

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
DEPARTAMENTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

BRUNO RODRIGUES FROZ

**INFERÊNCIA DAS EXPRESSÕES FACIAIS DE ALEGRIA E
TRISTEZA UTILIZANDO EMFACS E VÍDEO**

SÃO LUÍS
2013

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
DEPARTAMENTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

BRUNO RODRIGUES FROZ

**INFERÊNCIA DAS EXPRESSÕES FACIAIS DE ALEGRIA E
TRISTEZA UTILIZANDO EMFACS E VÍDEO**

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Aristófanés Corrêa Silva

SÃO LUÍS
2013

Froz, Bruno Rodrigues.

Inferência das expressões faciais de alegria e tristeza utilizando EMFACS e vídeos / Bruno Rodrigues Froz. – São Luís, 2013.

56 f.

Impresso por computador (fotocópia).

Orientador: Aristófanês Corrêa Silva.

Monografia (Graduação) – Universidade Federal do Maranhão, Curso de Ciência da Computação, 2013.

1. Visão computacional - EMFACS. 2. Facial Action Coding System - emoções. I. Título.

CDU 004

BRUNO RODRIGUES FROZ

**INFERÊNCIA DAS EXPRESSÕES FACIAIS DE ALEGRIA E
TRISTEZA UTILIZANDO EMFACS E VÍDEO**

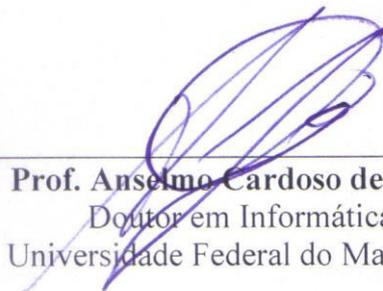
Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Aprovado em 01/02/2013

BANCA EXAMINADORA



Prof. Aristófanes Corrêa Silva (Orientador)
Doutor em Informática
Universidade Federal do Maranhão



Prof. Anselmo Cardoso de Paiva
Doutor em Informática
Universidade Federal do Maranhão



Prof. Geraldo Braz Junior
Mestre em Informática
Universidade Federal do Maranhão

AGRADECIMENTOS

Através deste trabalho, uma nova etapa da minha vida se inicia. Agradeço a toda a minha família por acreditar que eu posso vencer na vida e incentivar o meu crescimento pessoal e profissional. Agradeço ao meu pai Antônio e minha mãe Fátima, que me apoiaram incondicionalmente em todos os momentos e com certeza foram os atores principais em toda minha formação acadêmica.

Aos meus amigos, um agradecimento especial para os que começaram e terminaram essa jornada juntos, que foram o Luciano, o Júlio, o Paulo, o Pedro, o Tássio e a Suellen. Essa última foi parte importantíssima para a conclusão deste trabalho e agradeço pela paciência e boa vontade em ajudar.

Agradeço imensamente ao meu orientador Aristófanis, que ajudou sem hesitar em nenhum momento, e foi decisivo na conclusão deste trabalho.

Agradeço a todos que se envolveram de alguma forma para a minha vitória nesta batalha, que só está começando.

RESUMO

Este trabalho descreve um sistema baseado em uma metodologia que infere as emoções de Alegria e Tristeza utilizando vídeo em tempo real capturado por uma *webcam*. A inferência é feita através da utilização de um sistema de codificação chamado *Facial Action Coding System* (FACS), que classifica os movimentos musculares em Unidades de Ação, ou *Action Unit* (AU). Como cada AU representa uma deformação, a emoção é inferida através de uma derivação do FACS chamada *Emotion Facial Action Coding System* (EMFACS), que classifica uma emoção baseada na soma de uma ou mais AU's recorrentes. Os resultados encontrados através dos experimentos foram de uma média de acertos de 57,3% de inferências corretas para a emoção da Alegria e de 66,4% de inferências corretas para a emoção da Tristeza.

Palavras-chave: Emoção. Reconhecimento de emoções. Visão Computacional. FACS. EMFACS.

ABSTRACT

This work describes a system based on a methodology that infers the emotions of Joy and Sadness using real time video captured by a webcam. The inference is done through a codification system called Facial Action Coding System (FACS), which classifies muscle movements in Action Unit (AU). As each AU represents a deformity, the emotion is inferred through a derivation of FACS called Emotion Facial Action Coding System (EMFACS), which classifies the emotion based on the union of one or more AU's. The results from the experiments were an average of 57,3% of correct inferences to the Joy emotion and 66,4% of correct inferences to the Sadness emotion.

Keywords: Emotion. Emotion recognition. Computer Vision. FACS. EMFACS.

LISTA DE FIGURAS

Figura 2.1: Estrutura 3d resultante de algoritmos de reconstrução baseado em centenas de imagens 2d parcialmente sobrepostas. Fonte: snavely; szeliski, (2006).....	14
Figura 2.2: Usos no cotidiano da visão computacional. (a) rastreamento de carros em uma avenida. Fonte: potůček (2012?); (b) tomografia de um câncer de pâncreas. Fonte: matheus (2012?); (c) detecção de face em uma foto. Fonte: documentação do opencv (2001).....	15
Figura 2.3: Ilusão do xadrez. (a): o xadrez mostra dois quadrados rotulados de a e b. A ilusão formada pela iluminação e pela cor dos quadrados vizinhos, faz parecer que a e b tem cores diferentes. (b) a prova de que a e b tem a mesma cor. Fonte: adelson (1995).....	15
Figura 2.4: Formas mais comuns de características haar. Fonte: documentação do opencv (2001).....	17
Figura 2.5: Matrizes de uma imagem. (a) exemplo de uma imagem representada matriz das intensidades dos pixels; (b) exemplo de uma matriz da imagem integral obtida pela imagem do exemplo (a). Fonte: toscano (2011).....	18
Figura 2.6: FCP's e haralines. Fonte: jongh (2002).....	27
Figura 3.1: Etapas da metodologia proposta.....	28
Figura 3.2: Classificadores agindo em cascada. Fonte: campos (2011).....	30
Figura 3.3: Imagem em escala de cinza.....	31
Figura 3.4: Imagem em tom de cinza equalizada.....	31
Figura 3.5: Exemplo da detecção de face. (a) imagem original; (b) face detectada.....	32
Figura 3.6: Segmentação de elementos faciais. (a) imagem da face usada para análise; (b) elementos faciais segmentados.....	33
Figura 3.7: FCP's detalhadas. Fonte: oliveira; jaques (2008).....	34

Figura 3.8: Técnicas de <i>canny</i> (encontrar bordas), <i>erode</i> (erosão) e <i>dilate</i> (dilatação). (a) é a imagem da boca depois da técnica <i>canny</i> ; (b) é o efeito causado na imagem (a) após ter sido modificada pela sequência de métodos <i>erode</i> e depois <i>dilate</i>	35
Figura 3.9: Divisão da face em quatro partes	38
Figura 3.10: Fluxograma de inferência da alegria	44
Figura 4.1: FCP's e ponto origem desenhadas sobre a face do usuário. Os pontos azuis pequenos representam as fcp's das sobrancelhas e o ponto azul grande representa o ponto origem. Os pontos brancos representam as fcp's dos olhos. Os pontos verdes representam as fcp's da boca.....	46
Figura 4.2: Face do usuário em estado neutro. A barra amarela mostra que o estado está 100% neutro.....	47
Figura 4.3: Ocorrência da emoção da alegria na face do usuário. A barra verde demonstra que a alegria está acontecendo com a probabilidade de 100%	48
Figura 4.4: Ocorrência da emoção da tristeza na face do usuário. A barra vermelha mostra que a tristeza está ocorrendo com a probabilidade de 100%	48
Figura 4.5: Ocorrência das duas emoções simultaneamente. A barra verde está 100% cheia e a barra vermelha parcialmente cheia.....	49

