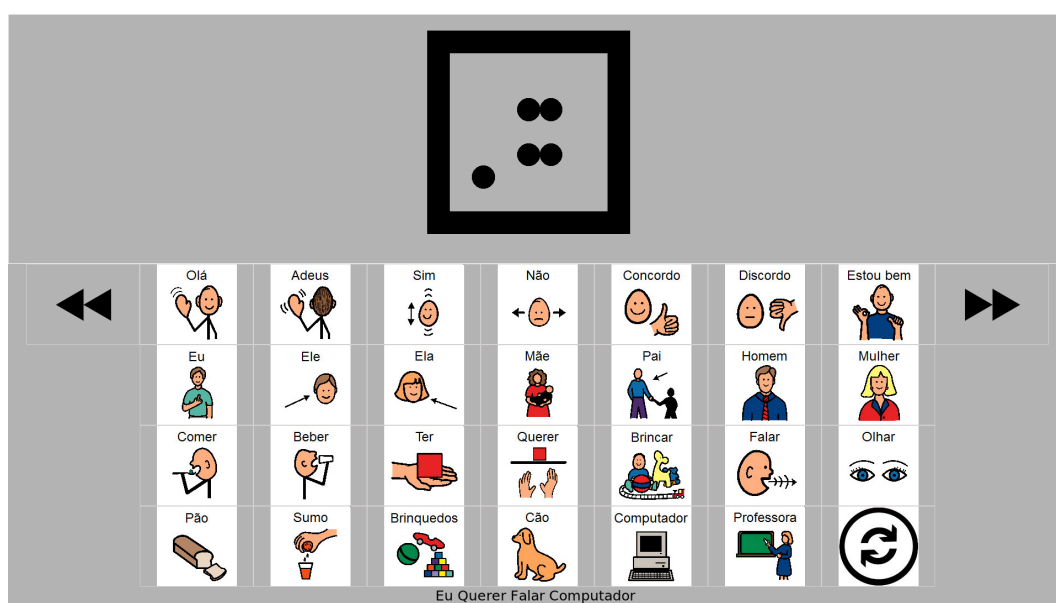


Figura 36 - Exemplo de um programa de comunicação assistida.

É de referir, que está a decorrer um trabalho em que se desenvolveu um sistema de comunicação assistida para ser usado por pessoas com paralisia cerebral sendo que também este sistema utiliza botões grandes, para que seja mais fácil e menos incómodo de utilizar. Este trabalho resultou

num artigo com o título “A gaze-based interaction system for people with cerebral palsy” entregue e aceite para apresentação em conferência (Galante & Menezes, 2012). Na Figura 36 vemos o exemplo de interface de um programa de comunicação assistida -cortesia de Adriano Galante-.

Acrescentando ainda que para uma aplicação do género das usadas para marketing, em que basta uma estimação qualitativa, e se permite algum erro, este sistema de estimação da direcção do olhar demonstrou ser capaz. Em aplicações comerciais deste género são marcados apenas os pontos para onde o utilizador focou. A focagem do utilizador num ponto implica contrações da pupila que os sistemas de *Seguimento dos olhos* conseguem captar porque têm muito boa resolução da zona da pupila. Como o sistema aqui implementado não tem capacidade para a medição das contrações da pupila, os pontos marcados na imagem são pontos por onde o olhar do utilizador passou e, nem tanto pontos para onde o utilizador focou. Portanto, o sistema aqui implementado pode apenas pelo número de ocorrências do ponto do olhar estimar a concentração do utilizador num determinado ponto da imagem.

Usando o capacete da SMI para a aplicação do código aqui desenvolvido, obtiveram-se resultados equivalentes quando comparado ao sistema comercial, com vantagens acrescidas. Grande parte dos trabalhos desenvolvidos no ISR/DEEC são implementados em sistemas operativos baseados em Unix, e o programa nativo da SMI é para Windows. Se for necessário juntar a aplicação da SMI a um trabalho em desenvolvimento é necessário estabelecer sistemas de comunicação para entre os dois sistemas, e por vezes necessário que os dois sistemas corram em computadores separados. O código criado para esta tese corre em sistemas baseados em Unix, facilitando assim a integração de um sistema de *Estimação da direcção do olhar* num outro em desenvolvimento.

Neste momento tem-se disponível na comunidade do ISR/DEEC código pronto a utilizar com o capacete da SMI, pelo que se pode deixar de usar o software providenciado pela marca. Pode ainda este código servir de ponto de partida para a construção de um sistema de *Estimação da direcção do olhar* aqui também no ISR/DEEC. Para isso será necessário a compra de duas *webcams* e LED de iluminação infravermelha. Numa câmara –a que capta o olho- retira-se o filtro de infravermelhos, que combinada com a iluminação vai evidenciar as características principais do olho. A outra câmara é colocada de forma a captar o mundo que o utilizador observa.

