



Universidade do Minho
Escola de Engenharia

Paulo Miguel Franco Correia de Brito

**Acessibilidade - Emulação do rato através de
movimentos faciais e voz -Mobility Camera**

Tese de Mestrado em Informática

Trabalho efectuado sob a orientação de
Professor Doutor António Ramires Fernandes
Professor Doutor Manuel João Ferreira

Outubro 2009

DECLARAÇÃO

Nome: Paulo Miguel Franco Correia de Brito

Endereço Electrónico: paulo.brito@alumni.uminho.pt **Telefone:** +351 939 302 239

Bilhete de Identidade: 10099297

Título da Dissertação de Mestrado: Acessibilidade - Emulação do rato através de movimentos faciais e voz -Mobility Camera

Orientadores:

Professor Doutor António Ramires Fernandes

Professor Doutor Manuel João Ferreira

Data de Conclusão: Outubro 2009

Designação do Mestrado: Mestrado em Informática

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA TESE, APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Universidade do Minho, ____ / ____ / _____

Assinatura: _____

Agradecimentos

Quero aqui expressar os meus agradecimentos a um conjunto de pessoas que me ajudaram durante a realização da minha dissertação.

Aos professores Dr. António Ramires Fernandes e Dr. Manuel João Ferreira, supervisores da minha dissertação da Universidade do Minho, pela preocupação e disponibilidade demonstrados e acima de tudo, pelo seu carácter e competência.

E a toda a minha família e amigos especialmente à minha namorada por todo o apoio e compreensão ao longo destes últimos meses de trabalho árduo.

Paulo Brito

Resumo

O acesso ao computador é tradicionalmente feito através de rato e teclado. Este tipo de acesso assume que o utilizador tem capacidade para operar estes dois dispositivos, o que se aplica no caso geral. No entanto, quer seja temporário ou permanente, é certo que existem utilizadores que não conseguem operar o rato e/ou teclado. Em relação ao teclado, se este for o único problema, existem soluções satisfatórias como o teclado virtual que permite escrever texto recorrendo somente ao rato. O rato representa sem dúvida um problema maior já que normalmente um utilizador incapaz de utilizar o rato também estará incapacitado para utilizar o teclado. Nestes casos a funcionalidade do rato é emulada através de um sistema alternativo que não recorra às mãos e/ou braços, membros recorrentemente utilizados para manipulação deste dispositivo. Entre potenciais exemplos de utilizadores de um sistema com estas características encontram-se pessoas que não possuem membros superiores, ou que não possuem controlo sobre estes, ou utilizadores temporariamente incapacitados no caso de acidentes ou lesões nos braços/mãos. Pretende-se neste trabalho estudar alternativas para a utilização do rato de forma a proporcionar o acesso ao computador a essas pessoas. O estudo incidirá sobre métodos de emulação do rato, através de movimentos faciais para controlar o movimento do rato e reconhecimento de voz para emular a funcionalidade dos botões do rato. Pretende-se que o sistema recorra somente a hardware encontrado em sistemas comuns, ou seja, um PC, com uma webcam e um microfone. Para tal é necessário recorrer a técnicas de tracking facial, recorrendo a uma stream de vídeo, capturada por uma webcam de forma a determinar o movimento da face. Estas técnicas recorrem ao processamento de imagem para determinar a posição de alguns elementos da face, nomeadamente o nariz e os olhos. É através da análise da variação das posições destes elementos que será determinado o percurso que o cursor terá de efectuar no ecrã.

Palavras-chave: tracking facial, interacção homem - computador, detecção de objectos, reconhecimento de padrões, processamento de imagem, acessibilidade, OpenCV.

Abstract

Access to the computer is traditionally done through mouse and keyboard. This type of access assumes that the user is able to operate these devices, which is true for the common user. However, whether temporary or permanent, it is also true that there are users who cannot operate the mouse and / or keyboard. Regarding the keyboard there are satisfactory solutions like the virtual keyboard that allows writing text using only the mouse. The mouse is certainly a major problem because typically a user unable to use the mouse will be unable to use the keyboard.

In these cases the functionality of the mouse has to be emulated by an alternative system that does not use the hands and / or arms, members recurrently used to handling this device. Among examples of potential users of a system with these characteristics are people who lost their arms or hands, or do not have control over them, or temporarily disabled users in case of accidents or injuries in the arms / hands. It is the goal of this work to study and implement alternatives to the use of the mouse. The study will focus on methods for the mouse emulation through facial movements to control the mouse movement, and voice recognition to emulate the functionality of the mouse buttons. It is intended that the system requires only hardware found in common systems, like a standard PC with a webcam and a microphone. To achieve this goal we use the techniques of facial tracking, on video streams, captured by a webcam in order to determine the movement of the face. Image processing algorithms are then used to determine the position of some elements of the face, particularly the nose and eyes. It is by analyzing the displacements of these elements from frame to frame, that we will determine the cursor's movement on the screen.

Key-words: face tracking, human-computer interaction, object detection, pattern recognition, image processing, accessibility, OpenCV.

Índice Geral

| | | |
|-------|---|----|
| 1 | Introdução | 1 |
| 1.1. | Contextualização | 1 |
| 1.2. | Motivação..... | 1 |
| 1.3. | Objectivos..... | 2 |
| 1.4. | Metodologia de Investigação | 3 |
| 1.5. | Estrutura do Relatório..... | 4 |
| 2 | Revisão Bibliográfica | 5 |
| 2.1. | Visão por Computador e Interacção Homem-Computador..... | 5 |
| 2.2. | Tracking Facial | 8 |
| 2.3. | Metodologias de Tracking Facial | 10 |
| 2.3.1 | Sistema de tracking baseado num modelo de cor estocástico com templates ajustáveis..... | 11 |
| 2.3.2 | Algoritmo de detecção robusta da cor da pele | 13 |
| 2.3.3 | Detecção facial recorrendo aos classificadores de Haar | 14 |
| 2.3.4 | Optical Flow..... | 20 |
| 2.4. | Soluções actuais..... | 21 |
| 2.4.1 | Camera Mouse 2009..... | 22 |
| 2.4.2 | uMouse | 23 |
| 2.4.3 | Sistemas de Acessibilidade com hardware específico..... | 24 |

| | | |
|---------|---|----|
| 2.4.3.1 | SmartNav | 25 |
| 2.4.3.2 | Tracker Pro..... | 26 |
| 3 | Descrição, Análise e Caracterização do Sistema proposto..... | 28 |
| 3.1. | Sistema de visão | 30 |
| 3.2. | Algoritmos | 32 |
| 3.2.1 | Tracking de pontos com Haar Cascade e mapeamento da área..... | 34 |
| 3.2.2 | Haar Cascade com interpolação de pontos e mapeamento da área | 37 |
| 3.2.3 | Haar Cascade com interpolação de pontos e posicionamento relativo ... | 39 |
| 3.2.4 | Detecção da cor de pele com Haar Cascade e posicionamento relativo . | 40 |
| 3.3. | Emulação dos botões do rato, Drag, Modo de Selecção e Escrita | 43 |
| 3.4. | Análise de desempenho e parâmetros | 45 |
| 3.5. | Especificação do Hardware e Software | 46 |
| 3.6. | Análise de outras soluções | 48 |
| 4 | Resultados experimentais | 50 |
| 4.1. | Pessoas sem limitações físicas..... | 50 |
| 4.2. | Simulação de limitações físicas | 54 |
| 4.3. | Análise dos resultados | 55 |
| 5 | Conclusões e trabalho futuro | 57 |
| 5.1. | Conclusões..... | 57 |
| 5.2. | Sugestões para trabalhos futuros..... | 59 |
| 6 | Anexos | 64 |

| | |
|---|----|
| 6.1. Templates dos questionários | 64 |
| 6.2. Imagens do percurso dos diferentes algoritmos..... | 66 |

Índice de Figuras

| | |
|---|----|
| Figura 1 - IHC é uma área científica multidisciplinar (adaptado de Deepamehta) | 7 |
| Figura 2 - Utilização da Gaussian Mixture Model (GMM) para a distribuição de cor da região de fundo e a região do rosto (modelo estatístico de aglomerados) | 12 |
| Figura 3. – Região facial detectada, indicado pela zona vermelha e o fundo detectado pelo laranja. Imagem adaptada de [8]. | 12 |
| Figura 4 – Algoritmo baseado na união cor de pele | 14 |
| Figura 5 - Exemplo de rectângulos de características. 2 Rectângulos de características são exibidos em A e B, 3 rectângulos de características são exibidos em C e 4 rectângulos em D | 16 |
| Figura 6 - Características seleccionadas pelo AdaBoost | 18 |
| Figura 7 - Resultado do algoritmo usando os classificadores de Haar | 19 |
| Figura 8 - Vector de movimento | 20 |
| Figura 9 - Movimentos de ZOOM-OUT, ZOOM-IN, DIREITA | 21 |
| Figura 10 - Image 1; Image 2 ; Hierarchical Lucas-Kanade – horizontal motion | 21 |
| Figura 11 - Utilização do Camera Mouse 2009 com utilizadores com restrições físicas, software sem custos | 22 |
| Figura 12 - Aplicação uMouse, software sem custos | 24 |
| Figura 13 - Dispositivo utilizado no SmartNav para o tracking | 25 |
| Figura 14 - Dispositivo utilizado no Tracker Pro para o tracking..... | 26 |
| Figura 15 - Material reflector utilizado no Tracker Pro | 26 |

| | |
|---|----|
| Figura 16 – Mobility Câmara com tracking facial recorrendo ao Classificador de Haar facial e identificação da cor da pele. | 28 |
| Figura 17 - Sistema de redundância | 29 |
| Figura 18 - Fluxo funcional do sistema | 30 |
| Figura 19 - Tracking de pontos com Haar Cascade e mapeamento de área | 31 |
| Figura 20 - Configuração de utilização..... | 31 |
| Figura 21 – Esquema das técnicas e algoritmos utilizados no algoritmo proposto | 32 |
| Figura 22 – Encadeamento dos algoritmos..... | 33 |
| Figura 23 - Área de acção do rato, o ponto central da face, a área da face | 35 |
| Figura 24 - Tracking 2 pontos dentro da área de acção do rato | 36 |
| Figura 25 – Detecção e tracking de faces..... | 37 |
| Figura 26 – Situação de tracking em que o rato está imóvel | 39 |
| Figura 27 – Tracking da cor da pele | 40 |
| Figura 28 - Algoritmo Camsift..... | 41 |
| Figura 29 - Sistemas de emulação dos botões do rato | 44 |
| Figura 30 - Configurações da aplicação..... | 46 |
| Figura 31 - Utilização do sistema | 47 |
| Figura 32 - SURF features | 48 |
| Figura 33 – Resultado do template matching : Procura B em A = C..... | 49 |
| Figura 34 – Algumas sessões de testes efectuadas | 50 |
| Figura 35 – Informação adicional da aplicação | 60 |

Índice de Tabelas

| | |
|---|----|
| Tabela 1 - Características dos métodos de tracking facial analisados..... | 11 |
| Tabela 2 - Referência de tempo de execução das tarefas propostas..... | 51 |
| Tabela 3 – Pontuação média resultante dos questionários das entrevistas | 54 |

Índice de Gráficos

| | |
|---|----|
| Gráfico 1 – Comparação de tempos entre tarefas e métodos utilizados | 52 |
| Gráfico 2 – Análise dos resultados mediante diferentes métodos estatísticos | 54 |

1 Introdução

1.1. Contextualização

Um estudo [22] realizado pela Forrester Research (2003) para a Microsoft Corporation apresenta estatísticas sobre a necessidade e a importância da acessibilidade da tecnologia. Estima-se que cerca de 17% (22,6 milhões) de utilizadores de computadores que sofrem de graves deficiências motoras são muito susceptíveis de beneficiar da tecnologia orientada para a acessibilidade. Também é referido que a necessidade de dispositivos com acessibilidade especial pode crescer devido ao aumento do computador nos utilizadores acima dos 65 anos de idade e do aumento da idade média de utilizadores de computador.

As estimativas sobre o número de utilizadores de computadores com deficiências graves sugerem que as interfaces inteligentes e os ambientes inteligentes são cruciais para uma fracção considerável da nossa população. Este trabalho apresenta uma visão geral do software de acessibilidade que foi especificamente concebido para pessoas com graves incapacidades de movimento. Para mais informação relativamente ao software de acessibilidade recomenda-se a leitura da seguinte referência [1].

1.2. Motivação

O acesso ao computador é tradicionalmente feito através de rato e teclado. Este tipo de acesso assume que o utilizador tem capacidade para operar estes dois dispositivos, o que se aplica no caso geral. No entanto, quer seja temporário ou permanente, é certo que existem utilizadores que não conseguem operar o rato e/ou teclado.

Em relação ao teclado, se este for o único problema, existem soluções satisfatórias como o teclado virtual que permite escrever texto recorrendo somente ao rato. O rato representa sem dúvida um problema maior já que normalmente um utilizador incapaz de utilizar o rato também estará incapacitado para utilizar o teclado.

Nestes casos a funcionalidade do rato é emulada através de um sistema alternativo que não recorra às mãos e/ou braços, membros recorrentemente utilizados para manipulação deste dispositivo. Entre potenciais exemplos de utilizadores de um sistema com estas características encontram-se pessoas que perderam os membros superiores, ou que não possuem controlo sobre estes, ou utilizadores temporariamente incapacitados no caso de acidentes ou lesões nos braços/mãos.

O recurso ao movimento da face para controlar o movimento do rato torna-se assim uma opção a explorar. Através de comandos vocais é possível emular a funcionalidade dos botões do rato.

1.3. Objectivos

O objectivo principal deste trabalho consiste na implementação de um sistema que possibilite a utilização por parte de utilizadores com deficiências motoras o uso de um PC. Como tal, pretende-se explorar uma solução alternativa ao uso do rato que faça uma emulação completa das suas funcionalidades. Para atingir este objectivo irá recorrer-se ao tracking facial para controlo do movimento do cursor e a comandos vocais para substituir as funcionalidades dos botões do rato. O sistema de tracking facial deverá utilizar características facilmente destacáveis na face do utilizador, como os olhos e nariz para conseguir acompanhar o movimento da face de forma robusta.

A robustez é um dos aspectos fundamentais do sistema. Dadas as limitações dos utilizadores potenciais de um sistema deste tipo, a falta de robustez torna-o inviável para uma utilização regular. A solução de tracking facial terá como objectivo maximizar essa robustez.

Sendo um sistema vocacionado para o aumento da acessibilidade, a sua usabilidade deve ser testada. Para este efeito, testes deverão ser realizados com utilizadores. Estes testes possibilitaram refinar o sistema e avaliar a sua robustez.

1.4. Metodologia de Investigação

A investigação começará por uma pesquisa bibliográfica com o objectivo de avaliar o estado da arte no campo de tracking facial e de sistemas que emulem o rato. Esta investigação será complementada por uma avaliação de APIs para processamento de imagem, e reconhecimento de voz. Estas APIs servirão de base para a implementação do sistema. Baseado no resultado desta investigação será delineado o sistema a implementar.

A implementação de uma primeira versão será o passo seguinte. Tratando-se de um produto no campo da acessibilidade importa realizar testes de usabilidade com um número significativo de utilizadores. Estes testes pretendem por um lado avaliar a robustez das soluções implementadas, assim como recolher opiniões dos utilizadores com vista a potenciais incrementos e/ou alterações à primeira versão.

Deste primeiro teste resultará uma possível lista de alterações à primeira versão, que uma vez implementadas conduziram à versão final. Um teste comparativo permitirá validar as alterações efectuadas, e verificar o seu real impacto no sistema, quer ao nível da robustez, quer ao nível da usabilidade.

Finalmente será efectuado uma análise comparativa com outros sistemas que apresentem funcionalidades similares.

Como tal, a metodologia de investigação que mais parece adequar-se é a metodologia designada na literatura inglesa como Action-Research, ou Pesquisa-Acção. A metodologia Pesquisa-Acção consta de uma família de metodologias de investigação que utilizam a acção (ou alteração) e a pesquisa (ou entendimento) ao mesmo tempo. Inicialmente realiza-se um processo cíclico ou em espiral em que se alterna entre acção e análise crítica e posteriormente ocorre um entendimento dos métodos e dados que estão continuamente a ser aperfeiçoados.

Portanto, é um processo emergente que adquire forma com o aumento da compreensão e é também um processo iterativo que converge para uma melhor percepção do que acontece. Na maioria das vezes também se torna num processo participativo (entre outras razões, a mudança é geralmente mais fácil de conseguir

quando as pessoas afectadas/interessadas pela mudança estão envolvidas) e qualitativo.

O método de Pesquisa-Acção realiza-se através das seguintes etapas:

- Identificação do problema;
- Análise do problema de forma a encontrar maneiras de o solucionar;
- Implementação de propostas de melhoria;
- Avaliação da implementação das propostas de melhoria;
- Modificação das formas de aprendizagem;
- Definição de novos projectos de Pesquisa-Acção.

Assim, com este trabalho pretende-se implementar um sistema que responda às seguintes questões:

- Qual o melhor sistema de tracking ou conjunto de sistemas de tracking?
- Como é que o sistema possibilita ao utilizador um controlo intuitivo?
- Como é que se pode aumentar a fiabilidade e robustez do sistema de reconhecimento e tracking facial?

1.5. Estrutura do Relatório

No capítulo 2 é feita a revisão bibliográfica com a apresentação dos diferentes domínios de ciência envolvidos, bem como algoritmos utilizados e soluções actuais. No capítulo 3 é efectuada a descrição e análise geral do sistema a implementar. No capítulo 4 são apresentados os resultados experimentais da utilização do sistema e é feita uma análise do desempenho. Por fim, no capítulo 5 são retiradas todas as conclusões inerentes à elaboração deste relatório.

2 Revisão Bibliográfica

Nesta revisão bibliográfica realiza-se uma análise ao estado da arte de vários temas, como Visão por Computador, Interação Homem-Computador, Tracking Facial, Metodologias de Tracking Facial e soluções actuais tais como Camera 2009, UMouse, SmartNav e Tracker Pro. Apresentando-se os princípios, ferramentas e mecanismos que permitem a implementação de sistemas de interface inovadores. O estado de arte reflecte o que se tem feito nestes domínios para a solução do problema da utilização de sistemas informáticos por parte de pessoas com deficiências motoras, e a sua aplicação prática em sistemas de emulação do rato através de movimentos faciais.

2.1. Visão por Computador e Interação Homem-Computador

É o domínio das ciências que consiste no estudo e aplicação de métodos que permitem aos computadores “compreender” o conteúdo de uma imagem. O termo “compreender” significa a extracção a partir de uma imagem da informação específica que é útil para um dado fim: seja para a apresentar a um operador humano (por exemplo, se células cancerosas forem detectadas numa imagem de um microscópio), ou para controlar algum processo (por exemplo, um robô industrial ou um veículo autónomo).

Como mecanismo de visão por computador é utilizado o OpenCV (Open Source Computer Vision Library). Originalmente, desenvolvida pela Intel, em 2000, é uma biblioteca multiplataforma, totalmente livre ao uso académico e comercial, para o desenvolvimento de aplicativos na área de Visão Computacional, bastando seguir o modelo de licença da BSD Intel. O OpenCV possui módulos de Processamento de Imagens e Vídeo I/O, Estrutura de dados, Álgebra Linear, GUI (Interface Gráfica do Utilizador) Básica com sistema de janelas independentes, Controlo do rato e teclado, além de mais de 350 algoritmos de Visão Computacional como: Filtros de imagem, calibração de câmara, reconhecimento de objectos, análise estrutural e outros. O seu processamento é em tempo real.

Na área de Interação Homem-Computador (IHC) os investigadores dedicam-se ao estudo da interação entre pessoas e computadores. É uma matéria multidisciplinar que relaciona a ciência da computação, artes, design, ergonomia, psicologia, sociologia, semiótica, linguística, e áreas afins. A interação entre homens e máquinas acontece através da interface do utilizador, a qual é formada por software e hardware. Ela é utilizada, por exemplo, para a manipulação de periféricos de computadores e grandes máquinas como aviões e centrais hidroelétricas.