LISTA DE QUADROS

Quadro 2.1: Sete expressões faciais básicas, ou emoções universais.....	19
Quadro 2.2: Lista das au's com base em movimentos musculares específicos.....	20
Quadro 2.3: Lista de au's sem base em movimentos musculares	24
Quadro 3.1: Combinação das AU's necessárias para inferir as emoções da Alegria e da Tristeza	43

LISTA DE TABELAS

Tabela 3.1: Equações de auxílio para encontrar a posição das coordenadas do eixo x das FCP's	36
Tabela 3.2: Deformações musculares da face	39
Tabela 3.3: Critérios de classificação das au's	40
Tabela 3.4: Emoção/au	41
Tabela 3.5: Pesos das au's e das combinações	43
Tabela 4.1: Experimentos com os usuários focados na alegria	51
Tabela 4.2: Experimentos com os usuários focados na tristeza	51

LISTA DE SIGLAS

AU	<i>Action Unit</i>
EMFACS	<i>Emotion Facial Action Coding System</i>
FACS	<i>Facial Action Coding System</i>
FCP	<i>Facial Characteristic Point</i>
IHC	Interface Humano-Computador
OpenCV	<i>Open Source Computer Vision Library</i>

SUMÁRIO

1 INTRODUÇÃO	11
2 FUNDAMENTAÇÃO TEÓRICA.....	13
2.1 VISÃO COMPUTACIONAL	13
2.2 OPENCV	16
2.3 CLASSIFICADORES HAAR.....	17
2.4 EMOÇÃO	19
2.5 FACS E EMFACS.....	20
3 METODOLOGIA.....	28
3.1 MODELO PROPOSTO	28
3.1.1 Detecção de face.....	29
3.1.2 Segmentação de elementos faciais	33
3.1.3 Detecção das FCP's.....	33
3.1.4 Classificação das AU's	38
3.1.5 Classificação da emoção.....	40
4 AVALIAÇÃO DOS RESULTADOS	45
4.1 FERRAMENTAS UTILIZADAS	45
4.2 ESTUDO DE CASO	45
4.3 ANÁLISE DOS RESULTADOS.....	50
4.3.1 Experimentos	50
5 CONCLUSÃO.....	53
REFERÊNCIAS	54

1 INTRODUÇÃO

Desde meados do século XX, o homem depende cada vez mais do computador e das operações computacionais para executar desde a mais simples até a mais complexa tarefa. Com o crescimento dessa dependência e com a evolução científica, fica cada vez mais claro que cada passo tomado na evolução científica influi diretamente nas emoções humanas, sendo para o lado bom ou para o lado ruim.

O ser humano tende a ser prejudicado de diversas formas no cotidiano devido às cargas emocionais. Formas que se manifestam no desempenho dele nas suas atividades diárias e nas relações físicas e emocionais com outros seres humanos.

Esses aspectos são estudados por um ramo científico chamado Ergonomia, que é o estudo das interações entre os seres humanos e elementos externos, ou interação entre ser humano e sistemas, onde se aplica teorias, princípios, dados e métodos a fim de aumentar o bem estar do ser humano e o desempenho individual do mesmo e desempenho global de sistemas (OLIVEIRA, 2010).

Essa área estuda também o comportamento do ser humano diante de um computador. Sua postura, seus movimentos, o tempo de exposição, entre outros fatores que devem ser considerados.

A preocupação com o envolvimento entre o ser humano e o computador pode ser estudada em uma área chamada Interface Humano-Computador, ou IHC, que é o estudo das interações entre o homem e o computador (BOOTH, 1989).

Computadores atuais estão ganhando a capacidade de interagir com as pessoas utilizando-se emoções como parâmetro. Eles podem tanto reconhecer uma emoção quanto reagir a elas. Esse estudo é chamado de Computação Afetiva (PICARD, 1995).

A Computação Afetiva estudada no contexto da IHC e da Ergonomia pode ser usada para o desenvolvimento de sistemas de reconhecimento de emoções, que podem ser usados de maneiras praticamente ilimitadas.

Com isso, utilizar as pessoas como objetos para reconhecimento de emoções e utilizar esse novo conhecimento como parâmetro para o desenvolvimento de sistemas de Computação Afetiva é uma maneira totalmente viável de propor melhorias nas condições humanas baseados nos estudos das áreas da Ergonomia, melhorando assim a qualidade de vida dos usuários do computador.

Sistemas de Computação Afetiva podem ser criados a partir dos conhecimentos de Visão Computacional e da extração de informações visuais que o computador capturar.

Com a captura de imagens e a extração de informações faciais, é possível reconhecer emoções baseadas em um sistema de codificação facial chamado *Facial Action Coding System* (FACS) (EKMAN; FRIESEN, 1978) e sua variação chamada *Emotion FACS* (EMFACS) (EKMAN; FRIESEN, 1983).

Esse trabalho tem por finalidade propor uma metodologia de reconhecimento das emoções da alegria e da tristeza em vídeo, utilizando as técnicas de Processamento de Imagem da Visão Computacional e EMFACS para a inferência mais aproximada da realidade possível, utilizando somente o rosto de uma pessoa em frente a uma *webcam* como parâmetro de entrada.

Nos próximos capítulos serão abordados os recursos necessários para a execução da metodologia proposta. O Capítulo 2 contém a fundamentação teórica, que explica sucintamente cada conhecimento utilizado para o desenvolvimento da metodologia. A metodologia é explicada no Capítulo 3, e cada etapa dela será descrita de forma detalhada, assim como sua execução demonstrada. No Capítulo 4 será mostrado um sistema desenvolvido baseado na metodologia proposta nesse trabalho, com a intenção de testar a metodologia e averiguar se a mesma pode ser utilizada para a inferência de emoções de forma consistente. E finalizando o trabalho, o Capítulo 5 contém uma conclusão acerca dos resultados obtidos e técnicas utilizadas para o desenvolvimento da metodologia, assim como sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Esse Capítulo descreve os conhecimentos prévios necessários para a execução deste trabalho. Será mostrado o conceito de Visão Computacional e sua importância. Depois será descrita a biblioteca OpenCV, em seguida, os classificadores Haar, que serão usados no sistema proposto. Também será tratado o conceito de emoção no contexto deste trabalho. Por último, será discutido os conceitos de FACS e EMFACS, assim como seus componentes auxiliares FCP.

2.1 Visão Computacional

No começo da década de 70, o conceito de Visão Computacional surgiu como um componente de percepção visual com o intuito de imitar a inteligência humana e embutir essa mesma em robôs. Os pioneiros da Visão Computacional acreditavam que resolver o problema de encontrar um parâmetro visual seria o passo mais fácil para resolver problemas de mais alto nível (SZELISKI, 2008).

O que distinguia a Visão Computacional de um campo que já existia chamado Processamento de Imagem Digital foi o desejo de recuperar a estrutura tridimensional do mundo a partir de imagens e usar isso como parâmetro para um entendimento mais generalizado. Winston, em 1975, Hanson e Riseman, em 1978, produziram duas coleções de artigos de muita importância para o contexto nessa época (SZELISKI, 2008).

Estudavam-se técnicas que almejavam, por exemplo, extrair bordas e depois inferir uma estrutura tridimensional de um objeto usando somente suas linhas bidimensionais (ROBERTS, 1965 apud SZELISKI, 2008).

Hoje nós temos técnicas que de forma confiável conseguem computar um modelo parcialmente 3D de um ambiente de centenas de imagens sobrepostas (SNAVELY; SZELISKI, 2006), como mostrada na Figura 2.1.



Figura 2.1: Estrutura 3D resultante de algoritmos de reconstrução baseado em centenas de imagens 2D parcialmente sobrepostas. Fonte: Snavely; Szeliski, (2006).

A dificuldade de estudar-se a visão é justamente devido a sua ideia de inversão de problemas. Nós devemos recuperar algumas incógnitas dadas informações insuficientes para assim especificar o problema, sendo necessária a utilização do recurso probabilístico e de modelos físicos para o objetivo ser concluído (SNAVELY; SZELISKI, 2006).