3.1 - Trabalho Futuro

Nos últimos tempos tem-se verificado que há cada vez mais a investigação e a procura sobre a utilização dos olhos e do olhar de um utilizador como um interface para uso de dispositivos portáteis. A utilização de sistemas de *Seguimento do olhar* tem sido aplicada com sucesso em pacientes com mobilidade limitada (de que são exemplo alguns doentes de paralisia cerebral) tanto em aplicações simples de interface, como também em sistemas de comunicação assistida. Na saúde, tem-se usado muitas vezes sistemas de *Seguimento do olhar* na investigação, de forma a tentar estabelecer relação entre movimentos e padrões de olhar com doenças, quer físicas quer psíquicas. No marketing, são cada vez mais as marcas que confiam em estudos efectuados com recurso a tecnologias de *Seguimento do olhar*, desde o design de produtos, à apresentação deles em prateleiras de supermercados. Nos automóveis verifica-se a introdução crescente de sistemas de ajuda ao condutor, e sistemas de segurança passiva, em que os sistemas de *Seguimento do olhar* são usados para verificar o nível de atenção na condução, ou por outro lado o cansaço do condutor. Por tudo isso, pode-se afirmar que os sistemas de *Seguimento do olhar* vão ter alguma evolução, desde a melhoria da sua precisão, à diminuição do seu tamanho, e que a sua incorporação em dispositivos de uso quotidiano vai ser crescente.

Com a evolução notória dos computadores portáteis, estima-se que dentro de pouco tempo estes tragam já incorporada uma webcam de alta resolução. Talvez passe também a ser *standart* a incorporação de iluminação infravermelha, e nesse caso a realização de um sistema de *Seguimento do olhar* num qualquer computador portátil é bastante provável. A este nível com recurso de uma câmara de alta resolução e a iluminação infravermelha, o erro de estimação do gaze do utilizador deverá ser de aproximadamente 0.5 grau (padrão na maioria dos sistemas de *Seguimento do olhar* remotos comerciais atuais). Pode-se prever a utilização do *Seguimento do olhar* em várias aplicações, em que a substituição do rato por um sistema de *Seguimento do olhar* pode ser uma realidade.

A utilização da estimação do gaze pode também ser usada na interação homem-máquina, na área da robótica, de forma a que um robô consiga determinar o nível de atenção e de compromisso de um utilizador em conversação. Durante o período de investigação e elaboração desta tese, surgiu a oportunidade de publicar um artigo para uma conferência (IEEE RO-MAN 2012, com o título “Context-based understanding of interaction intentions”), em que o olhar e o apontar servem de inputs para um robô de forma a que ele entenda intenções do utilizador. Num contexto em que estão vários indivíduos numa sala e um robô, um sistema de estimação da direção do olhar no robô permite a este identificar quando é que um utilizador está a falar para

ele, pela direção do gaze do utilizador. Por outro lado, se o robô comunica com o utilizador e este não encara o robô, podemos extrapolar que o utilizador está distraído com outra coisa e que por isso não está concentrado naquilo que o robô está a comunicar. Pode assim o robô ter a oportunidade de modificar o seu modo de operar. O robô pode assim optar por tentar chamar a atenção do utilizador, ou até mesmo tentar perceber para onde este está a olhar. No contexto de um robô guia num museu, permite ao robô fazer uma visita guiada interativa, falando e explicando as coisas que mais interesse provocam no utilizador. Mas conhecer o gaze do utilizador também pode ser utilizado para “ensinar” ao robô como este se há de comportar, e como ele deverá movimentar os seus interfaces, de forma a que o seu comportamento se torne mais semelhante ao humano. Por exemplo, quando estamos a descrever um objecto numa sala a alguém, o nosso olhar passa do objecto para a pessoa, de forma a captarmos características do objecto, transmiti-las à pessoa e confirmar que ela nos entende (ou segue). Atribuir estes comportamentos a um robô iria aumentar o conforto da comunicação entre uma pessoa e um robô. Um exemplo de um destes robôs é que se apresenta na Figura 37. Este robô tem vindo a ser desenvolvido no ISR para ser utilizado como robô guia.



Figura 37 – Hilário.

O artigo foi aceite para apresentação em conferência, demonstrando assim a pertinência e validade da suposição da utilização do olhar e do gaze para a determinação da atenção de um utilizador.

Referências

- Viola, M., & Jones, P. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. *Computer Vision and Pattern Recognition (CVPR2001)*. Kauai, HI, USA: IEEE Computer Society.
- Bay, H., Tuytelaars, T., & Van Goo, L. (2006). SURF: Speeded Up Robust Features. *European Conference on Computer Vision (ECCV2006)*. Graz, Austria.
- DeMenthon, D., & Davis, L. S. (1995). Model-Based Object Pose in 25 Lines of Code. *International Journal of Computer Vision*, 15, 123-141.
- Dogson, N. A. (2004). Variation and extrema of human interpupillary distance. *SPIE The International Society for Optical Engineering*, (pp. 36-46).
- Galante, A., & Menezes, P. (2012). A gaze-based interaction system for people with cerebral palsy. *CENTERIS 2012*. Vilamoura.
- Kaehler, A., & Bradski, G. (2008). *Learning OpenCV*. O'Rilley.
- Lambooi, M., IJsselsteijn, W., & Heynderick, I. (2007). Visual discomfort in stereoscopic displays: a review. *SPIE Electronic Imaging, Stereoscopic Displays and Virtual Reality*, 6490.
- Menezes, P., Barreto, J., & Dias, J. (2004). FACE TRACKING BASED ON HAAR-LIKE FEATURES AND EIGENFACES. *5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles*. Instituto Superior Técnico, Lisboa, P.
- Quinn, N. (1995). Parkinsonism-recognition and differential diagnosis. *EDUCATION & DEBATE*. London.
- Press, S. (1998). *Magill's Medical Guide*. Salem Press.
- Rosenhahn, B. *Foundations about 2D-3D Pose Estimation*.