O desempenho humano no uso de computadores e de sistemas de informação tem sido uma área de pesquisa e desenvolvimento que muito se expandiu nas últimas décadas. Os estudos recorrem a poderosas ferramentas computacionais na análise de dados colectados de acordo com métodos da Psicologia Experimental. Outras contribuições também advêm da Psicologia Educacional, do Design Institucional e Gráfico, dos Factores Humanos ou Ergonomia, e mais recentemente, da Antropologia e da Sociologia. As interfaces para utilizadores têm produzido importantes histórias de sucesso tais como a da Netscape, América Online, Universal Online, ou Yahoo.

A área de IHC começou com Donald Norman, psicólogo cognitivo que trabalhou o conceito de usabilidade. É possível referir 3 períodos durante a história da área de IHC:

- 1º Período virado para factores humanos. Estudo do utilizador como um conjunto de mecanismos de processamento de informação. Foco no indivíduo. Criação de guidelines para desenvolvimento de interfaces, métodos formais e testes sistemáticas baseados em métricas.
- 2º Período voltado para factores humanos. Foco em grupos. Abordagens qualitativas e não tanto quantitativas, desenvolvimento de protótipos e design contextual. Análise global e um entendimento geral dos fenómenos relativos a uma pessoa num dado ambiente.

- 3º Período com a ênfase em aspectos culturais e estéticos. Expansão do cognitivo ao emocional. Factores pragmáticos sociais da experiência. Tecnologias ubíquas, móveis e pequenas. A tecnologia extrapola os limites do contexto de trabalho e passa a fazer parte da cultura, vida e casa das pessoas.

Esta multidisciplinaridade de domínio pode-se visualizar na Figura 1.



Figura 1 - IHC é uma área científica multidisciplinar (adaptado de Deepamehta)

Resumindo, para um utilizador poder interagir com um determinado sistema é necessário que o controle e sobretudo que tenha conhecimento do seu estado. É a interface homem-computador que possibilita essa interacção, e representa-se pela camada que se encontra entre o sistema e utilizador humano. O objectivo da interface é fundamentalmente comunicativo. Esta comunicação só é possível com a existência de uma espécie de diálogo entre o utilizador e a máquina, basicamente orientado na forma como o utilizador pode interagir com a máquina.

2.2. Tracking Facial

As imagens contendo rostos de pessoas são muito utilizadas na interação homem-computador baseada numa visão inteligente. Devido a isto, vários grupos de investigação no processamento de imagens do rosto humano centram os seus esforços no reconhecimento facial, tracking facial, avaliação de posturas e reconhecimento de expressões. Em média, os algoritmos de detecção facial conseguem detectar rostos com precisão, relativamente à sua posição numa imagem. Um levantamento do estado da arte relativamente à detecção facial pode ser encontrada em [2].

Para desenvolver sistemas automáticos que analisam as informações contidas em imagens faciais, algoritmos de detecção facial eficientes e robustos são necessários. O objectivo da detecção facial, tendo em conta uma única imagem, passa por identificar todas as regiões na imagem que potencialmente contêm um rosto, independentemente da sua posição tridimensional, a orientação e condições de iluminação. Este problema é exigente porque os rostos não são rígidos e têm um elevado grau de variabilidade em tamanho, forma, cor e textura.

A detecção e o tracking facial é muitas vezes o primeiro passo em aplicações tais como vídeo vigilância e reconhecimento facial. Localização e tracking de rostos humanos é uma exigência do tracking facial e / ou análise de expressões faciais, embora, muitas vezes, é assumido que um rosto está disponível na imagem em causa.

As pessoas com restrições involuntárias de movimento sofrem frequentemente grandes dificuldades para interagir com o computador. Portanto, dispositivos centrados na tecnologia de acessibilidade têm sido desenvolvidos para diminuir estas limitações. No entanto, a maioria delas exigem dispositivos extra que são dispendiosos e desconfortáveis, sujeitos ao desgaste ou à utilização, como os infravermelhos, materiais reflectores e óculos.

Como referido em [3], a tecnologia de tracking facial tem diversas aplicações na área dos utilizadores de computador com deficiências fisiológicas que impeçam o uso normal de um dispositivo. Isto permite, por exemplo, aos utilizadores que têm

dificuldade em utilizar um dispositivo de interacção normal, tal como o rato, de serem capazes de manipular um cursor no monitor, movendo a cabeça.

Recentemente, devido ao melhoramento do desempenho de processamento do computador pessoal e do uso massivo de webcams, é possível construir sistemas de Interacção Homem-Computador (IHC), combinando os pontos fortes e descobertas de novos métodos de visão por computador. Estes sistemas permitem, com o recurso a uma câmara de vídeo, mapear o movimento da cabeça do utilizador para o cursor no monitor.

No entanto, para obter um tracking em tempo real, muitos problemas têm de ser resolvidos. Um problema importante reside na variedade de características existentes, tais como a cor da pele ou olhos, barba e óculos. Outro problema é o tempo de resposta do sistema. Não importa se um sistema é capaz de reconhecer e fazer o tracking de uma face se não é capaz de o fazer dentro de um tempo tolerável / utilizável. Além disso, o tracking tem que ser preciso e robusto para uso prático. Estes problemas aumentam a dificuldade para o reconhecimento e tracking das características do rosto em tempo real.

O problema do tracking facial através de uma câmara de vídeo faz com que a ênfase do presente trabalho recaia sobre a concepção de um sistema de visão que possibilite uma interface perceptual para o utilizador.

As aplicações que utilizam algoritmos de tracking facial requerem que o sistema de tracking seja rápido, acessível e, mais importante, preciso e robusto. Em particular, a precisão deve ser suficiente para controlar um cursor, enquanto a robustez deve ser alta o suficiente para permitir a um utilizador a comodidade e flexibilidade ao nível de movimentos da cabeça.

As abordagens para sistemas de tracking facial podem ser separadas em duas classes: com base nas imagens e abordagens baseadas nas características [4].

Com base nas imagens são utilizados indícios gerais do rosto, tais como cor da pele, cabeça, geometria e movimento. Estes indícios são robustos ao movimento de rotação, à escala, e não exigem alta qualidade de imagem. Por outro lado, estas abordagens

carecem de precisão e, portanto, não podem ser usadas para controlar o cursor com precisão.

Para obter um tracking preciso e suave é utilizada a abordagem das características [3]. Estas abordagens são baseadas no tracking individual de características faciais. O tracking das características pode ser obtido com precisão ao pixel, o que permite um mapeamento directo e preciso para a posição do cursor. A desvantagem destas abordagens é que estas geralmente requerem câmaras caras e de alta resolução. Além disso, estes não são robustos aos movimentos da cabeça, especialmente à rotação e à escala.

Ainda recentemente, foi demonstrado que a robustez do tracking baseado nas características individuais do rosto pode ser significativamente melhorada se em vez de usar-se as pontas baseadas em características tais como os cantos e bordas de olhos, boca e narina, se usarem características baseadas na curvatura do nariz [5]. Este facto cria uma nova gama de possibilidades interessantes para o tracking facial baseada nas características do rosto.

2.3. Metodologias de Tracking Facial

O tracking facial é um componente essencial para permitir que o computador consiga perceber o utilizador do PC num sistema de Interface Homem-Computador. Consequentemente é necessário analisar as soluções que possibilitem a capacidade de construir um sistema sólido e eficiente de tracking em tempo real e com base apenas numa câmara USB e um sistema desktop standard.

Nesta secção será apresentado o estudo / avaliação de alguns algoritmos de tracking facial e seu desempenho em relação à precisão, robustez, e capacidades de aprendizagem. As principais características e sistemas dos métodos analisados podem ser visualizadas na Tabela 1 - Características dos métodos de tracking facialTabela 1.

| Método | Características |
|--|--|
| 2.4.1 Modelo de cor estocástico com templates ajustáveis | Baseada na característica da detecção da cor da pele Mapeamento de templates/modelos ajustáveis |
| 2.4.2 Detecção robusta da cor da pele | Baseada na característica da detecção da cor da pele em diferentes espaços de cor |
| 2.4.3 Detecção facial recorrendo a classificadores de Haar | Baseada na detecção de características faciais |

Tabela 1 - Características dos métodos de tracking facial analisados

2.3.1 Sistema de tracking baseado num modelo de cor estocástico com templates ajustáveis

Este algoritmo foi proposto por [6] e permite tracking facial em tempo real. Começa com um tracking facial de um único rosto baseado num modelo estatístico para a classificação da cor e da utilização de um modelo no qual se consiga alterar a forma da área. O método descrito em [7] é usado para implementar o tracking facial.

Algoritmo:

Em primeiro lugar, para cada pixel na frame actual, é calculada a probabilidade de pertencer a cada uma das duas classes: a classe do rosto e a classe do background, como se pode ver na Figura 2. É utilizado o Gaussian Mixture Model (modelo estatístico para classificação/ identificação de clusters).

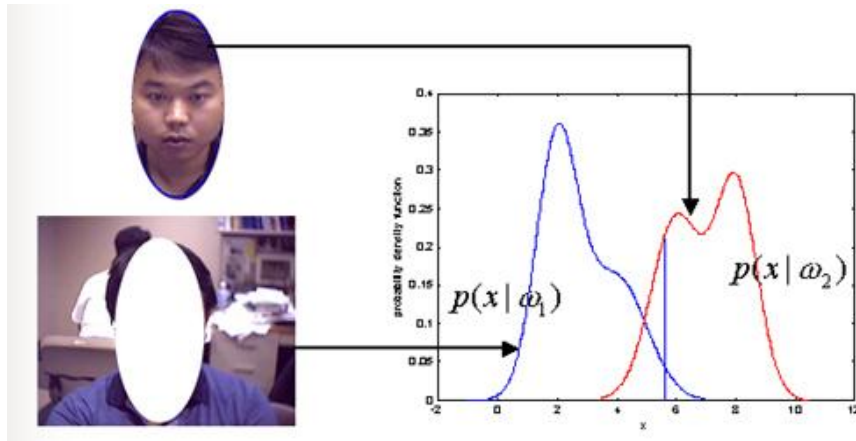


Figura 2 - Utilização da Gaussian Mixture Model (GMM) para a distribuição de cor da região de fundo e a região do rosto (modelo estatístico de aglomerados)

Depois é utilizado um modelo deformável para o grupo de píxeis mais provável a pertencerem à classe rosto. O modelo é deformado de maneira a que a área inclua o máximo número possível de píxeis pertencentes à classe do rosto e o menor número possível da classe do background. A melhor deformação pode ser encontrada usando uma pesquisa logarítmica.

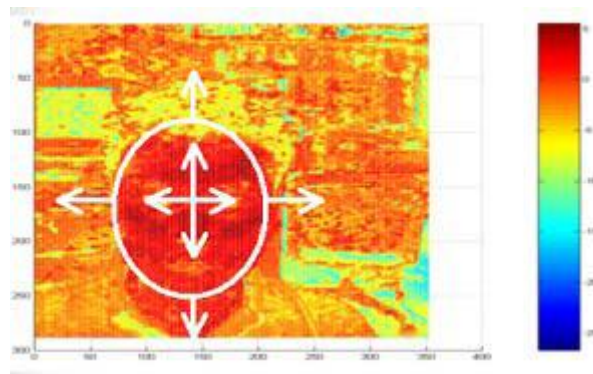


Figura 3. – Região facial detectada, indicado pela zona vermelha e o fundo detectado pelo laranja. Imagem adaptada de [8].

Este algoritmo foi utilizado em sequências de vídeo com diferentes padrões de oclusão e a conclusão é que este funciona bem para a maioria das sequências. Ainda assim, o processo de recuperação da perda de tracking pode ser melhorado, pois após a oclusão existe um desfasamento entre a detecção do rosto e o tracking facial [6].

2.3.2 Algoritmo de detecção robusta da cor da pele

Este algoritmo é baseado na detecção da cor da pele e na análise dos espaços de cores: RGB; YCbCr; e HSI [9]. Neste caso a solução é comparar os algoritmos baseados nestes espaços cor e combiná-los para obter um novo algoritmo de “face tracking” com maior precisão.

O estudo sobre a classificação da cor de pele tem vindo a adquirir cada vez mais atenção nos últimos anos, devido à investigação activa em termos de representação de conteúdos com base na imagem. A forma mais comum de localização facial é a utilização da informação relativa à cor, logo estimar as áreas que contêm a cor da pele é muitas vezes o primeiro passo dessa estratégia. Assim, a classificação da cor de pele tornou-se uma tarefa importante. Grande parte da investigação com base na cor de pele para a localização e detecção facial é baseado nos espaços de cor RGB, YCbCr e HSI.

A detecção facial com base na da cor de pele no espaço de cor é um dos mais simples algoritmos para detecção de píxeis que correspondem à cor de pele. A percepção da cor de pele humana varia em função da orientação relativa à iluminação. A região correspondente aos píxeis da pele pode ser detectada utilizando um histograma normalizado da cor, que pode ser otimizado para mudanças na intensidade de iluminação. O resultado é uma imagem de rosto detectada através da região da pele. Este algoritmo falha quando há mais regiões de pele na imagem tais como as pernas ou braços.

Combinando as regiões detectadas a partir dos três algoritmos baseados nos espaços de cor acima mencionados, todas as regiões da pele presentes na imagem são extraídas. Finalmente, uma caixa delimitadora é desenhada ao redor da região.

Perante uma imagem de uma face cada algoritmo encontra partes da face, podendo ou não ser capaz de encontrar a totalidade da área da pele da face. O resultado final será a união das três áreas detectadas.

Esta afirmação é baseada na ideia básica do diagrama de Venn da Teoria Set [10]. Se afirmamos que o resultado do espaço de cor RGB é a região "A", o resultado do espaço

de cor YCbCr é a região "B" e o resultado do espaço de cor HSI é a região "C", e se estas regiões identificarem individualmente partes de um rosto, o resultado da união dos três será, a área correspondente à imagem da face, como se pode ver na figura 4.

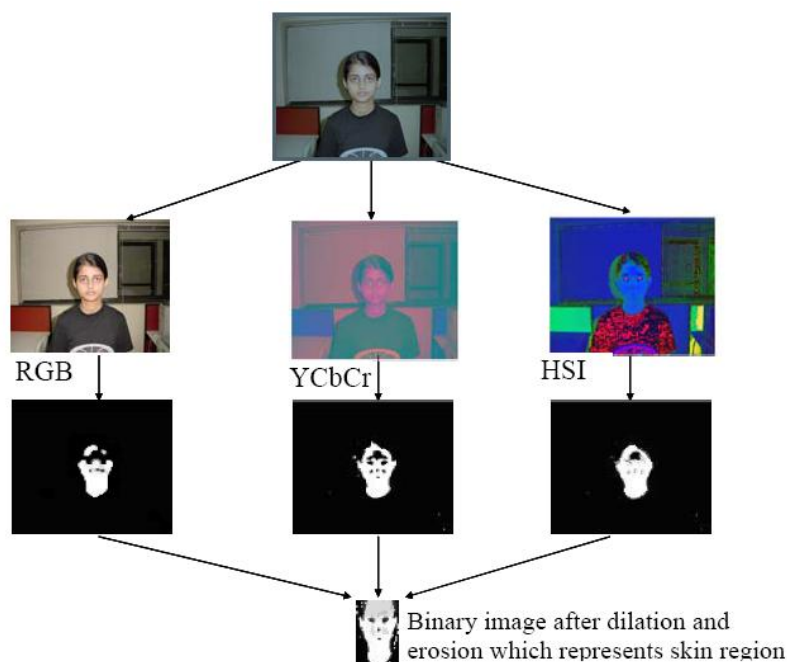


Figura 4 – Algoritmo baseado na união cor de pele

A avaliação desta técnica mostra que é suficiente precisa para localizar uma face humana numa imagem, atingindo valores para a precisão por volta dos 95% na detecção de regiões de pele e de faces.

2.3.3 Detecção facial recorrendo aos classificadores de Haar

Durante o início e meados dos anos 70 Chellappa [23] referiu que a identificação de faces era efectuada utilizando técnicas típicas da área de reconhecimento de padrões, onde as características extraídas eram medidas entre pontos retiradas da face. Como a extracção de características era executada manualmente, o sistema não apresentava alterações quando se efectuavam inclinações, rotações da cabeça, na qualidade e contraste da imagem. Em 1990 foi apresentado o método de análise em componentes principais (Principal Component Analysis PCA) que utiliza a transformada de Karhunen-

Loève [24]. Turk e Pentland [25] aplicaram este método no reconhecimento de faces, introduzindo a técnica denominada eigenfaces. João P. Hespanha [26] aplica o discriminante de Fisher no reconhecimento de faces (fisherfaces) e faz um paralelismo com a técnica de eigenfaces. Nos últimos anos muitas pesquisas concentraram-se nos problemas de segmentação e localização da face numa imagem, assim como na extracção de características da face.

Viola e Jones [11], introduziram um método preciso e eficiente (o uso de um sistema de tracking facial em tempo real) para detectar um rosto numa imagem. Este método pode ser adaptado para detectar com precisão características faciais [12]. No entanto, a área da imagem a ser analisada na procura de uma característica facial deve ser direccionada para a zona com maior probabilidade de conter a característica. Definindo várias áreas de detecção, eliminamos falsos positivos (quantidade de vezes que o algoritmo informou que existia uma face num local onde não existia) e aumentamos a rapidez da detecção devido à redução da área analisada.

Para a detecção de faces utilizamos as características extraídas através da transformada de Haar [11]. As características são definidas como rectângulos de tamanho variável que podem ser aplicadas em qualquer posição da imagem em tons de cinza. A descrição de uma característica de Haar compreende o tipo, o tamanho (i.e., largura e altura) e a posição na imagem.

Estas características, em vez de usar os valores de intensidade do pixel, utilizam a diferença de contraste entre valores de grupos rectangulares de píxeis adjacentes.

As variações do contraste entre grupos de píxeis são usadas para determinar áreas escuras e claras. As “Haar features” podem se adaptar ao tamanho do objecto a detectar pois a sua dimensão pode ser alterada (aumentando ou diminuindo o tamanho do grupo de píxeis em análise).

A extracção de características é baseada numa ideia simples, que não utiliza a intensidade do pixel, o que torna a tarefa de aprendizagem mais fácil, que é um dos motivos para a grande velocidade da técnica. Para obter uma característica é utilizada

a transformada de Haar [13] para segmentar regiões da imagem como mostrado na Figura 5 como descrito em [14] .

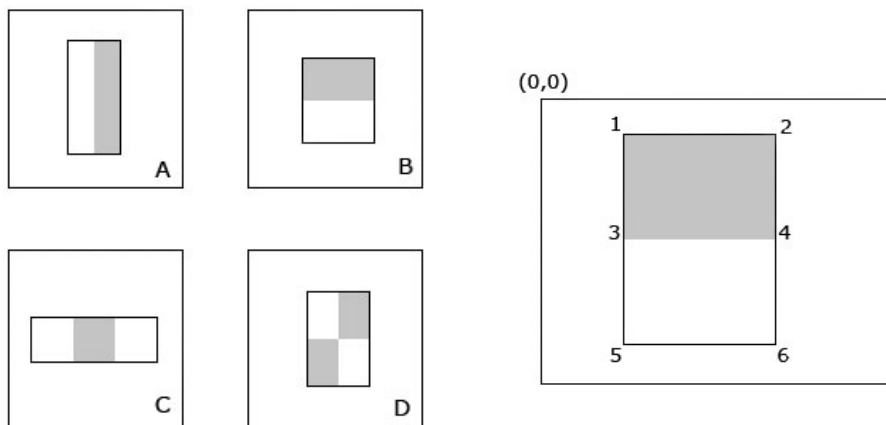


Figura 5 - Exemplo de rectângulos de características. 2 Rectângulos de características são exibidos em A e B, 3 rectângulos de características são exibidos em C e 4 rectângulos em D. Cálculo de uma característica.