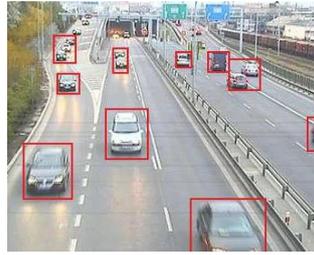
A Visão Computacional pode ser usada de maneira inteligente em praticamente todas as áreas de pesquisa. Alguns exemplos:

Rastreamento de carros: Podem-se usar técnicas de visão computacional para fazer um rastreamento de carros em vídeo (Figura 2.2a). Além disso, pode-se usa-la para reconhecimento de placas, cor, modelo de carro, entre diversas vertentes;

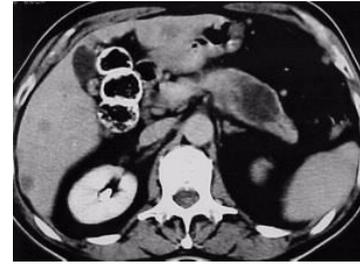
Imagem Médica: Imagens médicas hoje são usadas com diversas finalidades. Detecção de câncer, estudo cerebral, registro pre-operativo e pós-operativo, análises de exames laboratoriais como tomografia (Figura 2.2b), radiografia, etc.;

Detecção de objetos específicos: A detecção de certos objetos, pessoas, características das pessoas (Figura 2.2c), características do ambiente, entre outros, também podem ser usadas para fins distintos, como reconhecimento de padrões;

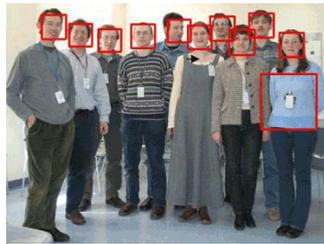
Sistemas de segurança: Pode ser usado vídeo para captura de movimento. Cada frame do vídeo é analisado, com o fim de detectar movimento em certo local. Isso pode ser usado como principal recurso de um sistema de segurança. Sensores de pessoas e objetos podem ser usados como sistemas de segurança de veículos, detectando pedestres e obstáculos nas estradas para evitar acidentes.



(a)



(b)

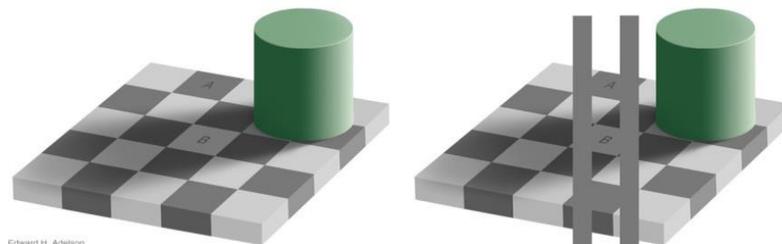


(c)

Figura 2.2: Usos no cotidiano da Visão Computacional. (a) Rastreamento de carros em uma avenida. Fonte: Potúček (2012?); (b) Tomografia de um câncer de pâncreas. Fonte: Matheus (2012?); (c) Detecção de face em uma foto. Fonte: Documentação do OpenCV (2001).

Os modelos utilizados em Visão Computacional geralmente são desenvolvidos na física, na matemática e na computação gráfica. Esses campos permitem que estudos como o da iluminação, de extrema importância na Visão Computacional, sejam utilizados da forma mais coerente possível para resolver os problemas inerentes a essa área.

O problema do Xadrez na Figura 2.3 mostra que a visão do ser humano pode ser limitada por diversos tipos de truques que causam ilusões de ótica. Não é um problema óptico do ser humano, mas sim, um problema referente à mente (MOTOKAWA, 1950).



(a)

(b)

Figura 2.3: Ilusão do Xadrez. (a): O Xadrez mostra dois quadrados rotulados de A e B. A ilusão formada pela iluminação e pela cor dos quadrados vizinhos, faz parecer que A e B tem cores diferentes. (b) A prova de que A e B tem a mesma cor. Fonte: Adelson (1995).

Um sistema de Visão Computacional consegue passar por cima de um problema de ilusão ótica humana, já que apesar de tentar imitar a visão humana, não tem a complexa rede neurológica dos seres vivos.

Apesar dessa minúscula vantagem sobre o cérebro humano, a Visão Computacional deve evoluir muito para se chegar perto de ser comparada com o mesmo. O limite do poder de processamento do cérebro humano e animal até hoje é desconhecido.

2.2 OpenCV

OpenCV – *Open Source Computer Vision Library* – é uma biblioteca *open source* que inclui centenas de algoritmos de visão computacional. Atualmente sua versão mais atual é a 2.4.3. Originalmente feito para as linguagens C/C++, o OpenCV pode ser usado com Python e Java também (INTEL, 2001).

As plataformas que suportam vão desde os desktops (Windows, Linux, MacOS, etc.) e até as mobiles (Android, Meamo e iOS).

O OpenCV tem vários módulos distintos, onde cada um contém vários métodos estáticos ou compartilhados. Os principais módulos são:

core: contém algumas estruturas básicas, incluindo o array multidimensional Mat;

imgproc: modulo de processamento de imagem que contém filtros lineares e não lineares, transformações geométricas, conversão de cor, histogramas, etc.;

video: módulo de análise de vídeo, que contém funções de estimação de movimento, subtração de background e algoritmos de rastreamento de objetos;

calib3d: contém elementos de reconstrução 3D, calibração de câmera, estimação de posição de objeto, etc.;

features2d: contém detector de características, descritores, etc.;

highgui: interface de captura de vídeo e de imagens;

gpu: contém algoritmos de aceleração de GPU;

A biblioteca OpenCV é de extrema importância para este trabalho, considerando que foi usado em todo o sistema proposto, o que facilitou bastante o desenvolvimento do mesmo.

2.3 Classificadores Haar

Os classificadores Haar foram desenvolvidos por Viola e Jones (2001) e são usados para detectar faces em uma imagem de forma rápida e precisa. Eles podem ser adaptados para reconhecimento de características faciais (WILSON, 2006).

Esse método têm como base para a detecção de objetos as características Haar. Essas características usam a variação de contrastes entre grupos retangulares de pixels, ao invés das intensidades de cada pixel separadamente. A Figura 2.4 mostra as formas de características Haar utilizadas normalmente.

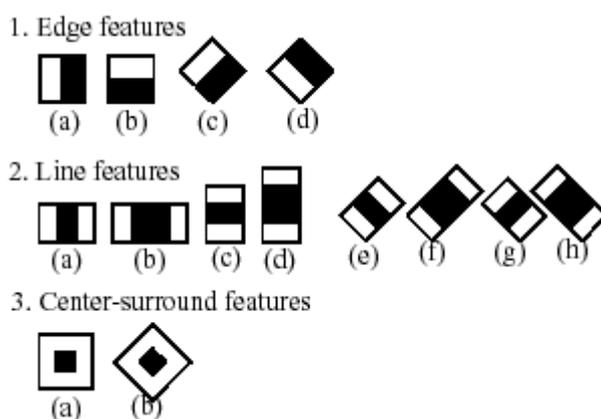


Figura 2.4: Formas mais comuns de características Haar. Fonte: Documentação do OpenCV (2001).

Essa variação de contrastes é usada para determinar áreas claras e escuras da imagem. Um pequeno grupo de dois ou três grupos adjacentes forma uma característica Haar.

Essas características podem variar a escala trabalhada somente aumentando ou diminuindo o tamanho dos grupos de pixels analisados, permitindo que objetos de diversos tamanhos sejam detectados (WILSON, 2006).