A característica nada mais é do que a soma de todos os píxeis da região branca subtraído da soma dos píxeis da região sombreada.

Para calcular as características de Haar rapidamente, com pouco processamento, utilizamos a Imagem Integral [11]. Nesta representação o valor de cada pixel (e.g. Figura 5) corresponde a soma de todos os seus antecessores (no eixo x e y). Com isso é possível calcular a soma dos píxeis de qualquer região da imagem original com apenas quatro acessos à imagem.

$$\text{Área Branca} = V6 + V3 - V4 - V5$$

$$\text{Área Sombreada} = V4 + V1 - V2 - V3$$

$$\text{Característica} = \text{Área Branca de A} - \text{Área Sombreada A}$$

Ao aplicarmos determinadas características de Haar em objectos semelhantes, como faces, os valores retornados tendem a ser próximos. Com isso é possível definir um limiar que identifica se o valor de uma característica numa imagem pode corresponder a uma face ou não.

Uma característica de Haar individualmente é muito simples. Ao aplicarmos um limiar no seu valor ela tem baixa expressividade, com uma taxa muito alta de falsos positivos. Então, a solução passa por combinar várias para formar classificadores fortes.

Um algoritmo de aprendizagem, permite ao computador aperfeiçoar seu desempenho numa tarefa (Adaptive Boosting [15]) de reconhecimento. Deste modo é efectuada uma fase de treino que define os conjuntos de características de Haar e limiares que melhor distinguem imagens de faces das não faces.

Esses conjuntos são chamados de classificadores. Para formar cada classificador, todas as características de Haar possíveis são testadas (todos os tipos, tamanhos e posições) com seus melhores limiares nas imagens de uma base de treino. E as características que apresentarem o menor erro de classificação entre face e não face são seleccionadas.

Na fase de detecção do algoritmo, uma janela de 24x24 percorre a imagem procurando as faces. Nesta janela são aplicados os classificadores na sequência em que foram treinados, ou seja, dos mais simples para os mais complexos. Com isso é possível eliminar grande parte dos candidatos a face rapidamente e com pouco processamento.

Os candidatos rejeitados por um classificador são directamente considerados não face.

Os autores foram capazes de atingir uma taxa de precisão 95% [11] [16] para a detecção de um rosto humano usando apenas algumas centenas de características simples. Com um computador a 2 GHz, um conjunto de classificadores de Haar pode detectar rostos humanos, a uma taxa de pelo menos cinco frames por segundo.

O resultado da utilização do algoritmo para a selecção de um conjunto reduzido de características pode ser demonstrado na Figura 6. É possível perceber que as características seleccionadas foram a de dois rectângulos na altura dos olhos com o

nariz e a de três rectângulos entre os olhos e a testa. No primeiro caso o algoritmo percebeu a diferença de luminosidade entre o nariz e a área por baixo dos olhos relativamente à região dos olhos, no segundo caso percebeu a semelhança de luminosidade entre os olhos e a diferença para a região superior do nariz.

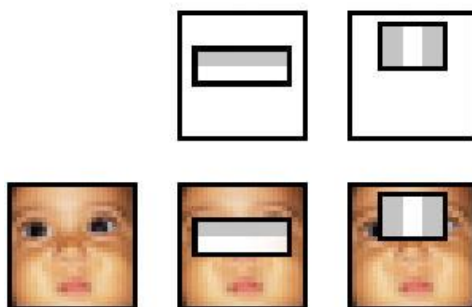


Figura 6 - Características seleccionadas pelo AdaBoost

O segredo para ter uma detecção rápida é ter uma sequência de classificadores. O primeiro classificador descarta a maioria das sub janelas da imagem como falsos positivos com pouquíssimo processamento pois realiza o cálculo de apenas 2 características (Figura 6).

A detecção de elementos da face, tais como a boca, olhos e nariz, exige que um conjunto de classificadores de Haar em cascatas seja treinado em primeiro lugar. De forma a treinar os classificadores, os algoritmos Adaptive Boosting e as características de Haar devem ser implementados.

Classificadores com mais características possuem uma taxa de falsos positivos menor e uma alta taxa de detecção, em contrapartida necessitam de mais processamento. A técnica ideal seria para cada classificador, seleccionar um limiar de separação entre as amostras positivo e negativamente classificadas de forma que minimize o número de características que serão utilizadas. Na prática uma técnica simples é utilizada para o treino: em cada etapa do treino é seleccionado um limite mínimo de reconhecimento e um limiar máximo de falsos positivos, o treino consiste em acrescentar rectângulos de características ate que esses parâmetros sejam obtidos.

Para treinar os classificadores são necessários dois conjuntos de imagens. Um conjunto respeita imagens ou cenas que não contém o objecto, neste caso a característica facial, que vai ser detectadas. Este conjunto de imagens é referido como as imagens negativas. O outro conjunto de imagens, as imagens positivas, contém uma ou mais instâncias do objecto. A localização dos objectos dentro do conjunto das imagens positivas é especificado por: nome imagem, o pixel superior esquerdo e da altura e largura do objecto.

O último classificador (e.g. Figura 7) utiliza apenas algumas centenas de características Haar [11] [12] devido ao agrupamento em cascata dos classificadores. Este tipo de agrupamento permite que apenas as sub-imagens com a maior probabilidade sejam analisadas por todas as características de Haar que distinguem um objecto. Também permite a variação da precisão de um classificador, isto é, permite aumentar tanto o parâmetro de falsos positivos como o parâmetro de respostas positivas, diminuindo o número de etapas. No entanto, atinge uma percentagem elevada com uma taxa relativamente baixa taxa de falsos positivos.

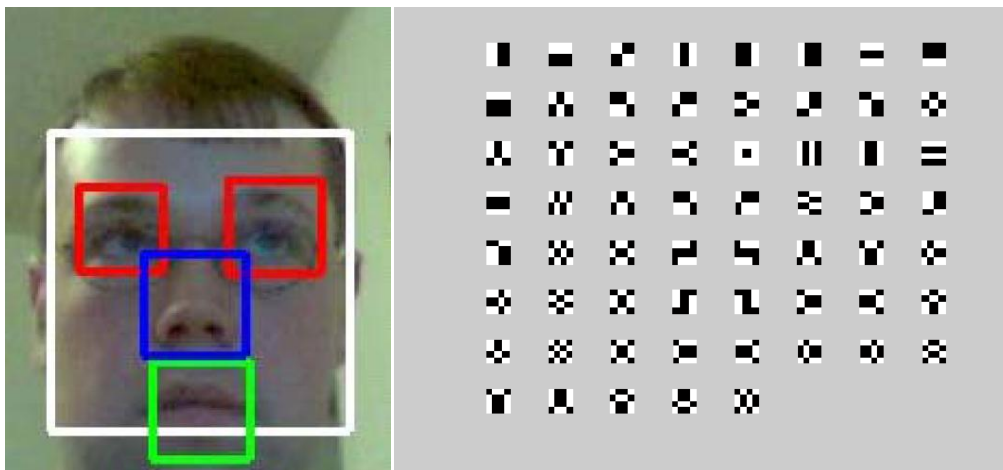


Figura 7 - Resultado do algoritmo usando os classificadores de Haar

2.3.4 Optical Flow

Há 20 anos atrás Horn e Schunck [27] publicaram um artigo relativamente ao cálculo do fluxo óptico. Este trabalho gerou um esforço substancial de investigação, produzindo uma abundância de técnicas para o cálculo do fluxo óptico.

A análise de movimento em vídeo permite mostrar características úteis, tais como a velocidade dos objectos ou do observador, a forma dos objectos e os seus percursos. Apenas com o estudo da intensidade dos píxeis ao longo de uma sequência de vídeo é possível detectar os objectos na sequência e fazer o seu reconhecimento.

Através de uma sequência de vídeo é possível determinar os vectores de movimento (basta apenas duas imagens para esse cálculo), um objecto com centro de massa (T_x, T_y) desloca o seu centro de massa numa outra imagem para a posição (H_x, H_y) . (T_x, T_y) é denominada a cauda do vector de movimento e (H_x, H_y) a cabeça do vector, como pode ser visto na Figura 8.

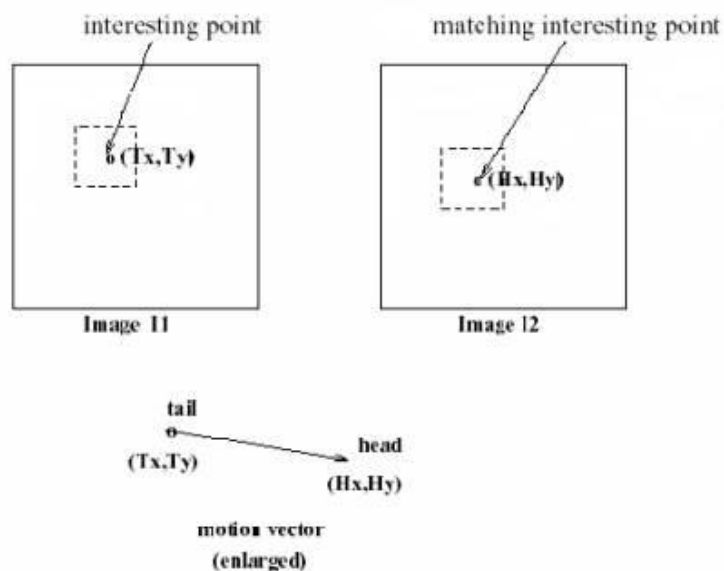


Figura 8 - Vector de movimento

Uns conjuntos de vectores 2D representando o movimento aparente (na imagem) de objectos são chamados de campo de movimento. Através do campo de movimento podemos ter informações que nos permitem distinguir e classificar movimentos por

exemplo o movimento de ZOOM-OUT, ZOOM-IN, DIREITA, mostrados pelos campos de movimento da Figura 9:

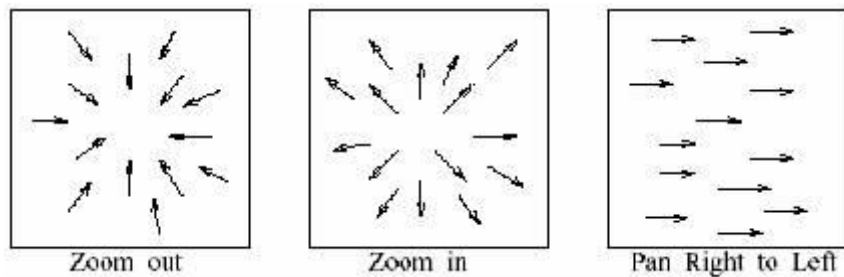


Figura 9 - Movimentos de ZOOM-OUT, ZOOM-IN, DIREITA

O fluxo óptico é o campo de movimento calculado sob a hipótese de que as intensidades em redor de pontos em correspondência nas duas imagens se mantêm constante. Existem vários algoritmos para o cálculo de fluxo óptico [28] [31], o por nós utilizado encontra-se no artigo [30].

O cálculo do fluxo óptico permite-nos dado um conjunto de pontos numa determinada imagem, encontrar esses mesmos pontos numa outra imagem (e.g. Figura 10). Uma questão fundamental que também é abordada na resolução deste problema é o cálculo dos pontos a monitorizar em cada imagem [32] [33].



Figura 10 - Image 1; Image 2 ; Hierarchical Lucas-Kanade – horizontal motion

2.4. Soluções actuais

Para além da implementação do conceito inerente à acessibilidade para pessoas com impedimentos motores, através da utilização de mecanismos de tracking, há necessidade de analisar as várias técnicas utilizadas em soluções existentes para aumentar a simplicidade e o desempenho da aplicação. Muitas destas soluções

baseiam-se em face tracking sem o recurso a hardware adicional, enquanto outras foram desenvolvidas com o apoio de mecanismos suplementares para realizar a tracking e posteriormente a emulação do rato.

Algumas das técnicas/soluções existentes actualmente são:

2.4.1 Camera Mouse 2009

O Camera Mouse 2009 é um sistema que usa uma câmara para monitorizar visualmente a ponta do nariz ou a ponta de um dedo, ou qualquer outra característica seleccionada do corpo movendo o cursor do rato em conformidade. As pessoas sem deficiência rapidamente aprendem a utilizar o sistema para interagir com o computador. Algumas pessoas com paralisia cerebral grave experimentaram o sistema com algum sucesso inicial. O objectivo deste sistema é o de proporcionar acesso ao computador a pessoas que não têm um controlo das mãos, mas que podem movimentar a sua cabeça. Pessoas com Paralisia Cerebral, atrofia muscular espinhal, ALS, esclerose múltipla, vários distúrbios neurológicos, tetraplégicos podem interagir normalmente com o computador, através da utilização de aplicações de sistemas de visão computacional como podemos observar na Figura 11.



Figura 11 - Utilização do Camera Mouse 2009 com utilizadores com restrições físicas, software sem custos

O sistema de visão do Camera Mouse 2009 funciona da seguinte maneira:

- o computador recebe 30 frames por segundo a partir de uma câmara montada acima ou em baixo do monitor. A imagem utilizada tem a dimensão de 320 por 240 píxeis.
- quando o rosto da pessoa está centrado na imagem, o utilizador utiliza o rato para clicar numa característica facial, por exemplo a ponta do nariz.
- a partir deste ponto uma sub imagem de 15 por 15 píxeis centrada à volta do ponto onde o utilizador clicou é comparada com as imagens futuras para fazer o tracking facial do cursor do rato

Esta aplicação não permite mover o rato com precisão suficiente para que possa ser utilizada no contexto de aplicações que contenham pequenos botões na sua interface. O Camera Mouse 2009 funciona melhor com os programas que exigem apenas um rato, um clique esquerdo e que não tem zonas pequenas de acção. Não existe a emulação do botão do lado direito do rato e o clique do botão do lado esquerdo faz-se com a paragem do cursor num ponto durante alguns segundos. Outro aspecto menos positivo é o facto da aplicação não ter uma política de recuperação em caso de falha do algoritmo de tracking e não suportar movimentos bruscos.

2.4.2 uMouse

O uMouse em conjunto com uma Web câmara pretende substituir o rato, o teclado convencional e pretende ser uma interface que possibilita a utilização do PC sem mãos. Possibilita o controlo do cursor e cliques apenas movendo a cabeça, mão ou dedo. O software fornece em tempo real, tracking visual e traduz isso em comandos para o rato. Além disso, o software é capaz de interpretar gestos, bem como mapear estas acções para eventos do teclado.

A nível de emulação dos cliques do rato o uMouse permite apenas executar o botão do lado esquerdo, não tendo opção para o botão do lado direito. O método do clique baseia-se em dois tipos: paragem prolongada – onde o clique é executado passado algum tempo de o cursor estar parado num ponto; e paragem com gestos – onde só é

possível realizar o clique através de um gesto e depois de estar com o cursor parado num ponto durante um determinado tempo.

Porém a complexidade (e.g. Figura 12) do sistema é bastante elevada e não foi possível obter e utilizar a aplicação devido à correcção de bugs que já dura desde o início desta tese sugerindo que as capacidades reais da aplicação não correspondem efectivamente à descrição da aplicação mencionada no primeiro paragrafo desta secção.

O facto da emulação do clique do rato utilizar um mecanismo de gestos para efectuar um clique pode restringir a sua utilização em pessoas com impedimentos físicos.

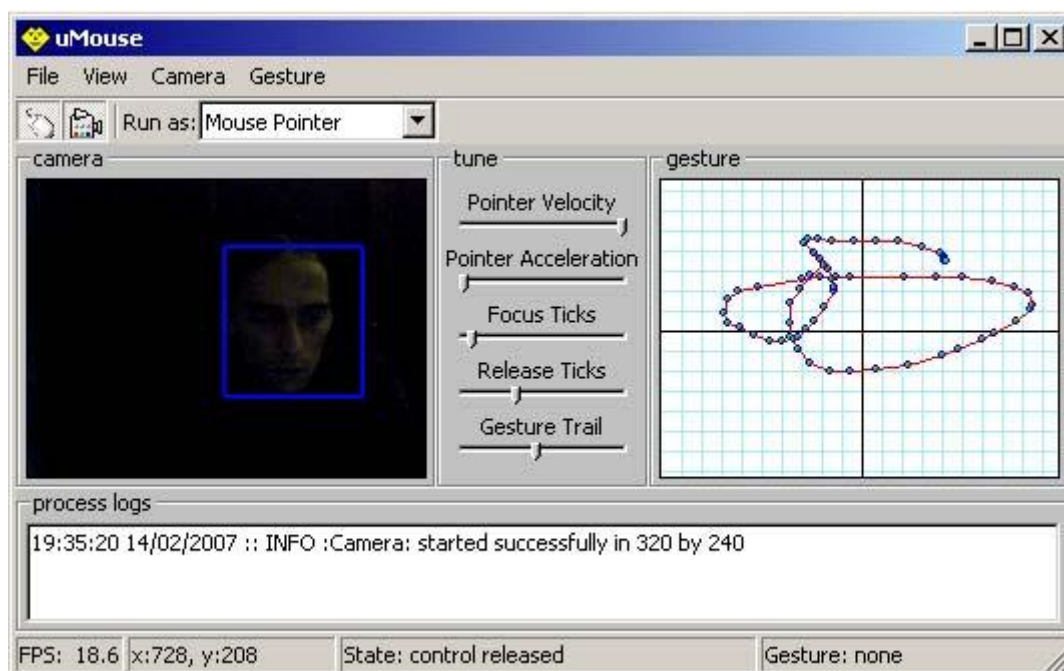


Figura 12 - Aplicação uMouse, software sem custos

2.4.3 Sistemas de Acessibilidade com hardware específico

Estes sistemas de acessibilidade são por norma mais eficazes mas baseiam-se na utilização de equipamentos específicos tais como câmaras de infravermelhos (IR) e materiais reflectores ou sinalizadores de IR. Embora esta tese se dedique a sistemas que não necessitem de hardware específico, apresenta-se aqui uma pequena lista de soluções com hardware dedicado, por uma questão de abrangência. Esta abordagem é

válida, mas limita o acesso a este tipo de soluções pelos mais necessitados pois torna-os mais caros.

Algumas das aplicações comerciais existentes actualmente são:

2.4.3.1 SmartNav

O SmartNav usa uma câmara de infravermelho (IR) para fazer o tracking dos movimentos da cabeça do utilizador (e.g. Figura 13). O sistema funciona da seguinte maneira: a luz infravermelha é emitida a partir dos LEDs e é reflectida de volta para o dispositivo por um material reflector (boné especial com reflectores, um reflector circular - ponto colocado no centro da testa). Esta luz reflectida é fotografada por um sensor CMOS e o sinal de vídeo é passado para a unidade de pré-processamento electrónica para elaborar o tracking do material que foi reflectido.

O SmartNav pode processar qualquer imagem oriunda de qualquer fonte IR; normalmente é o reflexo do material reflector ou uma fonte activa de IR, como um LED. O utilizador pode fazer o tracking de diversos objectos, colocando pontos ou LEDs reflectivos sobre o objecto. O SmartNav tem um campo de visão 45 graus e qualquer objecto que está a ser detectado deve permanecer nesse campo de visão. O preço da solução ronda os 500 dólares.



Figura 13 - Dispositivo utilizado no SmartNav para o tracking

2.4.3.2 Tracker Pro

O Tracker Pro é um sistema de mãos-livres para o rato que traduz movimentos da cabeça em movimentos do cursor do rato. Utiliza um sensor (e.g. Figura 14) que controla um minúsculo, reutilizável ponto reflectivo (e.g. Figura 15) e que é colocado na testa do utilizador. Quando a cabeça se move, o cursor move-se.

O sistema possibilita a utilização em Macintosh ou PC através da porta USB. Tarefas específicas, tais como desenho, jogar, navegar na internet e manutenção geral de computador são facilmente realizados com o Tracker Pro.



Figura 14 - Dispositivo utilizado no Tracker Pro para o tracking

O sistema funciona da maneira semelhante ao SmartNav: a luz infravermelha é emitida a partir dos LEDs e é reflectida de volta para o dispositivo por um material reflector. O preço do sistema é superior, cerca de 900 dólares.



Figura 15 - Material reflector utilizado no Tracker Pro

O Tracker Pro permite a utilização de sistemas complementares para realizar as funções de um clique do rato. Um dos programas utilizados é o Madentec's Magic Cursor para Windows ou o software similar para Macintosh.

Este sistema permite que pessoas com deficiências físicas e neurológicas graves com um uso limitado de suas mãos utilizem computadores no seu quotidiano.

3 Descrição, Análise e Caracterização do Sistema proposto

O sistema “Mobility Camera” tem como objectivo permitir o acesso aos sistemas computacionais a pessoas com restrições físicas graves. O sistema faz o tracking do movimento do rosto do utilizador do PC através de uma Web câmara e o converte para movimentos do rato.

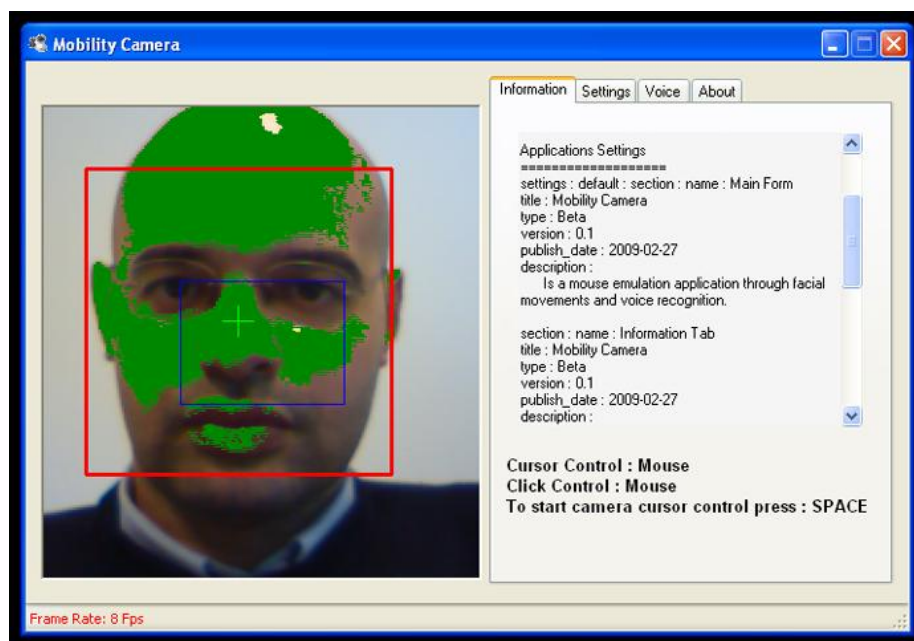


Figura 16 – Mobility Câmara com tracking facial recorrendo ao Classificador de Haar facial e identificação da cor da pele.

A nível de detecção facial e tracking, a aplicação utiliza um sistema que oferece o melhor compromisso entre precisão, usabilidade e robustez.

Um sistema é robusto quando continua a funcionar apesar de ter falhas de processamento. Logo, a solução deve utilizar vários métodos para garantir o seu bom funcionamento, mesmo com falhas dos algoritmos individuais. Por exemplo, podemos utilizar os classificadores de Haar para a detecção da face e o Optical Flow para o tracking, e como método alternativo para a recuperação automática na possibilidade do tracking ou a detecção falhar, um segundo algoritmo baseado nos classificadores de Haar e na detecção da cor de pele, como se pode ver na Figura 16. Os pontos fortes de cada algoritmo colmatam os fracos, e.g. o algoritmo baseado na detecção da cor de

pele é aplicado apenas no tracking, enquanto que os classificadores de Haar detectam a face.

O sistema de robustez [29] implementado é baseado no processo 1:N Redundância como se pode ver na Figura 17. 1: N é uma técnica utilizada na arquitectura de projectos onde existe um único objectivo para vários sistemas. Este objectivo é capaz de ser realizado por qualquer um dos sistemas activos. Esta técnica oferece redundância, a um custo muito menor por existir uma partilha de recursos entre as unidades primárias. Esta abordagem só funciona bem quando todas as unidades primárias têm funções muito semelhantes, permitindo assim a concretização do objectivo inicial de qualquer uma das unidades primárias, se um deles falhar.

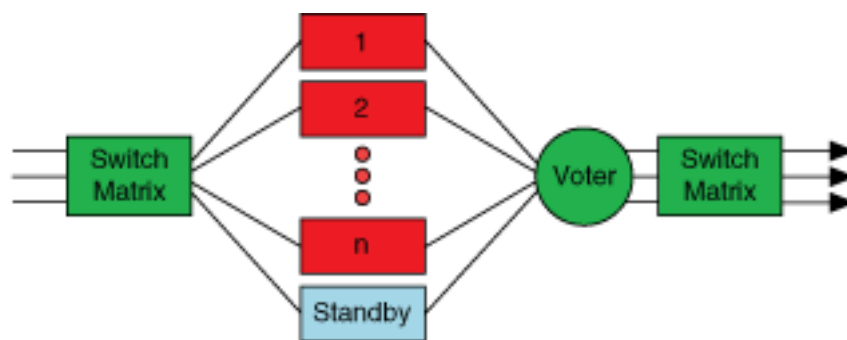


Figura 17 - Sistema de redundância

O sistema, de uma maneira geral, pode ser esquematizado através de actores, actividades e um conjunto de classes que representam as diversas componentes da aplicação, como ilustra a Figura 18. O sistema de redundância encontra-se implementado no processamento de imagem e na classe de algoritmos.

O método de processamento de imagens escolhe numa ordem pré definida qual o algoritmo que irá analisar a frame em causa delegando à classe algoritmos o processamento. Existem 4 algoritmos com diferentes técnicas que se alternam numa situação de falha na detecção ou tracking facial.

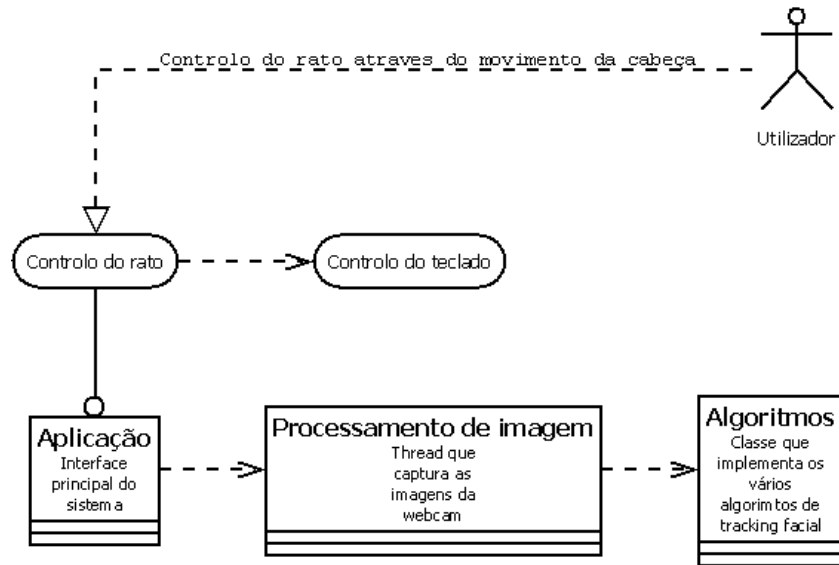


Figura 18 - Fluxo funcional do sistema

3.1. Sistema de visão

O sistema de visão por computador recebe e exibe um vídeo em tempo real do utilizador sentado na frente do computador. O vídeo é obtido através de uma câmara montada acima ou abaixo do monitor do utilizador. Visualizando o vídeo, o utilizador ou um assistente inicializa o processo de controlo do rato com a cabeça. Inicialmente deve-se ajustar o sistema de modo a que a câmara fica centrada no utilizador. O sistema de visão consequentemente determina as coordenadas iniciais automaticamente e depois, calcula-as em imagens subsequentes.



Figura 19 - Tracking de pontos com Haar Cascade e mapeamento de área

As coordenadas do ponto inicial são depois enviadas em cada frame para o computador do utilizador estabelecendo a posição do rato como podemos ver na Figura 19.

A Figura 20 ilustra a configuração padrão da utilização do sistema de tracking onde podemos visualizar o utilizador com o computador e câmara na posição frontal.

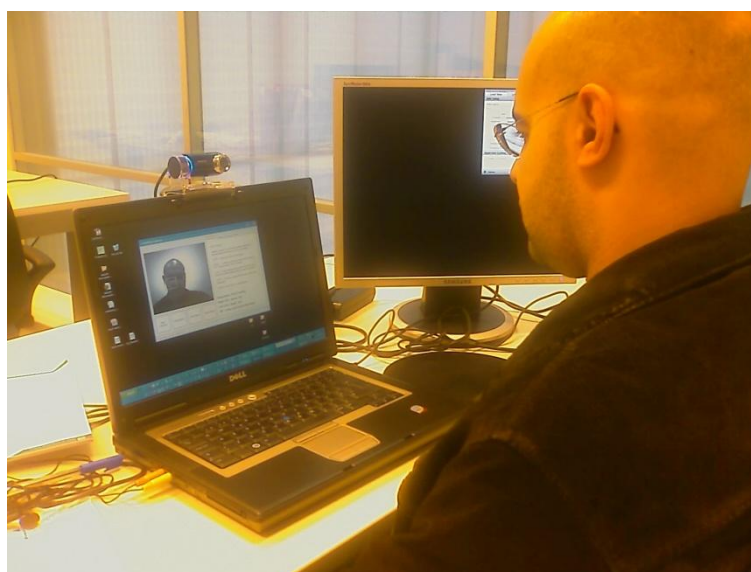


Figura 20 - Configuração de utilização

3.2. Algoritmos

A detecção dos movimentos do utilizador é feita utilizando uma webcam. Para capturar as imagens a partir da webcam utiliza-se um wrapper para C# do OpenCV - o EMGU [39].

Alguns dos algoritmos e estratégias experimentadas para detectar movimentos não tiveram o sucesso desejado como mencionado na secção 3.6 devido essencialmente à reduzida nitidez e contraste das imagens capturadas pelas webcams e ao tempo de processamento dos algoritmos na Framework Dot Net.

Contudo, os resultados obtidos levantaram questões essenciais, tais como o tempo de execução mínimo de FPS (Frames Por Segundo) de cada algoritmo, que implicaram novas abordagens para se chegar a uma solução adequada.

O algoritmo geral proposto e implementado baseia-se na detecção de faces utilizando as técnicas de detecção da cor da pele e o algoritmo HaarCascade (e.g. Figura 21). A nível de tracking as soluções utilizadas são o Optical Flow, isto é, o tracking de pontos de uma frame para outra, o Haar Cascade com interpolação de faces nas diferentes frames, o Haar Cascade com mapeamento relativo e a detecção da cor da pele com mapeamento relativo.

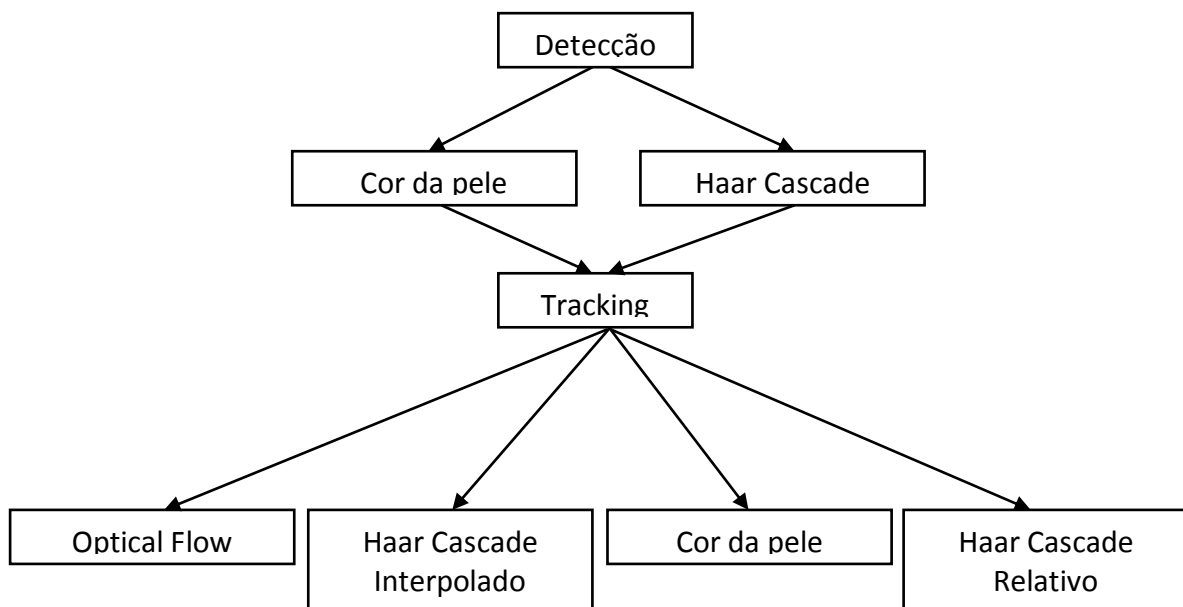


Figura 21 – Esquema das técnicas e algoritmos utilizados no algoritmo proposto

A classe algoritmos implementa 4 algoritmos que efectivamente são utilizados na análise de cada frame. Cada um desses algoritmos inclui um conjunto de técnicas para melhor detectar e realizar o tracking de uma face.

Estes 4 algoritmos estão encadeados num esquema (e.g. Figura 22) que garanta a robustez do algoritmo geral, isto é, estão organizados por ordem permitindo em situação de falha de detecção ou de tracking que o próximo algoritmo tente garantir o funcionamento normal, ou seja, a detecção e o tracking de faces. A situação de falha é controlada por um parâmetro configurável denominado "Error". O algoritmo inicial também pode ser definido por um parâmetro configurável.

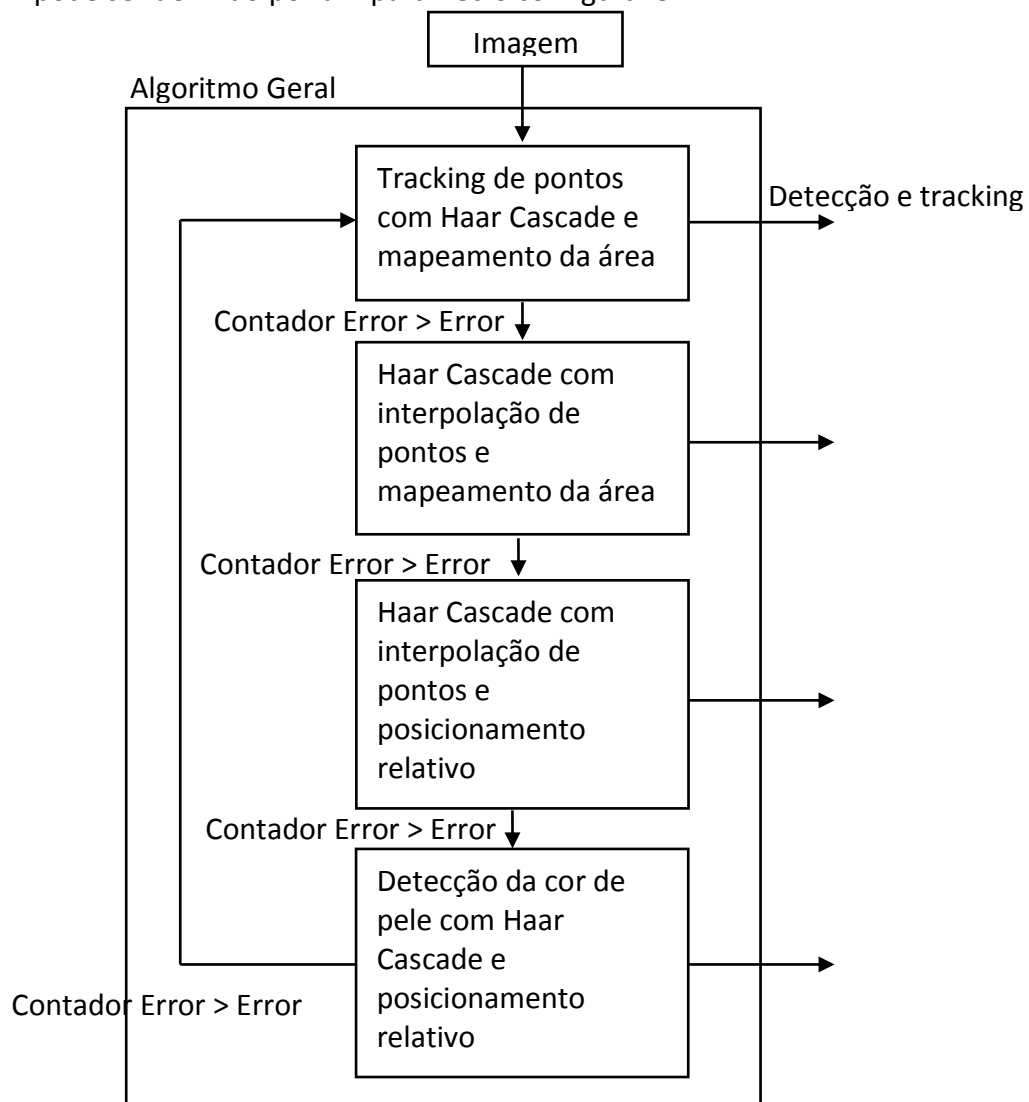


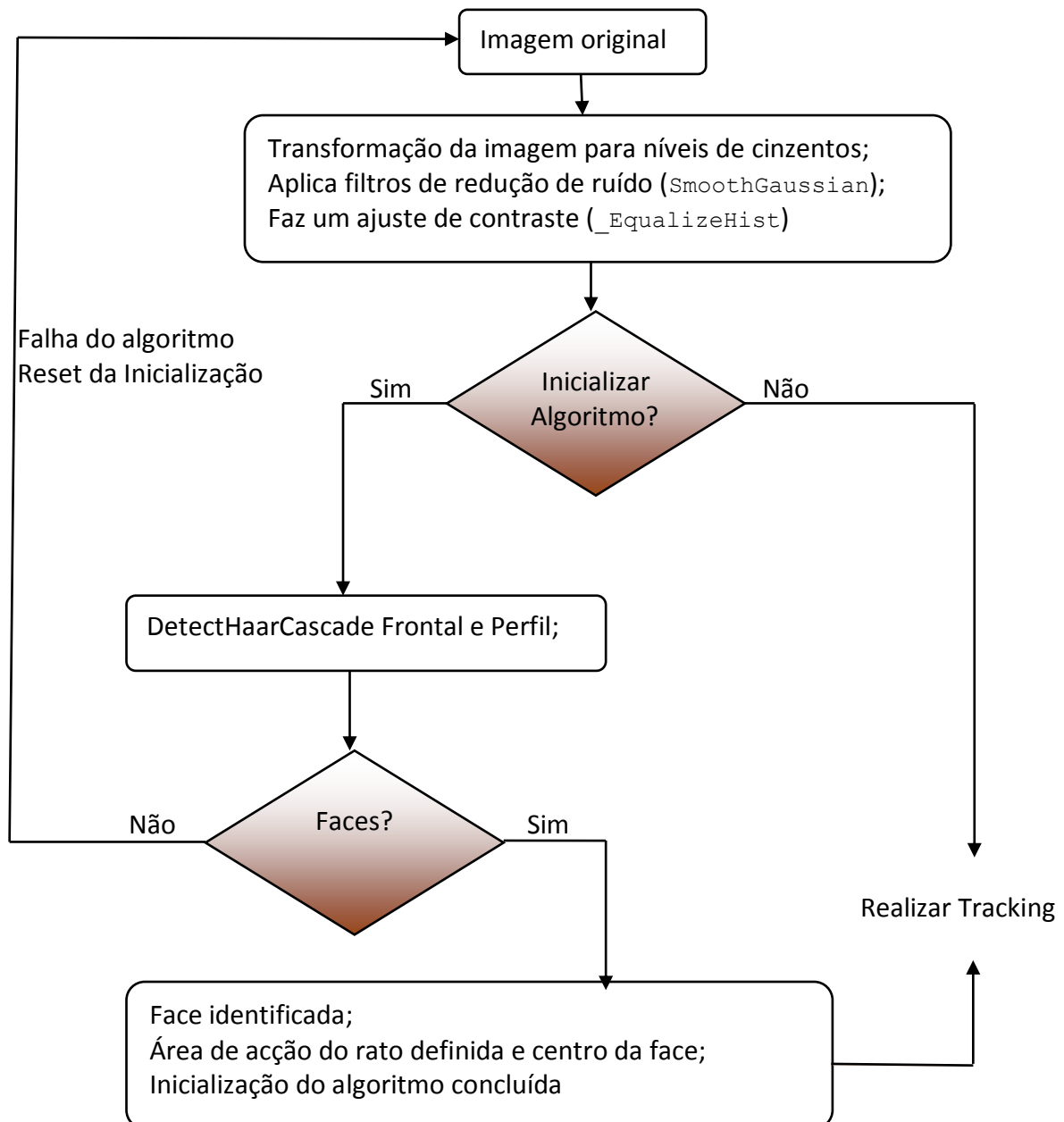
Figura 22 – Encadeamento dos algoritmos

Seque-se uma descrição mais pormenorizada dos algoritmos.