Para fazer a detecção dessas áreas retangulares, é utilizada uma forma intermediária e auxiliar de representação da imagem, chamada Imagem Integral (VIOLA; JONES, 2001) que consiste em uma matriz que contém a soma de todos os pixels à esquerda e acima de cada pixel. A Figura 2.5 mostra um exemplo de uma imagem sendo representada pelas intensidades de seus pixels (a) e uma imagem representando a imagem integral da mesma (b).

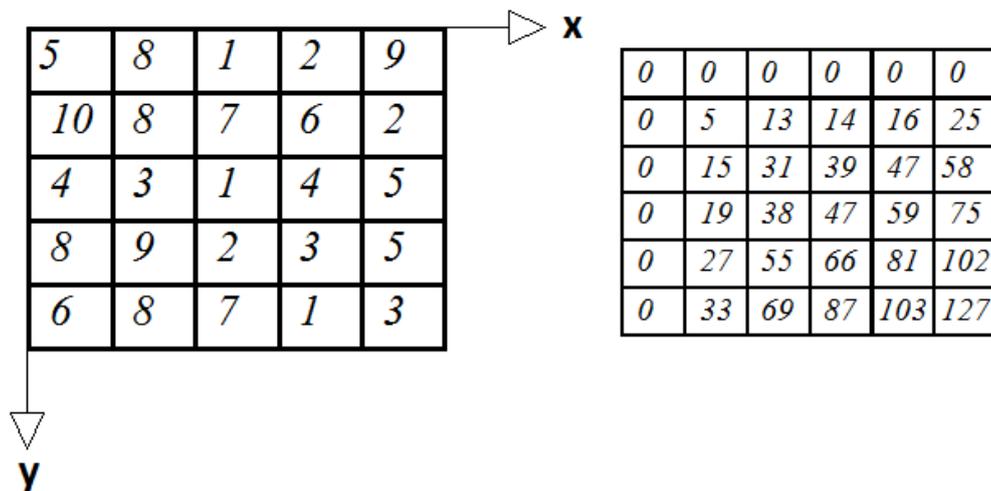


Figura 2.5: Matrizes de uma imagem. (a) Exemplo de uma imagem representada matriz das intensidades dos pixels; (b) Exemplo de uma matriz da imagem integral obtida pela imagem do exemplo (a). Fonte: Toscano (2011).

Considerando a matriz de intensidades de pixels da imagem original sendo “A” e a matriz da imagem integral sendo “I”, o valor de cada elemento de “I” é obtido através da Equação 2.1.

$$I[x,y] = \sum_{x' \leq x, y' \leq y} A(x',y') \quad (2.1)$$

Para características com o ângulo de 45°, como mostrado previamente nos itens 2(e), 2(f), 2(g), 2(h) e 3(b) da Figura 2.4, a equação de cálculo do valor dos elementos de “I” é a Equação 2.2.

$$I[x,y] = \sum_{x' \leq x, x' \leq x - |y-y'|} A(x',y') \quad (2.2)$$

A justificativa para essa mudança é que cada elemento de “I” é calculado através da soma dos pixels dos elementos às diagonais (45°) do elemento atual, e somente os valores à esquerda dele (LIENHART; MAYDT, 2002).

Para poder iniciar o cálculo em ambas as equações, é necessário adicionar uma linha e uma coluna no início da matriz, com o valor 0. Em outras palavras, se a matriz da imagem é NxN, a matriz da imagem integral é N+1xN+1, com a linha e a coluna inicial com todos os elementos com o valor 0.

Utilizando as imagens integrais em cada característica fica bem mais fácil de ser calculada a variação de intensidades de sub matrizes de uma imagem. Assim, usando a imagem integral e calculando a diferença entre algumas regiões da matriz, é possível achar

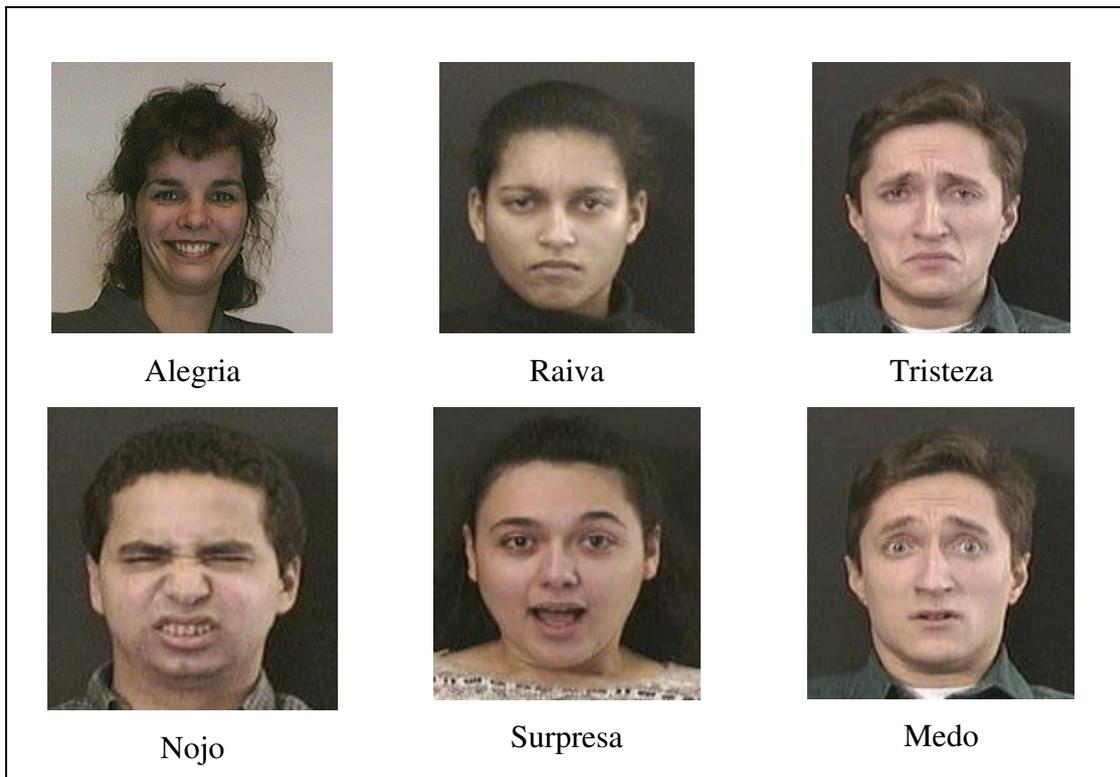
características de qualquer escala. Além disso, esse método é rápido e eficiente. Porém, calcular todas as características contidas em uma imagem não é nada prático (VIOLA; JONES, 2001).

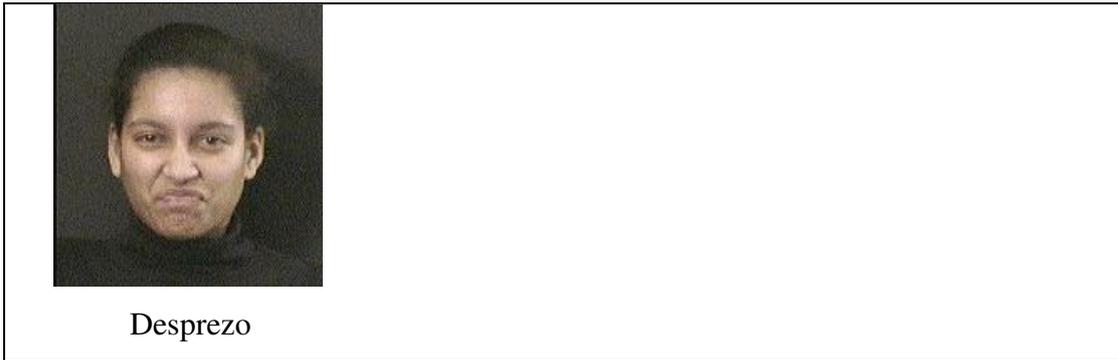
2.4 Emoção

Desde a publicação do famoso livro de Charles Darwin, *A Origem das Espécies*, sugerido pelo mesmo, era possível existirem emoções comuns a qualquer cultura. Alguns pesquisadores como Paul Ekman e Wallace Friesen (1983 apud RUSSEL, 1994) e Carroll Izard (1992) conduziram os primeiros estudos metodológicos sobre o assunto e concluíram que existem pelo menos seis emoções universais. São elas a Alegria, a Raiva, a Tristeza, a Surpresa, o Nojo e o Medo. Alguns pesquisadores concluem que o Desprezo também é uma emoção universal (RUSSEL, 1994).