3.2.1 Tracking de pontos com Haar Cascade e mapeamento da área

Este método permite tracking facial em tempo real. É o método inicial da aplicação e utiliza o Haar Cascade para a detecção da face frontal/perfil e para detectar a área de acção do rato. O tracking é realizado em cima dos dados iniciais obtidos através do Haar Cascade e baseia-se no tracking de um conjunto de pontos (Optical Flow) com determinadas características.

A fase de detecção baseia-se na aplicação de filtros de redução de ruído, de um ajuste no contraste e na transformação da imagem original a cores para tons de cinzentos. Depois se o algoritmo não estiver inicializado detectamos através do método do OpenCV - DetectHaarCascade as possíveis presenças de faces na imagem.



Se a detecção da face falhar existe um reset das inicializações e uma sinalização de erro que posteriormente irá decidir a mudança do algoritmo para o seguinte. Caso se detecte uma face inicializa-se a área de acção do rato, o ponto central da face, a área da face e dá se como concluída a inicialização como se pode ver na Figura 23.

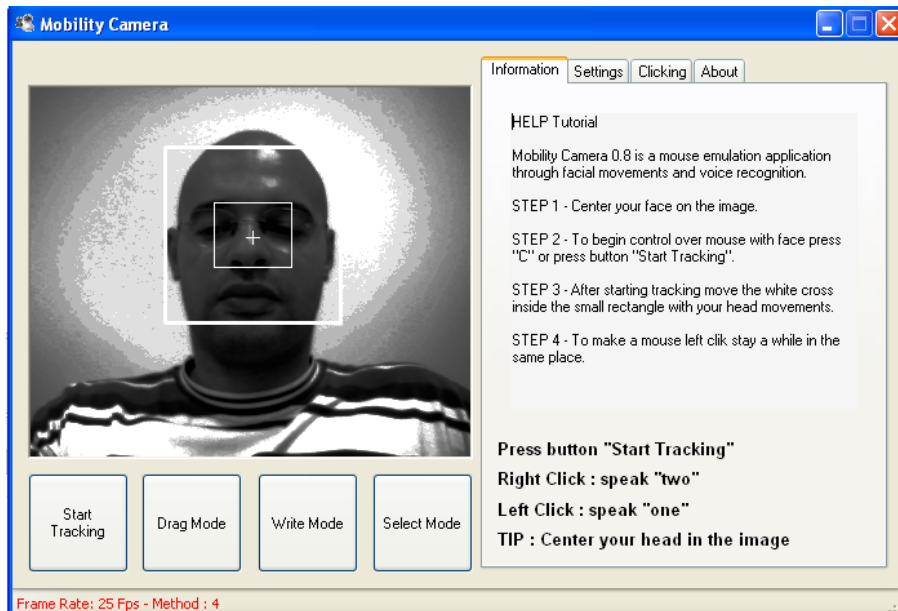
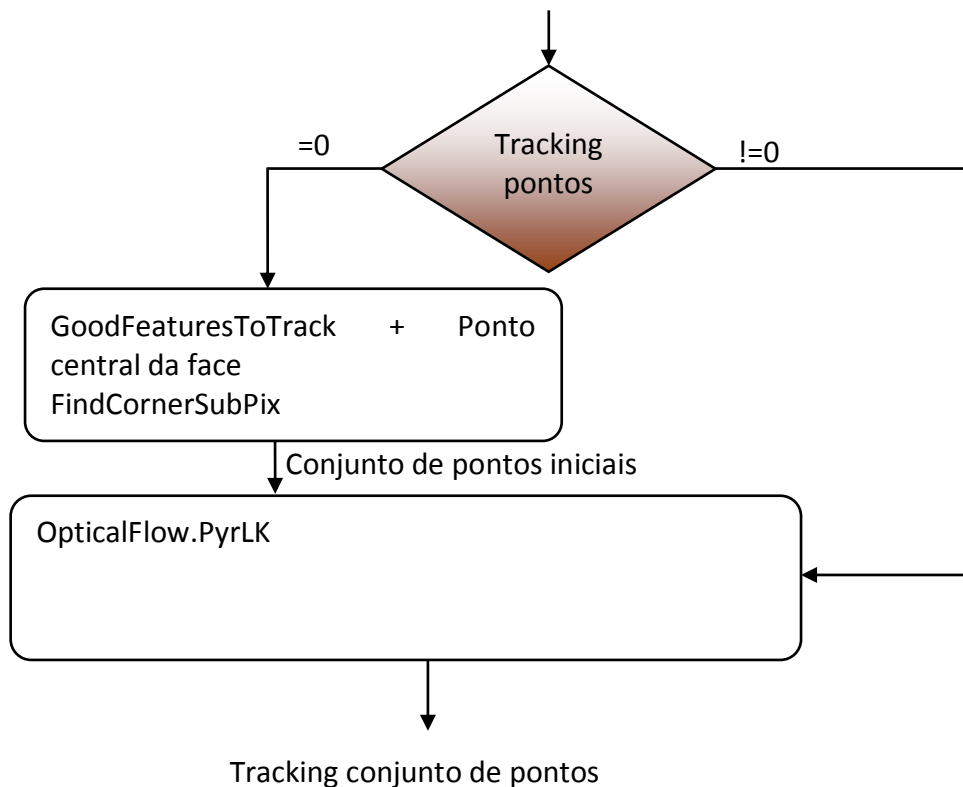


Figura 23 - Área de acção do rato, o ponto central da face, a área da face

Depois da inicialização do algoritmo o a passo seguinte é efectuar o tracking.



Se existirem pontos com tracking então a aplicação actualiza apenas a sua posição recorrendo ao método OpticalFlow.PyrLK [19]. Se não existirem pontos em tracking o algoritmo calcula um conjunto de pontos ideais através do método GoodFeaturesToTrack adicionando o ponto central da face e realizando uma optimização desses pontos com o método FindCornerSubPix. O resultado de cada iteração deste algoritmo é sempre um conjunto de pontos ideais para seguir como se pode ver na Figura 24.



Figura 24 - Tracking 2 pontos dentro da área de acção do rato

Caso não existam pontos a seguir dentro da área de acção do rato é sinalizado uma falha do algoritmo e um reiniciação do sistema.

Por fim é realizado o mapeamento da área de acção do rato para a área real do monitor. Isto é, depois de identificado o ponto de tracking ideal na frame é necessário normalizar a escala, para que esse ponto corresponda a um pixel do monitor, onde o cursor será colocado. O mapeamento é feito da seguinte forma:

Ponto mapeado X = ((ponto actual X – ponto inicial da área X) / amplitude da área X) * 1000;

Ponto mapeado Y = ((ponto actual Y – ponto inicial da área Y) / amplitude da área Y) * 1000;

Desta maneira o algoritmo devolve um ponto normalizado que depois é traduzido para as coordenadas do monitor. A posição do cursor é atribuída por este ponto.

3.2.2 Haar Cascade com interpolação de pontos e mapeamento da área

Este segundo sistema de tracking facial em tempo real baseia-se apenas na utilização do Haar Cascade para a detecção e tracking da face frontal/perfil e para detectar a área de acção do rato. O tracking inclui também a interpolação de pontos da face, ponto de origem e tamanho.

O mapeamento da área de acção do rato é idêntico ao algoritmo anterior facilitando a curva de aprendizagem do utilizador. A taxa de actualização do rato é controlada por um evento que é despoletado num determinado espaço de segundos, o que implica que a posição do rato apenas é actualizada com intervalo de tempo pré definido (0,5 segundos). Com esta aproximação conseguiu-se estabilizar mais a paragem do rato para efectuar cliques.

A fase de detecção baseia-se na aplicação de filtros de redução de ruído, de um ajuste no contraste e a transformação da imagem original a cores para tons de cinzentos. Depois se o algoritmo não estiver inicializado detectamos através do método do OpenCV - DetectHaarCascade as possíveis presenças de faces na imagem como podemos ver na Figura 25.

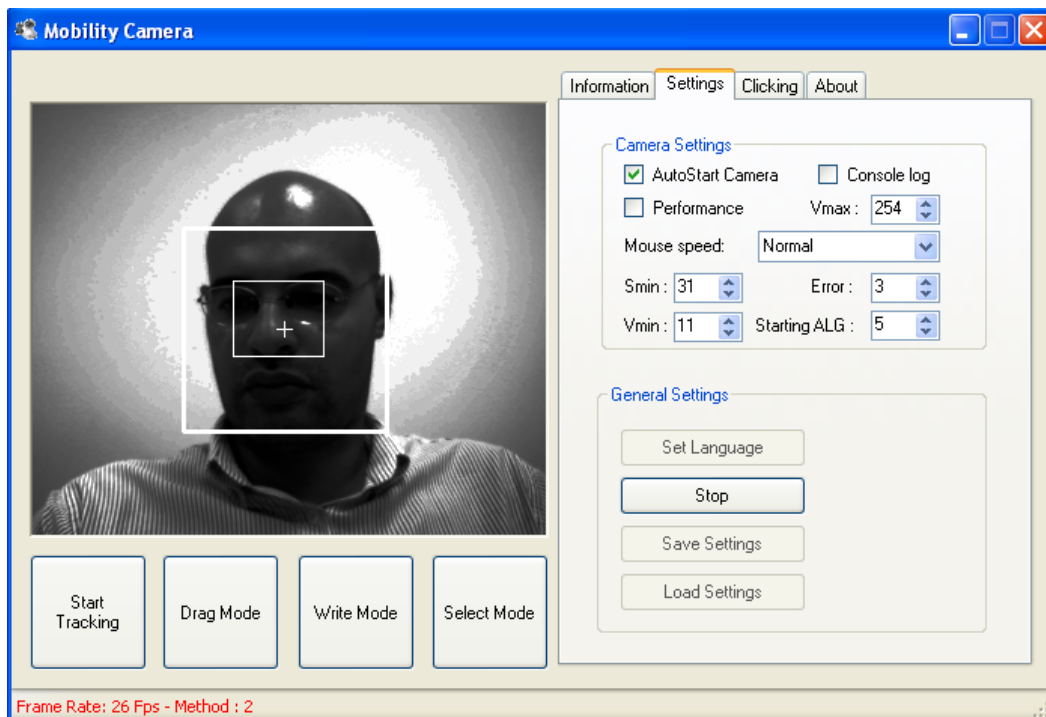
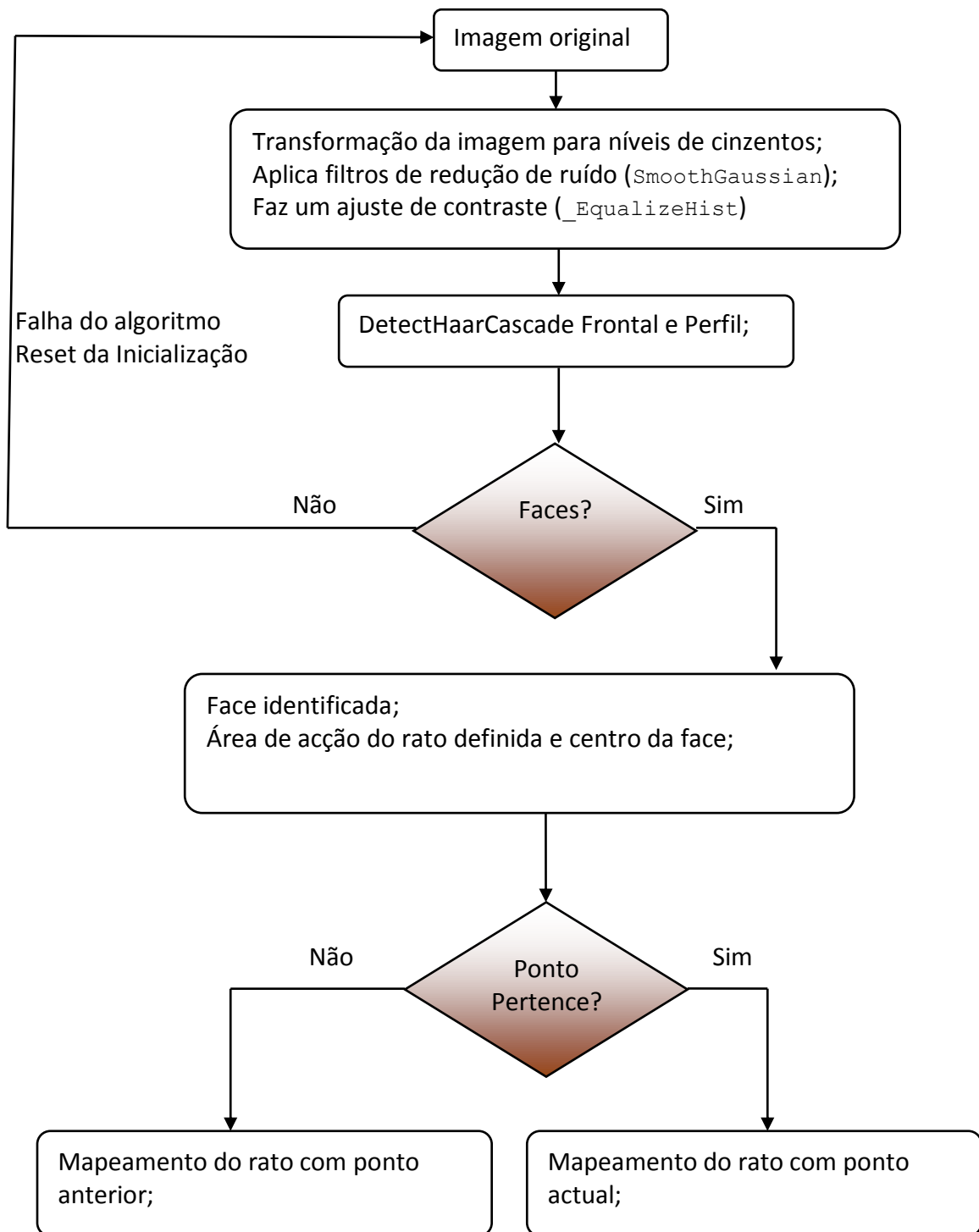


Figura 25 – Detecção e tracking de faces



Desta maneira o algoritmo devolve um ponto normalizado que depois é traduzido para as coordenadas do monitor. A posição do cursor é atribuída por este ponto. A sinalização de falha no algoritmo tem um limite que quando é atingido faz com que haja a mudança automática de algoritmo. Ele é avaliado segundo: Número de falhas do algoritmo > Factor predefinido * Margem de erro (parâmetro da aplicação).

3.2.3 Haar Cascade com interpolação de pontos e posicionamento relativo

Este terceiro algoritmo de tracking facial em tempo real utiliza o Haar Cascade para a detecção e tracking da face frontal/perfil. O tracking inclui também a interpolação de pontos da face, ponto de origem e tamanho. Para detectar a área de acção do rato é utilizada o centro da imagem original.

O mapeamento da área de acção do rato é relativo à posição do rato e à área central de acção. A taxa de actualização do rato é controlada pelo algoritmo mas em compensação o algoritmo tem uma pequena área no meio da imagem para parar o rato como se pode ver na Figura 26.

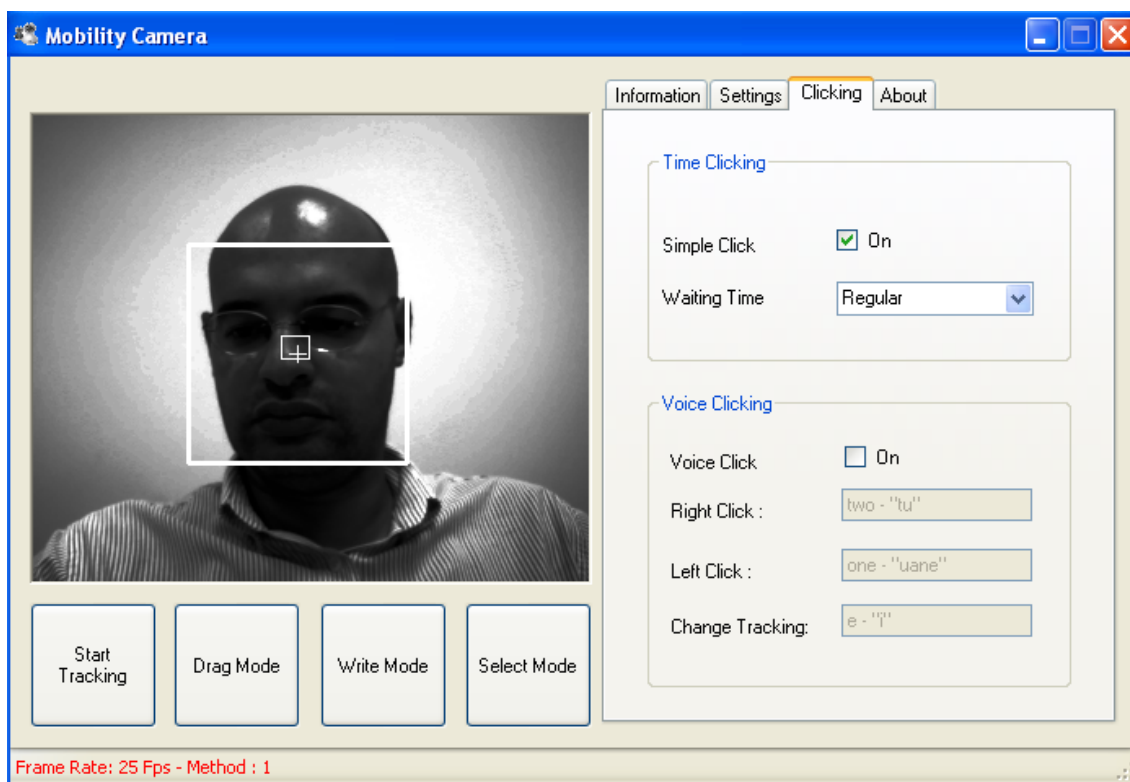


Figura 26 – Situação de tracking em que o rato está imóvel

O utilizador consegue mexer o rato movendo a cabeça a cruz de tracking para a zona exterior ao pequeno rectângulo no central da imagem. Quando a cruz do ponto de tracking é detectada como estando fora do rectângulo o rato começa a receber actualizações da posição.

O algoritmo é semelhante ao anterior mudando apenas a forma como mapeia o movimento do rato. A condição de falha do algoritmo continua a ser a falha na detecção da cara.

3.2.4 Detecção da cor de pele com Haar Cascade e posicionamento relativo

Este quarto algoritmo de tracking facial em tempo real utiliza a detecção da cor da pele para o tracking e o Haar Cascade para a detecção da face frontal/perfil. Para detectar a área de acção do rato é utilizada o centro da imagem original.

O mapeamento da área de acção do rato é relativo à posição do rato e à área central de acção. A taxa de actualização do rato é controlada pelo algoritmo mas em compensação o algoritmo tem uma pequena área no meio da imagem para parar o rato como se pode ver na Figura 27.

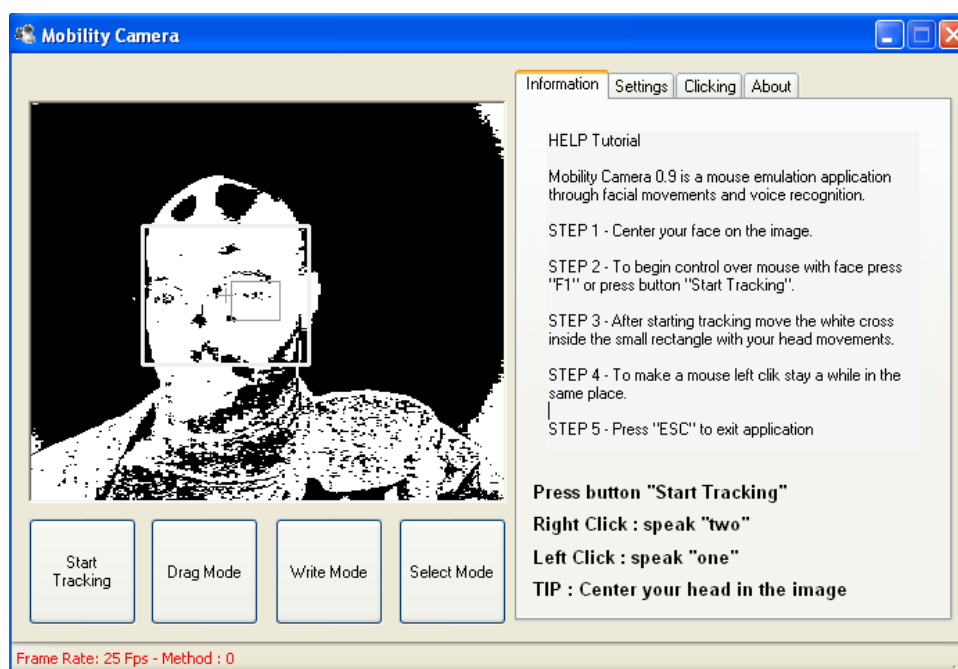


Figura 27 – Tracking da cor da pele

Este algoritmo é o que é mais afectado pelas condições de iluminação pois baseia-se na análise do espaço de cor HSI. A vantagem deste algoritmo é que é estável e eficiente quando reúne as condições necessárias para ter um bom desempenho, ou seja, uma fonte de luz branca radial fria posicionada numa zona acima do utilizador

O algoritmo faz uma análise do espaço de cor HSI, inicializa algumas variáveis tais como a zona da cara e depois faz o tracking através do método do CamShift.

A Intel Open Source Computer Vision Library [33] contém uma implementação do algoritmo Continuously Adaptive Mean Shift Algorithm (CamShift) [35] que faz o tracking do movimento da cabeça e face usando um histograma de uma dimensão constituído pela quantidade de valores nos canais do espaço de cor HSI (e.g. Figura 28).

Como o algoritmo é projectado para consumir o menor número de tempo de CPU possível [36], um único canal (tonalidade) é considerado no modelo de cor.

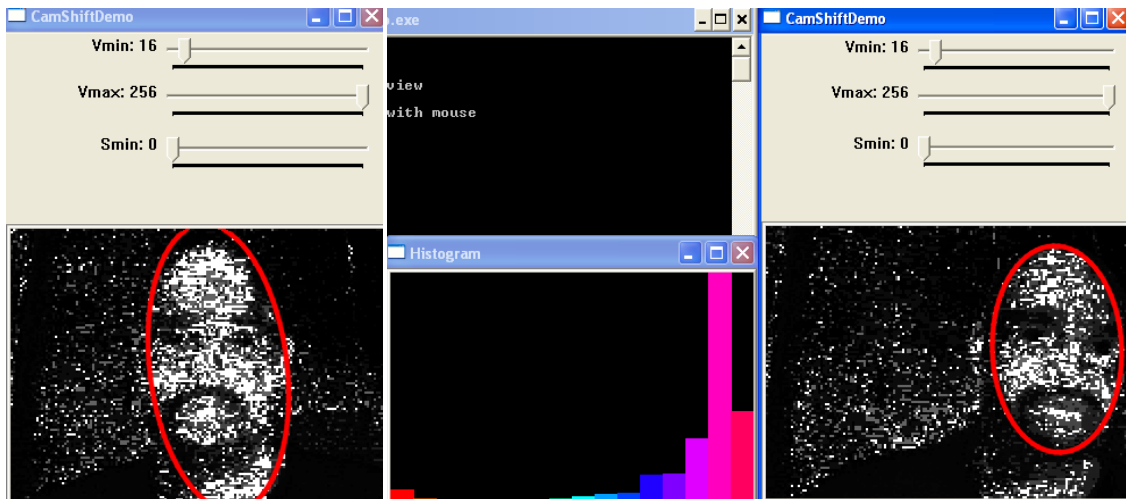
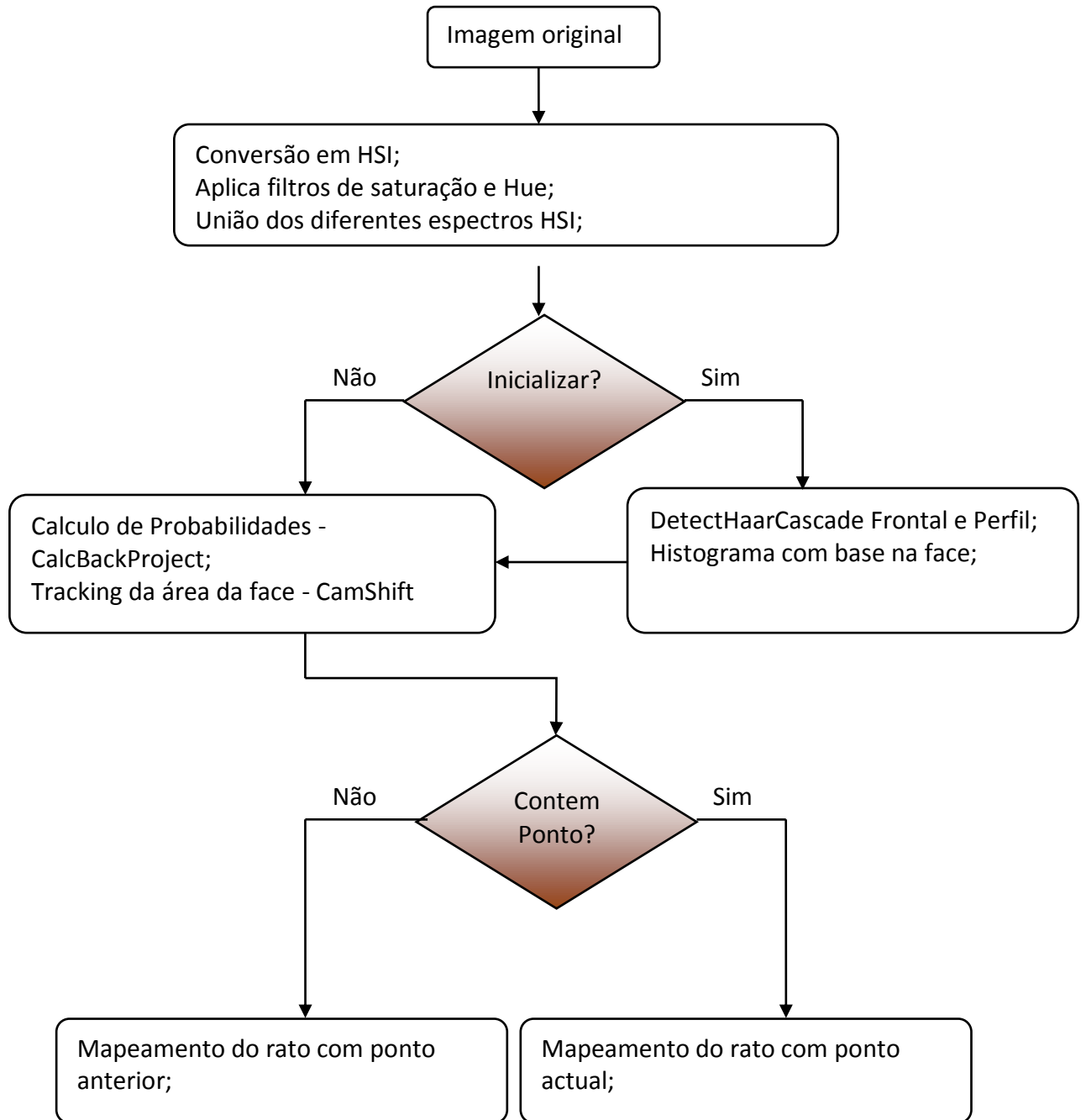


Figura 28 - Algoritmo Camsfift



3.3. Emulação dos botões do rato, Drag, Modo de Selecção e Escrita

A simulação dos cliques do rato ficou assente em duas estratégias mutuamente exclusivas que providenciam ao utilizador duas maneiras de interagir com os botões do rato. Essas estratégias são explicadas de seguida.

Assim, escolheram-se utilizar o reconhecimento de voz para simular os eventos do rato, recorrendo ao Microsoft Speech SDK 5.1. Este SDK inclui um motor de reconhecimento em inglês, o que facilita muito a utilização deste tipo de tecnologia, pois permite reconhecer comandos vocais do utilizador de forma simples e eficaz. O utilizador poderá, por exemplo, efectuar um clique com o botão esquerdo do rato dizendo “left”, ou fazer um clique com o botão direito do rato dizendo “right”. Como comandos podem usar-se quaisquer palavras inglesas. Logo é possível alterar o ficheiro contendo as palavras mencionadas na emulação do clique para colocar palavras ou sons mais simples, por exemplo uma única letra.

O outro sistema de interacção com os botões do rato é um método mais simples e que prevê a impossibilidade do utilizador não conseguir vocalizar sons correctamente ou estar num ambiente com muito ruído sonoro. O sistema apenas consegue realizar o clique do botão do lado esquerdo do rato e funciona da seguinte maneira: o utilizador ao manter o rato parado numa posição relativamente constante com uma margem de variação máxima pré definida (8 pixéis), durante um determinado tempo consegue realizar um clique na posição actual do rato. O tempo é configurável entre Short (2s), Regular (5s) e Long (8s) segundos como podemos ver na Figura 29.

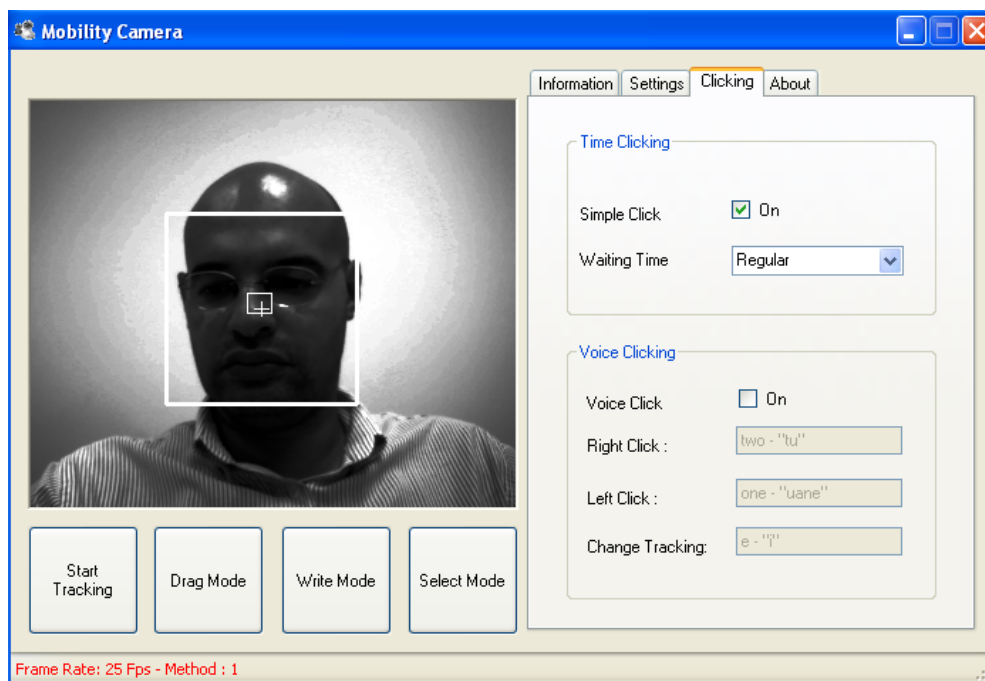


Figura 29 - Sistemas de emulação dos botões do rato

O “Start Tracking ” inicializa a utilização dos algoritmos de tracking mencionados anteriormente no capítulo 3.2.1.

O “Drag Mode” possibilita arrastar objectos, isto é, é uma acção que contempla um clique e um mover do rato. Ao entrar neste modo o utilizador pode clicar num objecto e depois ao clicar num outro ponto o objecto desloca-se para lá e sai do modo Drag.

O “Write Mode” foi uma melhoria sugerida quando os primeiros testes foram realizados e faz com que o mapeamento do rato se desloque de 10 em 10 pixéis, na vertical ou na horizontal e possibilita a paragem do rato de uma maneira mais eficiente. O movimento do rato é calculado através da diferença entre a última posição e a posição actual, se a diferença na horizontal for superior a diferença vertical o rato anda para a esquerda ou para a direita consoante o valor é negativo ou positivo. O modo de escrita também abre uma instância do teclado de ecrã automaticamente.

O “Select Mode” realiza uma selecção da área de trabalho efectiva, isto é, reduz o tamanho do monitor para valores escolhidos pelo utilizador. Isto afecta o mapeamento e a zona de acção do rato no monitor. Para utilizar o modo de selecção basta clicar uma vez no botão e os próximos dois cliques irão estabelecer a área rectangular de

acção. O modo de selecção é desactivado quando se clica depois do segundo clique no botão "Select Mode".

3.4. Análise de desempenho e parâmetros

O desempenho do tracking na aplicação Mobility Camera depende de vários factores e parâmetros, tais como a iluminação ambiente, o tempo de exposição da câmara Web, o número de frames processadas e a capacidade do processador do computador. Uma boa e constante iluminação faz com que as FPS (Frames Per Second) subam consideravelmente, já o tempo de exposição da câmara Web, quanto mais baixo melhor, devido ao tempo que as câmaras levam no pós-processamento interno. As frames processadas (de 3 em 3 frames) ajudam a aliviar a carga do processado tornando aplicação mais rápida.

O tamanho da imagem inicial também é um factor critico pois quanto maior a área, maior o tempo de processamento necessário, embora aumente o número de características para análise. Quanto menor a área em análise mais FPS obtemos.

A referência bibliografia [20] analisa esta questão em detalhe e explica que a dimensão da imagem de aquisição não é o único factor, mas mais importante para o desempenho que a dimensão, é a complexidade da imagem.

Existem um conjunto de parâmetros que podem ser configurados via as configurações da aplicação e que possibilitam uma configuração do desempenho até um determinado grau como pode-se ver na Figura 30.

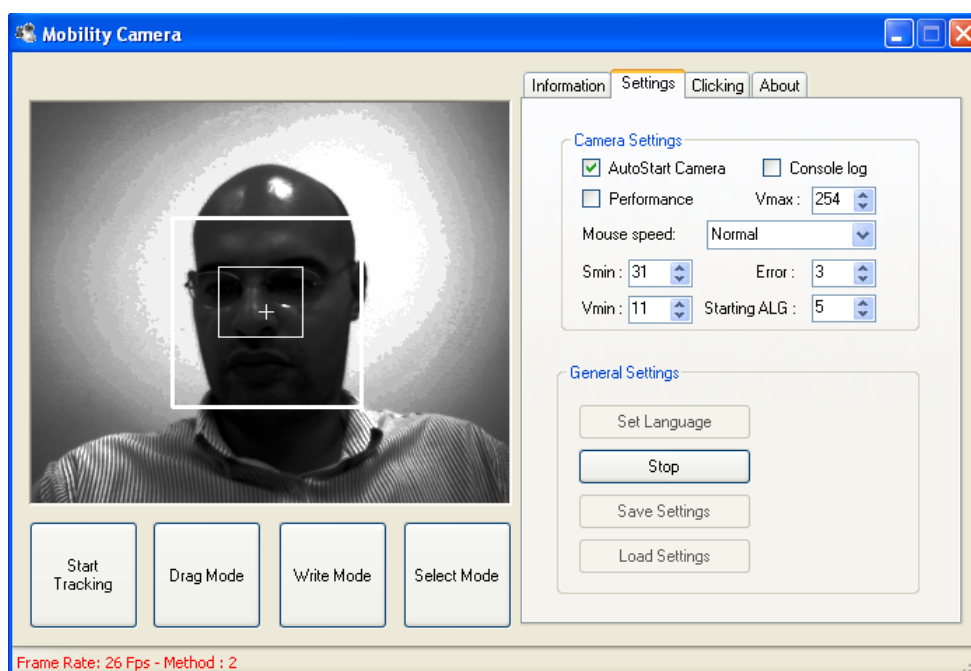


Figura 30 - Configurações da aplicação

Estes valores controlam a velocidade do rato em modo normal, a escolha do algoritmo inicial ou mesmo se o algoritmo muda após falhar consecutivamente (e.g. parâmetro Error). Também se pode controlar o desempenho eliminando a visualização da imagem, porém esta opção é recomendada apenas para utilizadores com muito treino de utilização da aplicação.

3.5. Especificação do Hardware e Software

Foram utilizados equipamentos para teste com características de hardware sensivelmente diferenciadas. O sistema mais fraco que foi utilizado para teste e que apresentou um desempenho aceitável foi um portátil Acer Aspire 5634, Intel Core Duo T5600 1,83 GHz com 4 Gb de memória com uma webcam externa Logitech QuickCam Messenger e com sistema de microfone incorporado. A nível de software foi testado em sistemas Windows XP, DOT NET Framework 3.5 e com Sapi SDK 5.1. O sistema de captura de visão utilizou um wrapper do OpenCV, o EMGU e a linguagem de desenvolvimento foi o C#.

As condições de utilização também sofreram variações de iluminação mas como base foram utilizados ambientes que possuam características que permitam capturar imagens com a melhor qualidade possível. A condição mais importante é a colocação do sistema num ambiente com boa e constante iluminação. A iluminação deve ter pouca influência da luz natural vinda do exterior, para se manter constante e impedir a webcam de se tentar adaptar às mudanças de luminosidade. A luz deve ser, de preferência, branca; o melhor funcionamento é conseguido com uma fonte de luz fluorescente colocada no tecto sobre o utilizador.

Devem ser evitadas fontes de luz ao nível do utilizador e da webcam. Essas luzes produzem reflexões e sombras indesejadas sobre a cara do utilizador, fazendo com que determinados algoritmos presentes na aplicação não funcionem correctamente. A última condição é fácil de garantir, a webcam deve estar ao nível da cara do utilizador e centrada como podemos ver na Figura 31.



Figura 31 – Condições de utilização do sistema

3.6. Análise de outras soluções

Algumas das soluções analisadas e implementadas que após testes de desempenho não foram incluídas na versão final da aplicação, serviram para ter perspectivas diferentes acerca da detecção e tracking facial.

As mais significativas foram a utilização de SURF features [21] para a detecção ou possível tracking de uma zona ou face e o template matching. Ambos algoritmos servem para detectar uma face ou realizar o tracking, entretanto ambos demonstraram-se lentos para o tracking.