Uma demonstração dessas sete emoções básicas está no Quadro 2.1.

Quadro 2.1: Sete expressões faciais básicas, ou emoções universais.





Fonte: Jongh (2002).

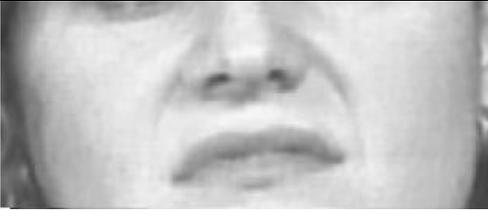
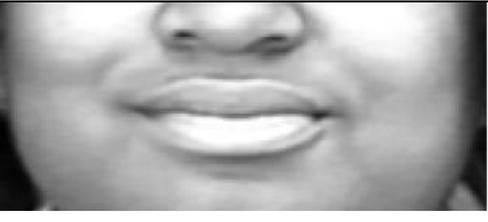
2.5 FACS e EMFACS

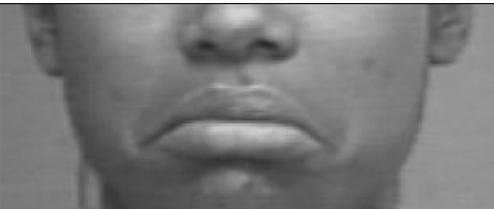
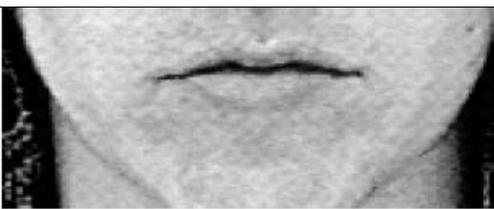
Facial Action Coding System (FACS) é um sistema baseado em observações humanas, desenvolvido para detectar mudanças súbitas em características faciais (EKMAN; FRIESEN, 2005). Cada movimento muscular é rotulado. Esses rótulos se chamam *Action Units* (AU), que correspondem a movimentos musculares simples ou um conjunto de movimentos musculares simultâneos.

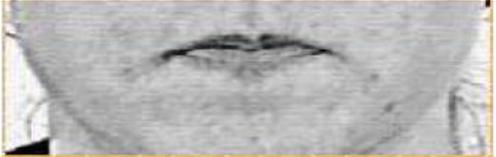
Existem 44 AU's na codificação do FACS. Dessas, 30 são relacionadas diretamente com contrações de um ou mais músculos combinados. As outras 14 não são relacionadas com os músculos, mas com movimentos que independem da posição dos mesmos. O Quadro 2.2 mostra as 30 AU's relacionadas a movimentos musculares específicos e o Quadro 2.3 mostra os 14 que são indiretamente relacionados aos movimentos.

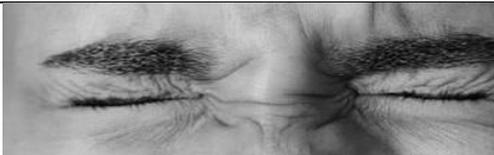
Quadro 2.2: Lista das AU's com base em movimentos musculares específicos.

AU	DESCRIÇÃO	EXEMPLO
1	Sobrancelha interna levantada	
2	Sobrancelha externa levantada	

4	Sobrancelhas abaixadas e juntas	
5	Pálpebra superior levantada	
6	Bochechas levantadas	
7	Pálpebras semi-fechadas	
9	Músculo labial superior e músculo nasal levantados	
10	Lábio superior levantado	
11	Músculo nasolabial aprofundado	
12	Cantos do lábio puxados	

13	Bochechas afofadas	
14	Bochecha pressionada contra os molares	
15	Cantos do lábio abaixados	
16	Lábio inferior abaixado	
17	Queixo levantado	
18	Lábio empurrados (movimento do “Beijo”)	
20	Lábio esticados	
22	Lábio afunilado	

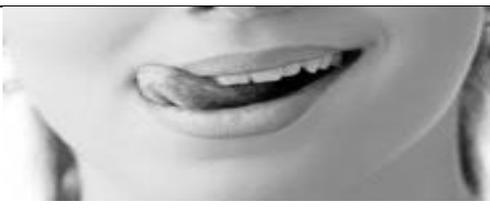
23	Lábio reduzido	
24	Lábio pressionado	
25	Boca relaxada e semi-aberta	
26	Queixo caído	
27	Boca aberta	
28	Lábio sugado	
41	Pálpebras caindo	
42	Pálpebras levemente fechadas	

43	Olhos fechados	
44	Estrabismo	
45	Piscar dois olhos	
46	Piscar um olho	

Fonte: Adaptado de Oliveira; Jaques, (2008).

Quadro 2.3: Lista de AU's sem base em movimentos musculares.

AU	DESCRIÇÃO	EXEMPLO
8	Lábios em direção um ao outro	
19	Língua pra fora	
21	Pescoço apertado	
29	Mandíbula elevada	

30	Mandíbula de lado	
31	Mandíbulas fechadas e boca aberta	
32	Lábio mordido	
33	Assoprar	
34	Inflar bochechas	
35	Bochechas mordidas	
36	Língua pra fora e espessa	
37	Lamber lábio	

38	Narina dilatada	
39	Narina comprimida	

Fonte: Adaptado de Oliveira; Jaques (2008).

Paul Ekman propôs que a combinação das AU's do FACS representavam protótipos de expressões de emoção (EKMAN; FRIESEN; HAGER, 2002). Essas expressões faciais não são rotuladas dentro do FACS, mas em um sistema separado chamado *Emotion Facial Action Coding System* (EMFACS) (EKMAN; FRIESEN, 1983). FACS é puramente descritivo. Não contém nenhum tipo de inferência pra emoções. Mas ao converter os rótulos dos movimentos musculares no sistema EMFACS, é possível inferir expressões de emoções básicas, como alegria, tristeza, raiva, surpresa, etc (EKMAN; ROSENBERG, 2005).

Como exemplo, ao detectar a ocorrência das AU's 12 e 15, a emoção detectada é a de “embaraçamento”. A AU 12 sozinha demonstraria a emoção da alegria, pois é equivalente ao sorriso, levando o lábio a ser esticado pelo movimento do levantamento da bochecha. Unindo a AU 15, que é mostra o canto do lábio levemente abaixado, modifica completamente a interpretação da emoção da alegria.

Outro exemplo é a ocorrência da AU 15, mostrando a depressão dos cantos do lábio juntamente com a AU 1, que mostra o levantamento da parte interna da sobrancelha e AU 4, que junta as sobrancelhas. Essa combinação demonstra a tristeza normalmente.

A detecção das AU's não é uma tarefa fácil. Kobayashi e Hara (1991) propuseram um modelo facial que mapeia a face em 30 pontos específicos. Esses pontos são chamados de *Facial Characteristic Points* (FCP). Essas 30 FCP's representam as 30 AU's que são movimentos musculares que fazem parte das emoções básicas.

As FCP's são relacionadas com contornos dos olhos, sobrancelhas e boca. Devido a experimentos, Kobayashi e Hara escolheram esses pontos, pois observaram que as pessoas só prestam atenção para a posição e tamanho dos olhos, sobrancelhas e boca para inferir emoções.

A Figura 2.6 representa as 30 FCP's da face. As linhas verticais se chamam *Haralines*, que tem suas posições fixas e dependem da posição das FCP's a1, a2, a3, a4, a19 e a20. Elas

servem para fixar as coordenadas do eixo X das FCP's relacionadas (KOBAYASHI; HARA; TANGE, 1995).