O SURF features necessita de uma imagem inicial detalhada e com um tamanho mínimo como podemos ver na Figura 32 para realizar a detecção. Logo, não servia para esta fase pois necessitaria de uma imagem inicial indicativa da zona da face. Por outro lado, também não poderia ser usado no tracking pois tem um limite inferior do tamanho da imagem fixo e a implementação do método no EMGU não é a mais eficaz causando de vez em quando a corrupção da memória.



Figura 32 - SURF features

Em relação ao template matching [37][38] a questão problemática é o tempo de execução da método que não é o ideal. Este desempenho é relativo à velocidade que necessitamos para executar a aplicação de uma maneira eficiente. Temos que ter em conta que o nosso algoritmo está a analisar 25 fps, logo não pode demorar na análise de cada frame.

O template matching tenta procurar uma sub-imagem predeterminada tendo como base uma imagem maior como podemos ver na Figura 33.

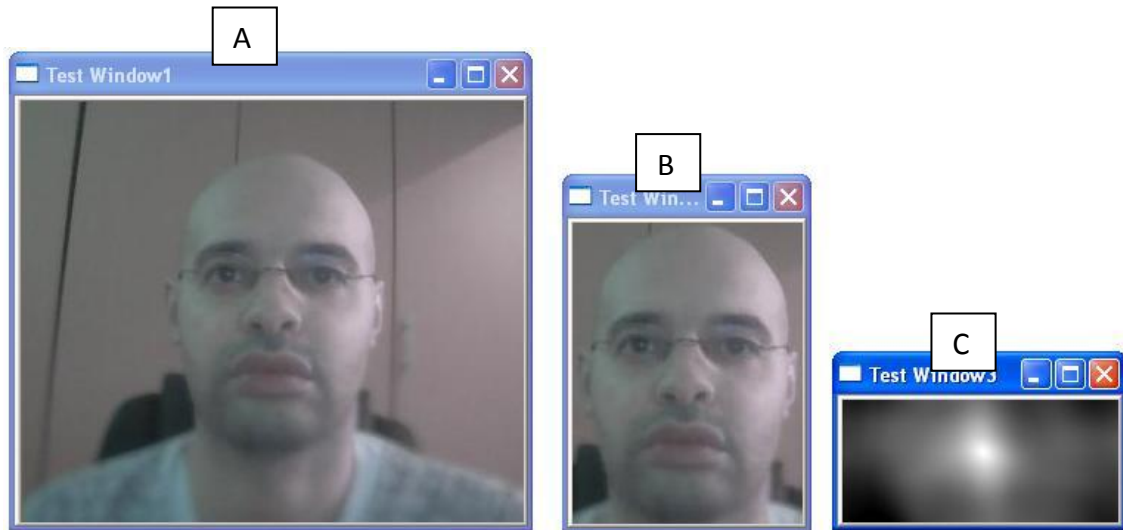


Figura 33 – Resultado do template matching : Procura B em A = C

4 Resultados experimentais

Este capítulo apresenta os resultados experimentais efectuados em diferentes tipologias de público-alvo e faz uma análise do desempenho.

A análise do grupo heterogéneo de pessoas que testaram a aplicação em diferentes ambientes de trabalho foi realizada através de entrevistas. A entrevista consistia numa fase inicial em formação para o utilizador, no preenchimento de questionários (Anexo 1) e gravações de vídeo. Após análise dos dados foram identificadas as mais-valias e os aspectos negativos na utilização da aplicação. Os indivíduos inquiridos apresentaram algumas propostas de melhorias que podem contribuir para uma melhor satisfação e motivação das pessoas que utilizam a aplicação.

4.1. Pessoas sem limitações físicas

O grupo de pessoas (e.g. Figura 34) sem limitações físicas que testaram a aplicação é composto por diversões escalões etários e com vários níveis de qualificação. A observação do desempenho das pessoas bem como a análise dos comentários e do questionário foram a base destes resultados experimentais.

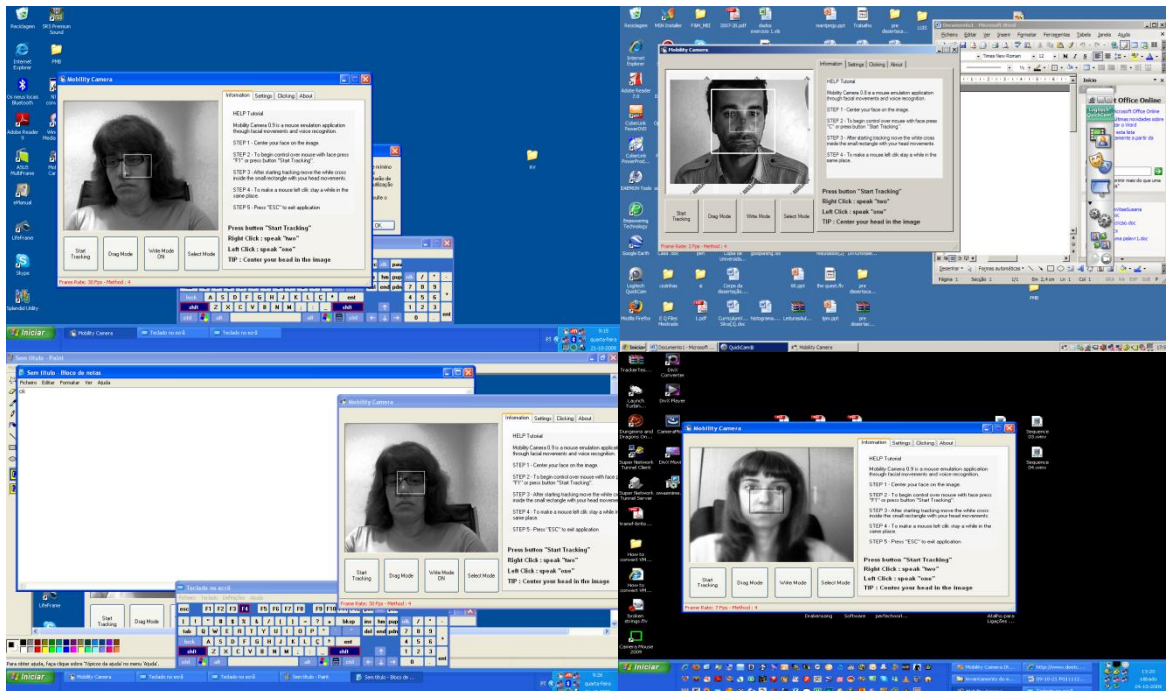


Figura 34 – Algumas sessões de testes efectuadas

O número de pessoas sem impedimentos físicos que participaram nos testes é relativamente baixo (cinco pessoas) e deve-se a um conjunto de factores tais como o atraso na codificação de alguns algoritmos da aplicação final, nos prazos apertados referentes à entrega da tese e dificuldade de encontrar candidatos disponíveis.

| Método | Tarefas | Tmp Apl |
|--|---------------------------------------|---------|
| Haar Cascade with Point Tracking and scale mapping | Abrir teclado de ecran | 65s |
| | Abrir\Escriver\Navegar no browser | 660s |
| | Abrir\Escriver texto curto no Notepad | 105s |
| Haar Cascade with Point Interpolation and scale mapping | Abrir teclado de ecran | 150s |
| | Abrir\Escriver\Navegar no browser | 0s |
| | Abrir\Escriver texto curto no Notepad | 0s |
| Haar Cascade with Point Interpolation and relative position | Abrir teclado de ecran | 45s |
| | Abrir\Escriver\Navegar no browser | 540s |
| | Abrir\Escriver texto curto no Notepad | 135s |
| Skin Color Detection with Haar Cascade and relative position | Abrir teclado de ecran | 0s |
| | Abrir\Escriver\Navegar no browser | 0s |
| | Abrir\Escriver texto curto no Notepad | 0s |
| Camera 2009 | Abrir teclado de ecran | 30s |
| | Abrir\Escriver\Navegar no browser | 150s |
| | Abrir\Escriver texto curto no Notepad | 30s |
| Rato Normal | Abrir teclado de ecran | 5 |
| | Abrir\Escriver\Navegar no browser | 35 |
| | Abrir\Escriver texto curto no Notepad | 5 |

Tabela 2 - Referência de tempo de execução das tarefas propostas

A combinação de algoritmos, o tipo de tarefas e o tempo de cada tarefa foram um dos pontos de análise como se pode ver na Tabela 2, o outro foram as sugestões dos utilizadores e a pontuação do questionário. Os dados da Tabela 2 confirmam o que os questionários indicam, isto é, que a utilização deste tipo de aplicações por parte de pessoas que não tem impedimentos físicos é improvável pois não estão dispostos a perder tanto tempo a realizar as tarefas. Este problema é devido à comparação com a precisão que os utilizadores têm com o controlo do rato com a mão.

A disparidade de tempo dispendido a realizar as tarefas é visível no Gráfico 1. Este gráfico mostra o tempo das 3 tarefas analisadas nos diferentes métodos da aplicação, com a aplicação Camera 2009 e com a utilização do rato normalmente. Todos os

valores que sejam 0 implicam que neste momento é impossível de realizar a tarefa com o algoritmo em causa devido ao tempo e paciência que é necessário.

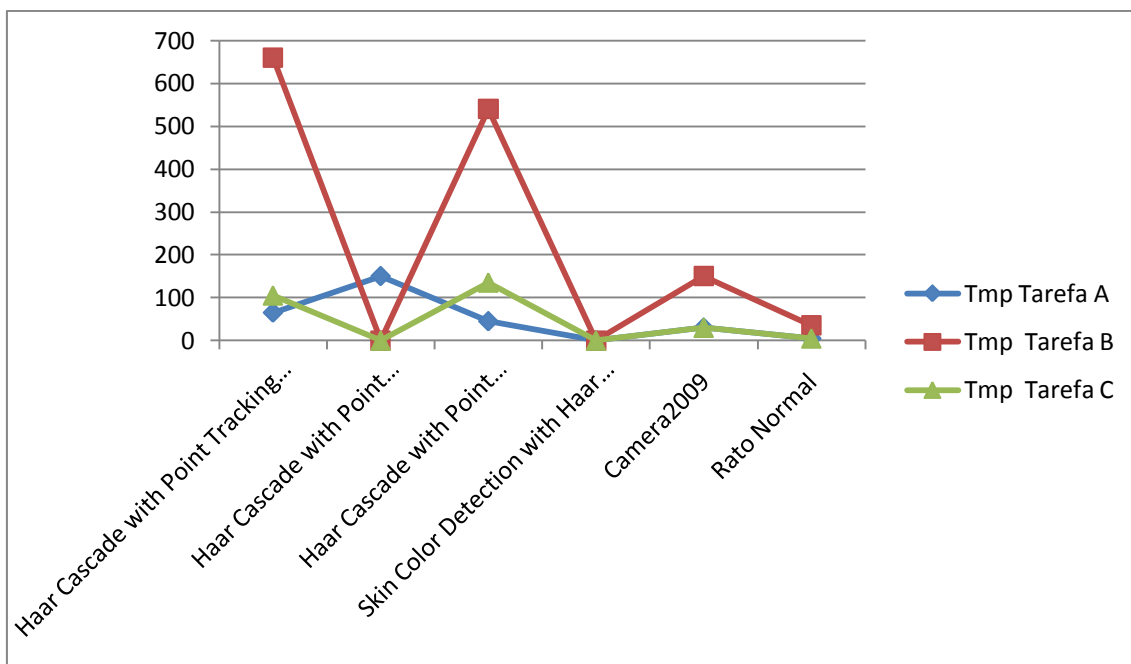


Gráfico 1 – Comparação de tempos entre tarefas e métodos utilizados

Dos questionários e da observação directa das entrevistas algumas questões a resolver foram levantadas: a dificuldade em parar o rato, o tamanho do teclado virtual. Um dos aspectos positivos foi a preferência ao utilizarem o clique por som em vez do sistema alternativo de temporização. Na Tabela 3 podemos visualizar a pontuação média referente ao questionário.

Escala referência: 1-Nada 2-Pouco 3-Razoável 4- Muito 5-Totalmente

| 1. Avaliação Global | | 1 | 2 | 3 | 4 | 5 | |
|---|------|---|---|---|---|---|------------|
| 1.1. Globalmente a aplicação apresentada está bem desenvolvida? | Nada | | | X | | | Totalmente |
| | | | | | | | |

| 2. Facilidade de navegação | | 1 | 2 | 3 | 4 | 5 | |
|---|------|----------|----------|----------|----------|----------|------------|
| 2.1. Conseguiu navegar na aplicação com facilidade? | Nada | | | X | | | Totalmente |
| 2.2. Conseguiu executar as tarefas facilmente? | Nada | | | X | | | Totalmente |
| 3. Simplicidade | | 1 | 2 | 3 | 4 | 5 | |
| 3.1. A apresentação do interface da aplicação é simples? | Nada | | | X | | | Totalmente |
| 4. Relevância do objectivo da aplicação | | 1 | 2 | 3 | 4 | 5 | |
| 4.1. O objectivo da aplicação é útil e conciso? | Nada | | | | X | | Totalmente |
| 4.2. Os objectivos estão de acordo com as necessidades do utilizador. | Nada | | | | X | | Totalmente |
| 5. Manter a consistência | | 1 | 2 | 3 | 4 | 5 | |
| 5.1. Os diferentes sistemas de controlo do rato apesar de diferentes são úteis? | Nada | | | | X | | Totalmente |
| 6. Tempo suportável | | 1 | 2 | 3 | 4 | 5 | |
| 6.1. O tempo para executar uma tarefa é aceitável? | Nada | | | | X | | Totalmente |
| 7. Foco no utilizador | | 1 | 2 | 3 | 4 | 5 | |

| | | | | | | | |
|--|------|--|--|---|--|--|------------|
| 7.1. A aplicação adapta-se facilmente ao ambiente do utilizador permitindo executar rapidamente uma nova tarefa? | Nada | | | X | | | Totalmente |
|--|------|--|--|---|--|--|------------|

Tabela 3 – Pontuação média resultante dos questionários das entrevistas

No Gráfico 2 podemos ver os resultados do questionário numa perspectiva estatística.

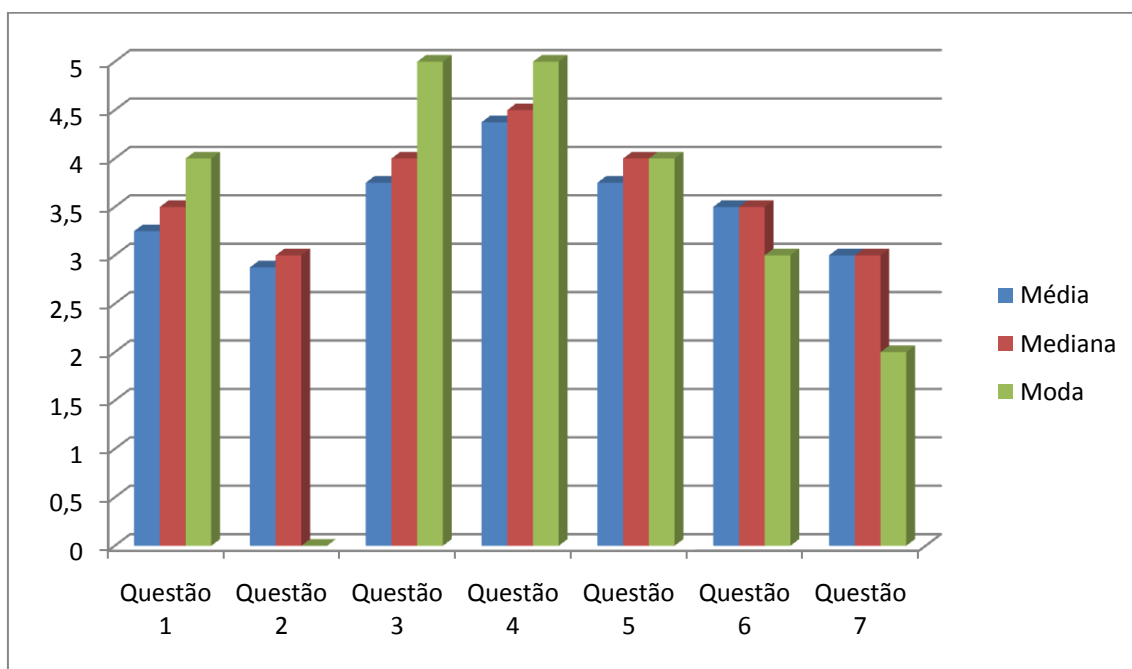


Gráfico 2 – Análise dos resultados mediante diferentes métodos estatísticos

Estes resultados permitiram efectuar algumas melhorias imediatas na aplicação pois não implicavam grande alteração em termos gerais da aplicação, por exemplo a criação do Modo de Escrita (“Write Mode”) para resolver o problema do tamanho do teclado virtual e de estabilização do rato. Este modo apesar de permitir um controlo mais estável do rato faz com que os tempos de execução das tarefas aumentem devido ao facto de se mover apenas um número reduzido de pixéis em cada actualização do cursor.

4.2. Simulação de limitações físicas

Não foi possível efectuar testes nem recolher dados relativamente a pessoas com limitações físicas devido à impossibilidade de se testar no Hospital de Guimarães e na Associação Portuguesa de Deficientes em Braga. Uma das evidências que foi visível e

referida nos testes com pessoas sem limitações é que o nível de motivação seria certamente superior e que provavelmente iriam tolerar melhor o tempo necessário para executar as tarefas.

Outros aspectos importantes a considerar seriam a disponibilidade e conhecimentos das pessoas deste grupo. Pois poderiam adulterar completamente os resultados obtidos, isto é, a vontade de utilizar a aplicação e os conhecimentos relativos a um sistema computacional são aspectos críticos na selecção de pessoas para o teste.

4.3. Análise dos resultados

De uma forma geral, os resultados obtidos na sequência dos testes efectuados foram satisfatórios e encontram-se dentro dos objectivos traçados para este trabalho. Atingiu-se uma taxa de utilização da aplicação no desempenho das tarefas elevada, tendo os tempos de execução atingido um valor médio 19 vezes superior ao tempo médio de execução das tarefas realizadas com a mão.

É, no entanto, notório que estes resultados estão muito dependentes do treino efectuado pelos utilizadores. Com efeito, verificou-se que os resultados obtidos por pessoas com um nível de conhecimento superior e com algumas sessões de treinos eram superiores às pessoas com menos qualificações ou sem afinidade à área computacional. Verificou-se também que, os utilizadores tendem para o primeiro e terceiro algoritmo na utilização da aplicação, o que nos indica que cada utilizador tem uma percepção diferente da utilização da aplicação.

A análise do grupo heterogéneo de pessoas que testaram a aplicação em diferentes ambientes de trabalho e que foi realizada através de entrevistas indicou-nos algumas conclusões. Com os preenchimentos de questionários e gravações de vídeo foram identificadas as causas que influenciam negativamente a utilização da aplicação em determinados algoritmos: paragem e estabilidade do rato em zonas muito pequenas (e.g. botões do teclado de ecrã), curva de aprendizagem elevada na utilização da aplicação e um tempo de execução de tarefas elevado em comparação com a utilização do rato com as mãos. Foram apresentadas algumas propostas de melhorias

que podem contribuir para uma melhor satisfação e motivação das pessoas que utilizam a aplicação.

Como resultados positivos temos a utilização da aplicação com sucesso por pessoas sem limitações e a validação da utilidade da aplicação bem como o desempenho da solução de robustez e eficiência dos algoritmos implementados. Especificamente temos os comentários positivos em relação à utilização do som para emular os botões do rato e aos vários tipos de soluções apresentadas (redundância - vários tipos de algoritmos e sistema de detecção de falha) que permitem adaptar-se a um utilizador com uma determinada característica.