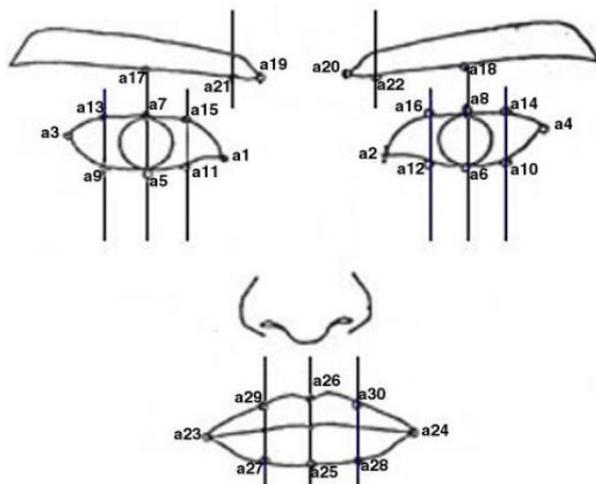


Figura 2.6: FCP's e Haralines. Fonte: Jongh (2002).
As técnicas envolvidas na identificação das FCP's serão discutidas no Capítulo 3.

3 METODOLOGIA

Este Capítulo mostra os passos necessários para a execução da metodologia proposta. Começamos demonstrando uma modelagem da metodologia de forma generalizada, depois cada etapa será descrita. Por fim, será mostrada cada etapa em execução, juntando todo o conhecimento reunido no Capítulo 2, utilizando as funções de Visão Computacionais estudadas.

3.1 Modelo proposto

Esta metodologia é dividida em cinco etapas principais, como mostrado na Figura 3.1.

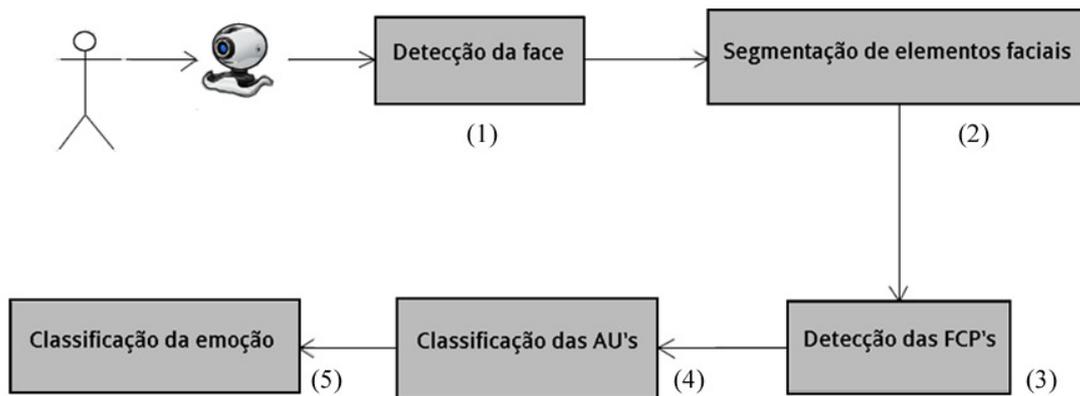


Figura 3.1: Etapas da metodologia proposta.

Através de uma *webcam*, o vídeo a ser analisado é capturado. O processo da metodologia é feito em cada frame.

O objetivo da primeira etapa é segmentar a face do sujeito analisado. O motivo dessa segmentação é focar os recursos do sistema para área específica da face, que será a única parte analisada em todo o frame.

Após essa etapa, uma nova segmentação acontece, mas somente na face. São identificadas as posições da boca, dos olhos e das sobrancelhas.

Depois cada FCP é identificada, para ser possível a classificação das AU's, através da identificação das deformações faciais.

Ao classificar as deformações, é possível classificar uma emoção. Para a classificação da mesma, é necessário fazer uma comparação das deformações do rosto no frame atual do vídeo com um frame posterior que representa a neutralidade do rosto, que será obtido manualmente.

As técnicas de cada etapa serão descritas na próxima seção.

3.1.1 Detecção de face

Através da *webcam* o vídeo é capturado e, em tempo real, as etapas começam a ser executadas para cada *frame*.

O primeiro objetivo é encontrar a face do usuário do sistema. Para isso, foram utilizados recursos da biblioteca OpenCV.

Para isso, usar Classificadores Haar é a escolha ideal (VIOLA; JONES, 2001). Somente algumas das Características Haar são necessárias pra determinar de uma região da imagem contém o objeto procurado (MENEZES; BARRETO, 2004). O objetivo do classificador é justamente selecionar as características que serão utilizadas na detecção, descartando as características com falsos positivos.

O algoritmo AdaBoost, de aprendizado de máquina, é utilizado em uma forma variante para determinar quais as características importantes para a detecção e suas limiares (FREUND; SCHAPIRE, 1995).

O algoritmo no fim consiste em, a cada etapa, utilizar um classificador variante do AdaBoost para decidir se um característica é uma candidata a ser parte do objeto procurado, ou refutar a mesma. As características passam por vários classificadores, onde a cada ciclo, utiliza as características refutadas como aprendizado, para ajudar a decidir se uma característica candidata anteriormente ainda é uma candidata ou é refutada. A Figura 3.2 mostra como esses classificadores agem. O término e decisão das características dependem muito da forma que o algoritmo foi desenvolvido.

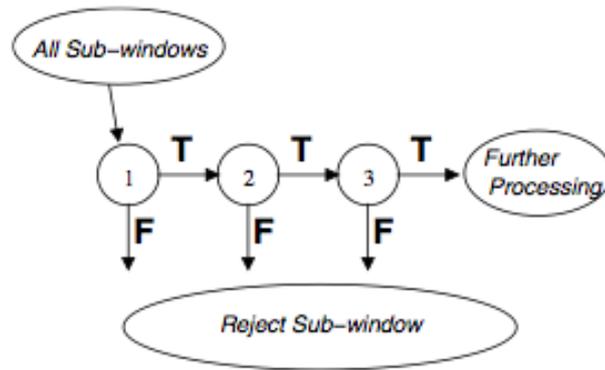


Figura 3.2: Classificadores agindo em cascada. Fonte: Campos (2011).

A biblioteca utilizada neste trabalho, o OpenCV, já possui uma implementação desse classificador de Haar, sendo que somente devem ser ajustados seus parâmetros.

Para treinar os classificadores, é necessário dois conjuntos de imagens. As imagens negativas, que são imagens que não contém o objeto desejado a ser detectado, e as imagens positivas, que contém o objeto procurado em uma ou mais localizações. As imagens positivas devem ter seu nome, pixel superior esquerdo e pixel inferior direito do objeto especificados, como citado no manual do OpenCV (2001).

Quanto mais imagens são utilizadas, menor a probabilidade de que falsos positivos sejam encontrados. Geralmente, essa quantidade ultrapassa os milhares, para cada grupo positivo e negativo. Existem classificadores já treinados na documentação do OpenCV para cada tipo de características mais comuns a serem procuradas, como por exemplo, faces, olhos, bocas e narizes.

O uso recomendado é sobre uma imagem com tons de cinza. Então fazemos primeiramente uma conversão das cores do frame analisado para a escala de cinza. A Figura 3.3 mostra um frame de um vídeo na escala de cinza.



Figura 3.3: Imagem em escala de cinza.

Após essa conversão, aplica-se um equalizador de Histograma, com função já disponível no OpenCV, para aguçar o contraste da imagem e facilitar a futura detecção de face. A Figura 3.4 mostra um dos frames com tom de cinza depois da equalização.



Figura 3.4: Imagem em tom de cinza equalizada.

No OpenCV existe uma classe correspondentes aos Classificadores Haar, como mostra o manual do OpenCV (2001). Para a classe ser utilizada é necessário treinar o objeto utilizando seu método *load*. Para o treino são necessários arquivos XML que contém tanto as imagens positivas e negativas já treinadas. Alguns exemplos desses arquivos já estão presentes na ferramenta desde a instalação e o arquivo referente à face é utilizado.

Após treinar o objeto com o arquivo XML referente à face, é utilizado o método *detectMultiScale*, também presente na classe, passando como parâmetro a imagem anteriormente equalizada para a detecção de possíveis faces. Essas ocorrências serão colocadas em um vetor com todas as posições de todas as faces encontradas (podem ser detectadas mais de uma face, incluindo os falsos positivos).

Como o sistema foi construído com base na informação de que somente uma face será encontrada, na prática esse vetor de faces retornará somente um resultado, que é a face do usuário detectada.

A Figura 3.5 mostra o resultado esperado na detecção da face.

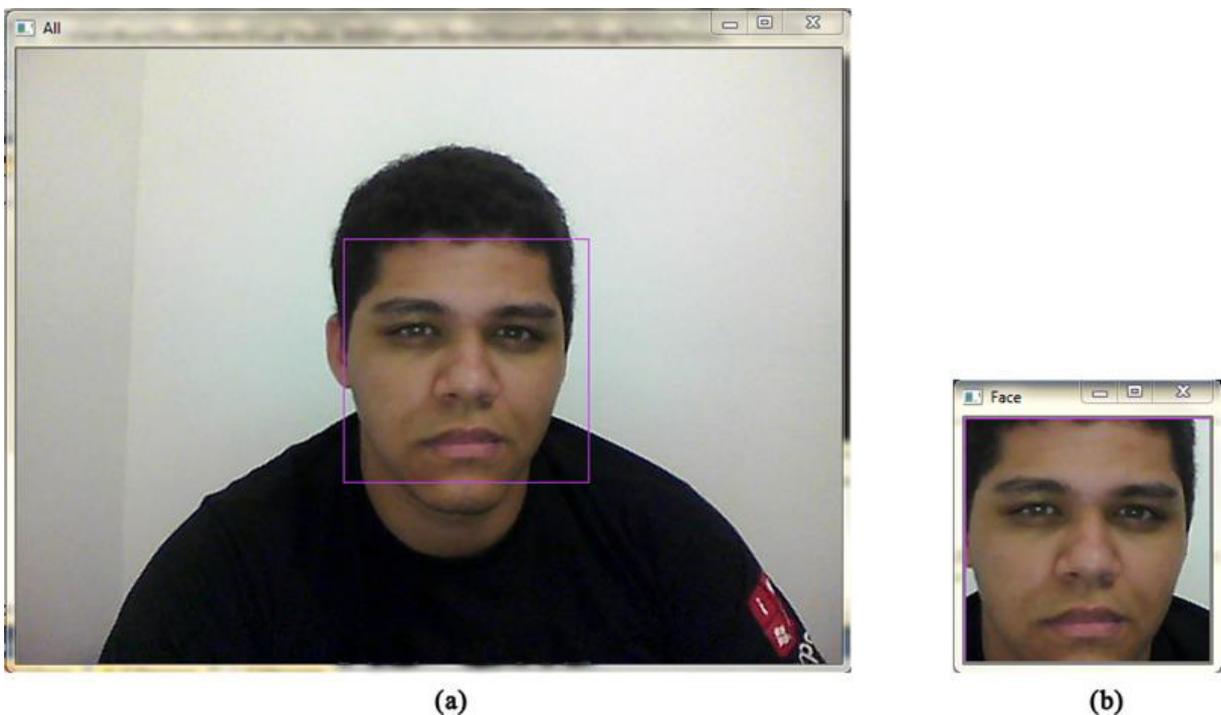


Figura 3.5: Exemplo da detecção de face. (a) Imagem original; (b) Face detectada.

3.1.2 Segmentação de elementos faciais

Nessa etapa, analogamente à detecção da face, a segmentação dos elementos faciais necessários é feita da mesma forma, utilizando o Classificador Haar. A diferença é que novos objetos serão treinados usando arquivos XML diferentes, voltados para cada elemento procurado.

Os elementos que o sistema busca é a boca, os olhos e a sobrancelha. Cada objeto é treinado usando um arquivo XML correspondente a cada elemento.

Mais uma diferença nessa etapa é que a imagem usada para fazer essa busca será o retângulo que representa a região da face, que foi descoberta na etapa anterior. O motivo disso é pra reduzir os falsos positivos que podem ser encontrados.

O método *detectMultiScale* é utilizado novamente, mas em cada objeto treinado, e como saída terá um vetor com todas as posições de cada objeto encontrado.

A Figura 3.6 mostra algumas das características faciais segmentadas.

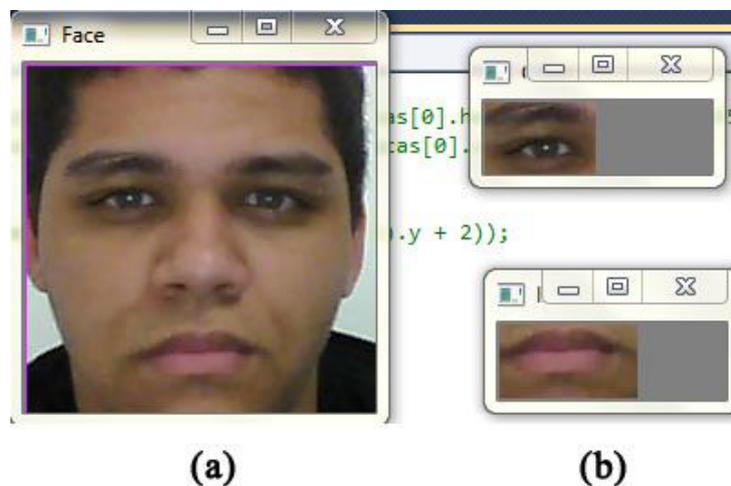


Figura 3.6: Segmentação de elementos faciais. (a) Imagem da face usada para análise; (b) Elementos faciais segmentados.

3.1.3 Detecção das FCP's

As FCP's são pontos específicos do rosto, definidos por Kobayashi e Hara (1991), observados no rosto humano como sendo pontos importantes para a identificação de movimentos musculares únicos.

A Figura 2.6, no Capítulo 2, mostra a posição de cada ponto definido e suas respectivas *Haralines*.

A metodologia propõe a identificação automática desses pontos, o que não é uma tarefa simples.

Inicialmente, cada parte segmentada da face – boca, olhos e sobrancelhas – passam pelo mesmo processo para a identificação das FCP's, porém, algumas FCP's devem ser identificadas antes para as regras serem aplicadas. A Figura 3.7 mostra as regras necessárias para a identificação das FCP's.