Comparativamente às outras soluções existentes não comerciais podemos indicar que o reconhecimento de gestos da uMouse apesar de sair do âmbito desta solução é um ponto com interesse a nível de acessibilidade, já a complexidade a nível da interface do sistema é um aspecto claramente negativo.

Em relação ao Camera Mouse 2009 podemos referir que o sistema de emulação do rato é limitado e o facto de não suportar movimentos bruscos nem ter uma política de recuperação de tracking eficaz como pontos inferiores relativamente à aplicação Mobility Camera. No entanto a nível de rapidez na execução de determinadas tarefas tem um desempenho superior devido à estabilização do ponteiro do rato através de algum pós-processamento.

5 Conclusões e trabalho futuro

5.1. Conclusões

Apresenta-se, neste capítulo, as principais conclusões obtidas ao longo do trabalho, assim como um sumário dos principais aspectos desenvolvidos.

A motivação principal para o presente trabalho prendeu-se essencialmente com o desenvolvimento de um algoritmo e de uma aplicação que permitissem o reconhecimento e tracking de faces com o objectivo de criar um novo meio de interacção entre o utilizador e o sistema computacional.

Para tal, foi necessário encontrar uma solução de reconhecimento e tracking facial que se traduzisse numa robustez e eficiência significativa, penalizando ao mínimo a fluidez da interacção entre utilizador e a máquina. Esta necessidade traduziu-se na procura de algoritmos suficientemente leves, de um ponto de vista computacional, mas que ainda assim possibilitassem resultados que não comprometessem o seu funcionamento.

Neste contexto foi desenvolvida uma aplicação que permite a interacção entre um utilizador e o sistema computacional baseado apenas nos movimentos da cabeça, sem qualquer intervenção dos membros superiores.

Com o intuito de exemplificar o funcionamento desta aplicação, bem como no sentido de efectuar testes com utilizadores, foram desenvolvidos 4 algoritmos que combinam várias técnicas, das quais salientamos o Haar Cascade, a análise de espaços de cor com o intuito de identificar a cor da pele e o Optical Flow, que foram encapsulados num algoritmo geral de maneira a tornar a aplicação mais resistente à falha.

Posteriormente foram efectuados testes com utilizadores, bem como entrevistas com o recurso a questionários, que permitiram a obtenção de métricas referentes ao desempenho da aplicação implementada.

Assim, com este trabalho pretendeu-se implementar um sistema que respondesse às seguintes questões:

- Qual o melhor sistema de tracking ou conjunto de sistemas de tracking?

Nitidamente a combinação de várias técnicas de tracking e detecção são efectivamente uma solução superior à utilização de uma técnica apenas. A análise de determinadas características – áreas de interesse – e características faciais são técnicas sólidas para a solução do problema.

- Como é que o sistema possibilita ao utilizador um controlo intuitivo?

Tendo uma interface simples e uma eficiência e robustez elevada, isto implica que a aplicação deve possibilitar a realização de tarefas num tempo exequível.

- Como é que se pode aumentar a fiabilidade e robustez do sistema de reconhecimento e tracking facial?

Combinando varias técnicas que resistam às alterações de iluminação e possibilitam várias alternativas para diferentes tipos de impedimentos físicos.

A solução apresentada, isto é, a combinação de diferentes algoritmos, cada um com um conjunto de técnicas diferentes de tracking e detecção facial diferentes é uma opção válida, eficaz e robusta. A implementação do algoritmo baseado nos classificadores de Haar é bastante robusta pois é capaz de detectar uma face num curto espaço de tempo mesmo depois de perder o tracking.

Algumas das soluções testadas e não implementadas na aplicação final parecem promissoras, tais como o SURF features se conseguir obter uma optimização a nível de rapidez na execução do algoritmo. Notou-se que o algoritmo apesar de robusto não tem um desempenho a nível de tempo em tarefas minuciosas desejável comparativamente à aplicação Camera Mouse 2009. Este facto deve-se à necessidade de ter um algoritmo de pós processamento de estabilização do cursor do rato em alguns algoritmos da solução.

5.2. Sugestões para trabalhos futuros

Perante os resultados obtidos em função da implementação da proposta inicial apresentada sugere-se que se efectuem alguns melhoramentos. Estes melhoramentos advêm de sugestões e críticas na utilização da aplicação e podem servir de base a um trabalho futuro.

Um estudo da utilização da aplicação em pessoas com impedimentos físicos e o posterior desenvolvimento de um algoritmo de pós-processamento inteligente baseado em dados estatísticos em tempo real para uma movimentação do rato mais estável e orientada para as necessidades de quem utiliza a cabeça como meio de movimento.

Estabelecer uma ajuda visual baseada numa camada superior ao nível do sistema operativo de maneira a que o utilizador nunca perca a referência visual. Ou seja, ter uma referência visual da área de acção e do ponto de tracking na camada superior de interface.

Um outro trabalho futuro seria o upgrade da aplicação Mobility Camera com a futura versão do wrapper do OpenCV - Emgu que corresponda à versão do OpenCV 2.0 que foi disponibilizada a 30 de Setembro de 2009. Os melhoramentos presentes na nova versão do OpenCV poderiam melhorar a robustez e eficiência da aplicação. A inclusão de novas técnicas descritas por [17] e [18] ou do melhoramento de alguns algoritmos existentes poderiam ser uma solução provável.

Para analisar desenvolvimentos e melhoramentos na aplicação podem se dirigir ao site www.3dcgc.com para acompanhar a evolução da aplicação bem como vídeos ilustrativos da utilização da mesma. Esta informação está disponível na aplicação como se pode ver na Figura 35.

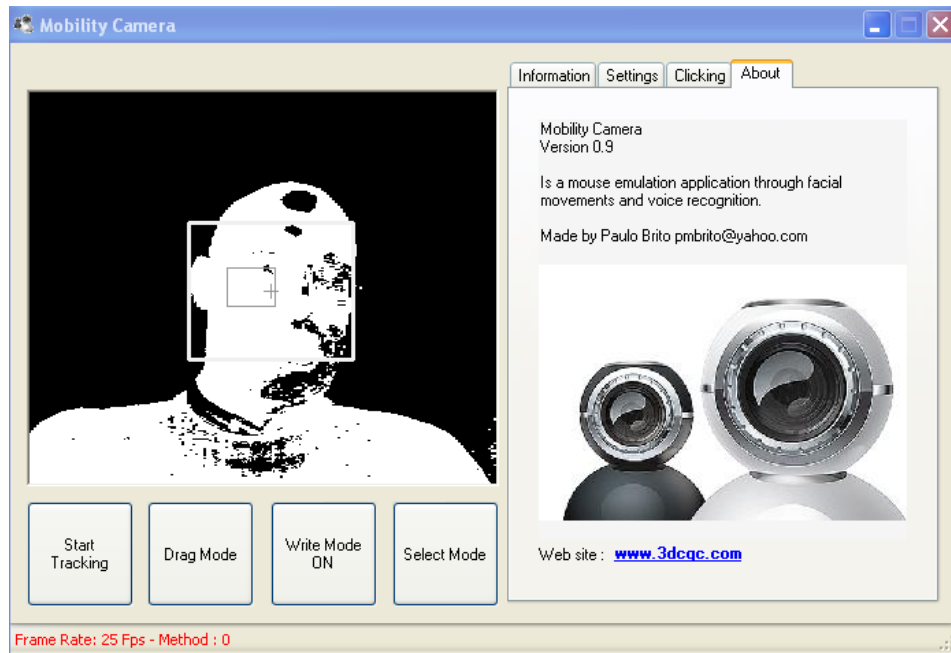


Figura 35 – Informação adicional da aplicação

Bibliografia

- 1 M. Betke. "Camera-based Interfaces and Assistive Software for People with Severe Motion Impairments." Department of Computer Science Technical Report BUCS-2008-025, Boston University. October 1, 2008
- 2. Hjelmås, E., Low, B.K.: Face detection: A survey. CVIU 83 (2001)
- 3. K. Toyama. Look, Ma – no hands! Hands-free cursor control with real-time 3D face tracking. In Proc. Workshop on Perceptual User Interfaces (PUI'98), San Fransisco, November 1998.
- 4. Yang, M., Kriegman, D., Ahuja, N.: Detecting faces in images: A survey. PAMI 24 (2002)
- 5. D. Gorodnichy. On importance of nose for face tracking. In Proc. Intern. Conf. on Automatic Face and Gesture Recognition (FG'2002), Washington, D.C., May 2002.
- 6. F. J. Huang and T. Chen, "Tracking of Multiple Faces for Human-Computer Interfaces and Virtual Environments", IEEE Intl. Conf. on Multimedia and Expo., New York, July 2000
- 7. R. Rao and R. Mersereau, "On Merging Hidden Markov Models with Deformable Templates," Proceedings of the Int'l Conf. on Image Processing, Washington D.C., 1995.
- 8. Advanced Multimedia Processing Lab - Carnegie Mellon University, <http://amp.ece.cmu.edu/projects/FaceTracking/>
- 9. Sanjay Kr. Singh, D. S. Chauhan, Mayank Vatsa, Richa Singh, "A Robust Skin Color Based Face Detection Algorithm" Tamkang Journal of Science and Engineering, Vol. 6, No. 4, pp. 227-234 (2003)
- 10. Venn Diagram Wikipedia, http://en.wikipedia.org/wiki/Venn_Diagram
- 11. Viola, P. and Jones, M. Rapid object detection using boosted cascade of simple features. IEEE Conference on Computer Vision and Pattern Recognition, (2001)
- 12. Wilson, P.I., Fernandez, J.: Facial feature detection using Haar classifiers. Journal of Computing Sciences in Colleges. Volume 21 , Issue 4, Pages: 127 - 133 (2006)
- 13. HORTA, T. C. de O. Um estudo das Transformadas Wavelets de Haar e Gabor para Detecção de Objectos. Dissertação (Mestrado) — Universidade de Brasília, 2007
- 14. PAPAGEORGIU, M. O. C.; POGGIO, T. A general framework for object detection. International Conference on Computer Vision, 1998

- 15. KUNCHEVA, L. I. *Combining Pattern Classifiers, Methods and Algorithms*. [S.l.]: Wiley-Interscience, 2004
- 16. SCHNEIDERMAN, H.; KANADE, T. A statistical method for 3d object detection applied to faces and cars. *International Conference on Computer Vision*, 2000
- 17. DELAC, K.; GFGIC, M. *Face Recognition*. [S.l.]: I-TECH Education and Publishing, 2007
- 18. COSTA, G. H. *Detecção Facial: Autofaces versus Antifaces*. Dissertação (Mestrado) -Universidade Federal de Santa Catarina, 2003
- 19. Radgui, A.[Amina], Demonceaux, C.[Cédric], Rziza, M.[Mohammed], Mouaddib, E.M.[El Mustapha], Aboutajdine, D.[Driss], An adapted Lucas-Kanade's method for optical flow estimation in catadioptric images, OMNIVIS08(xx-yy). 0810
- 20. M. Betke and N. C. Makris, "Recognition, resolution and complexity of objects subject to affine transformation," *Int. J. Comput. Vis.*, vol. 44, no. 1, pp. 5–40, Aug. 2001
- 21. Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, "SURF: Speeded Up Robust Features", *Computer Vision and Image Understanding (CVIU)*, Vol. 110, No. 3, pp. 346--359, 2008
- 22. *The Wide Range of Abilities and Its Impact on Computer Technology – Forrester Research Inc.*, 2003
- 23. Chellappa. R.; Wilson,C.L. & Sirohey,S: Human and machine recognition of faces: A survey. *Proceedings of the IEEE* 83, 5 (May 1995), 703-740
- 24. Kirby, M. & Sirovich, L. Application of the karhunen-loève procedure for the characterization of human faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 1, 12 (January 1990), 103–108
- 25. Turk, M. A. & Pentland, A. P. Face recognition using eigenfaces. In *Proc. Of the IEEE con Computer Society Conference*. IEEE Computer Society Press (1991)
- 26. P. N. Belhumeur, J.P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", *IEEE Tran. on PAMI*, Vol. 19, No. 7, pp. 711-720, July 1997
- 27. B.K.P. Horn and B.G. Schunck. *Determining optical flow*. AI Memo 572. Massachusetts Institue of Technology, 1980
- 28. FREEMAN, W. T. et al. Computer vision for interactive computer graphics. *IEEE Computer Graphics and Applications*, v. 18, n. 3, p. 42-53, 1998
- 29. *Redundant System Reference Design*, <http://www.ni.com/> (search RedundantReference)

- 30. B. D Lucas and T Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," in Proceedings of the International Joint Conference on Artificial Intelligence, April 1981, pp. 674–679
- 31. J. L Barron, D. J Fleet, and S. S Beauchemin, "Performance of Optical Flow Techniques," International Journal of Computer Vision, vol. 12, no. 1, pp. 43–77, 1994
- 32. J.-Y Bouguet, "Pyramidal Implementation of the Lucas- Kanade Feature Tracker: Description of the algorithm," Tech. Rep., OpenCV Document, Intel Microprocessor Research Labs, 2000
- 33. D. G Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004
- 34. Intel Corporation (2001): Open Source Computer Vision Library Reference Manual, 123456-001
- 35. Bradski , G. R. (1998): Computer vision face tracking for use in a perceptual user interface. Intel Technology Journal, 2nd Quarter, 1998
- 36. Allen, J.G., Xu, R.Y.D. and Jin, J.S. (2004). Object Tracking Using CamShift Algorithm and Multiple Quantized Feature Spaces. In Proc. 2003 Pan-Sydney Area Workshop on Visual Information Processing (VIP2003), Sydney, Australia. CRPIT, 36. Piccardi, M., Hintz, T., He, S., Huang, M. L. and Feng, D. D., Eds. ACS. 3-7
- 37. R. Brunelli, Template Matching Techniques in Computer Vision: Theory and Practice, Wiley, ISBN 978-0-470-51706-2, 2009
- 38. H. Y. Kim and S. A. Araújo, "Grayscale Template-Matching Invariant to Rotation, Scale, Translation, Brightness and Contrast," IEEE Pacific-Rim Symposium on Image and Video Technology, Lecture Notes in Computer Science, vol. 4872, pp. 100-113, 2007
- 39. EMGU Wrapper do OpenCV, <http://www.emgu.com>

6 Anexos

6.1. Templates dos questionários

| | | | | | | | |
|---|------|------------------------|---|---|---|---|------------|
| TESTE DE USABILIDADE "Tese de Mestrado" | | Mobility Camera | | | | | |
| Projecto: Mobility Camera | | Data: ___/___/___ | | | | | |
| Entidade: _____ | | Local: _____ | | | | | |
| Área: _____ | | Duração: _____ | | | | | |
| Moderadores: _____ | | Tester: _____ | | | | | |
| No âmbito da Tese de Mestrado Mobility Camera, este plano foi desenvolvido de forma a aferir a aplicação das regras de usabilidade da aplicação desenvolvida. | | | | | | | |
| Escala referencial: 1-Nada 2-Pouco 3-Razoável 4-Muito 5-Totalmente | | | | | | | |
| 1. Avaliação Global | | 1 | 2 | 3 | 4 | 5 | Totalmente |
| 1.1. Globalmente a aplicação apresentada está bem desenvolvida? | Nada | | | | | | |
| Comentários: | | | | | | | |
| 2. Facilidade de navegação | | 1 | 2 | 3 | 4 | 5 | Totalmente |
| 2.1. Conseguiu navegar na aplicação com facilidade? | Nada | | | | | | |
| Comentários: | | | | | | | |
| 2.2. Conseguiu executar as tarefas facilmente? | Nada | | | | | | |
| Comentários: | | | | | | | |
| 3. Simplicidade | | 1 | 2 | 3 | 4 | 5 | Totalmente |
| 3.1. A apresentação da interface da aplicação é simples? | Nada | | | | | | |
| Comentários: | | | | | | | |

Entidade: _____

| | | | | | | | |
|--|------|---|---|---|---|---|------------|
| Comentários: | | | | | | | |
| 4. Relevância do objectivo da aplicação | | 1 | 2 | 3 | 4 | 5 | Totalmente |
| 4.1. O objectivo da aplicação é útil e conciso? | Nada | | | | | | |
| Comentários: | | | | | | | |
| 4.2. Os objectivos estão de acordo com as necessidades do utilizador? | Nada | | | | | | |
| Comentários: | | | | | | | |
| 5. Manter a consistência | | 1 | 2 | 3 | 4 | 5 | Totalmente |
| 5.1. Os diferentes sistemas de controlo do rato apesar de diferentes são óbvios? | Nada | | | | | | |
| Comentários: | | | | | | | |
| 6. Tempo suportável | | 1 | 2 | 3 | 4 | 5 | Totalmente |
| 6.1. O tempo para executar uma tarefa é aceitável? | Nada | | | | | | |
| Comentários: | | | | | | | |
| 7. Foco no utilizador | | 1 | 2 | 3 | 4 | 5 | Totalmente |
| 7.1. A aplicação adapta-se facilmente ao ambiente do utilizador permitindo executar rapidamente uma nova tarefa? | Nada | | | | | | |
| Comentários: | | | | | | | |

Entidade: _____

| | |
|--|---------------------------------------|
| Aspectos positivos a salientar | Aspectos negativos a salientar |
| | |
| Sugestões de acções de melhoria | |
| | |

Entidade: _____

| TESTE DE USABILIDADE "Tese de Mestrado" | | Mobility Camera |
|---|-------------------|-----------------|
| Projecto: Mobility Camera | Data: ___/___/___ | |
| Entidade: _____ | Local: _____ | |
| Área: _____ | Duração: _____ | |
| Moderadores: _____ | Tester: _____ | |
| <p>No âmbito da Tese de Mestrado Mobility Camera, este plano foi desenvolvido de forma a aferir a aplicação das regras de usabilidade da aplicação desenvolvida.</p> <p>GUIÃO:</p> <p>- Apresentação</p> <ul style="list-style-type: none"> • Explicar a finalidade da aplicação • Mostrar como se manuseia o rato na aplicação – ver vídeo de ajuda • Deixar bem claro que é o protótipo que está a ser avaliado e não o utilizador <p>- Pedir para verbalizar os pensamentos e as acções durante a interacção – anotar essa informação em papel para posterior análise.</p> <p>- Pedir autorização para filmar, explicando a sua finalidade.</p> <p>- Ler cada tarefa antes de a executar em voz alta.</p> <p>TAREFAS:</p> <ol style="list-style-type: none"> 1. Iniciar o programa e ler informação de ajuda. 2. Iniciar o sistema de controlo do rato através dos movimentos da cabeça. 3. Ambientar-se durante algum tempo aos movimentos da cabeça que controlam o rato e clicar no botão do lado esquerdo e direito do rato. 4. Alterar sistema de controlo do rato (<i>tracking system</i>) e ambientar-se ao sistema mais uma vez. Repetir esta tarefa até voltar a imagem de início sem controlo do rato. 5. Escolher o sistema de controlo do rato que melhor funcione em relação ao ambiente de iluminação. 6. Clique no menu Iniciar (Start) -> Todos os programas -> Acessórios -> Acessibilidade -> Teclado de ecrã. Ou pode utilizar o <i>Write Mode</i> se quiser para escrever o texto. Este teclado permite escrever apenas recorrendo ao rato. 7. Abra um browser de internet escreva o seguinte endereço "www.3dcgc.com", navegue nas diferentes secções do site. Pode utilizar o <i>Write Mode</i> se quiser para escrever o texto. | | |

8. Abra a aplicação "Notepad" ou "Microsoft Word" e escreva o seguinte texto: " A aplicação Mobility Camera pode ser uma ajuda substancial para as pessoas que não conseguem por algum motivo ter controlo das mãos." Pode utilizar o *Write Mode* se quiser para escrever o texto.

| Notas: |
|---|
| <ul style="list-style-type: none"> • Onde sentiu mais dificuldades? • O que funcionou melhor? • Que outras tarefas seriam realistas? • Há tarefas que não faria aqui? • Há mais algo que consideraria importante contemplar? |

6.2. Imagens do percurso dos diferentes algoritmos