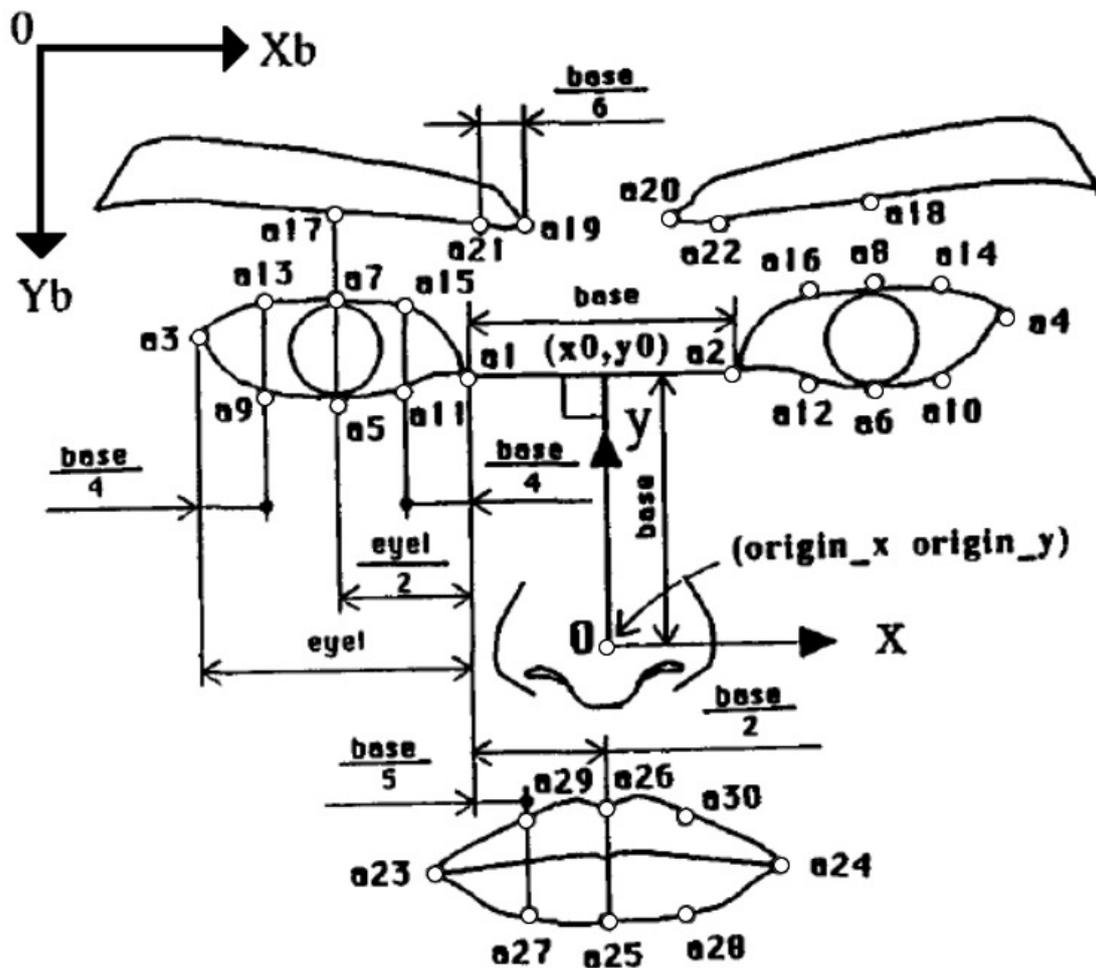


Figura 3.7: FCP's detalhadas. Fonte: Oliveira; Jaques (2008).

O valor *base* é calculado pela distância entre a FCP a1 e a FCP a2 através da Equação 3.1.

$$base = \sqrt{(a2.x - a1.x)^2} + \sqrt{(a2.y - a1.y)^2} \quad (3.1)$$

O valor da base é de muita importância, pois ele servirá principalmente para fazer uma correção de escala das coordenadas dos FCP's, algo que será discutido ainda nesse tópico. Além disso, será necessário esse valor para encontrar algumas FCP's. Então é preciso primeiramente encontrar esse valor.

Para isso, é necessário encontrar as FCP's a1 e a2. Essas FCP's podem ser encontradas usando as imagens segmentadas dos olhos direito e esquerdo.

A metodologia proposta utiliza diversas técnicas já desenvolvidas na biblioteca OpenCV para encontrar esses pontos. A detecção de bordas *Canny* encontra todas as bordas da imagem analisada automaticamente. Para a correção de ruídos na imagem resultante da detecção de bordas, é usada uma junção das técnicas *erode* e *dilate*. A função *erode* reduz todas as bordas dos grupos de pixels com valor igual a 1 em um valor de intensidade arbitrário, e a função *dilate* faz o oposto, aumenta as bordas desses grupos em um valor de intensidade também arbitrário. O efeito que causa na imagem é o apresentado na Figura 3.8.



Figura 3.8: Técnicas de Canny (Encontrar Bordas), Erode (Erosão) e Dilate (Dilatação). (a) É a imagem da boca depois da técnica *Canny*; (b) É o efeito causado na imagem (a) após ter sido modificada pela sequência de métodos *Erode* e depois *Dilate*.

Ao utilizar essas duas técnicas em sequência – *erode* e depois *dilate* – as bordas ficam mais consistentes, e os grupos de pixels que eram ruídos somem depois que a técnica *erode* é executada.

A FCP a1 representa o ponto extremo a esquerda do olho direito e a FCP a2 representa o ponto extremo a direita do olho esquerdo. Para identificá-los é só utilizar a imagem resultante das técnicas anteriores de detecção de borda e procurar os pontos extremos das imagens, percorrendo os pixels da mesma. Com isso é possível analogamente descobrir os pontos a3 e a4, que são os extremos opostos.

As mesmas técnicas são usadas para descobrir os pontos a23 e a24, que são os extremos da boca, e os pontos a19 e a20, que são as pontas internas das sobrancelhas.

Algumas FCP's são encontradas com o auxílio das *Haralines* (KOBAYASHI; HARA, 1993), que são linhas verticais que determinam a posição das FCP's que são dependentes umas das outras. As FCP's a5 e a7, por exemplo, tem uma *Haraline* entre elas, que é localizada exatamente no centro do olho direito. A FCP a5 é o ponto que a *Haraline* passa na parte inferior da borda do olho direito, e a FCP a7 é o ponto que essa mesma *Haraline* passa na parte superior da borda do olho direito. Os pontos a6 e a8 são encontrados da mesma forma, porém no olho esquerdo.

Os pontos restantes são encontrados com o uso da *base*, como citado anteriormente. A Tabela 3.1 mostra uma forma resumida de como encontrar o valor da coordenada x das FCP's restantes.

Tabela 3.1: Equações de auxílio para encontrar a posição das coordenadas do eixo x das FCP's.

FCP	Coordenada X
a9	$a3.x + \frac{base}{4}$
a10	$a4.x - \frac{base}{4}$
a11	$a1.x - \frac{base}{4}$
a12	$a2.x + \frac{base}{4}$
a13	$a3.x + \frac{base}{4}$
a14	$a4.x - \frac{base}{4}$
a15	$a1.x - \frac{base}{4}$
a16	$a2.x + \frac{base}{4}$
a17	a5.x ou a7.x
a18	a6.x ou a8.x
a21	$a19.x - \frac{base}{6}$
a22	$a28 + \frac{base}{6}$
a25	$a1.x + \frac{base}{2}$
a26	$a1.x + \frac{base}{2}$

a27	$a1.x + \frac{base}{5}$
a28	$a2.x - \frac{base}{5}$
a29	$a1.x + \frac{base}{5}$
a30	$a2.x - \frac{base}{5}$

As coordenadas x das FCP's listadas devem ser encontradas com os cálculos mostrados na Tabela 3.1, pelas regras de Kobayashi e Hara (1991) Para encontrar as coordenadas y desses pontos, o processo usado foi semelhante aos anteriores. Foi utilizados os mesmos recursos da biblioteca OpenCV, mas limitando as áreas de procura com o auxílio das coordenadas x já encontradas.

Um ponto importante a ser descoberto é o ponto de origem, que será utilizado para facilitar o uso das FCP's para cálculos no futuro. As coordenadas desse ponto (X_o, Y_o) são descobertas através da Equação 3.2 e, na prática, fica aproximadamente na ponta do nariz.

$$X_o = \frac{(a1.x + a2.x)}{2} \quad | \quad Y_o = \frac{(a1.y + a2.y)}{2} + base, \quad (3.2)$$

onde (X_o, Y_o) são as coordenadas da origem.

Com todas as coordenadas das FCP's e do ponto de origem descobertas, os pontos das FCP's serão transladados em relação à origem. A explicação pra isso é que quando os pontos forem trasladados, cada FCP terá um valor que caracterizará a sua posição na imagem, pois a face estará dividida em quadrantes. Por exemplo, uma FCP com a coordenada $x > 0$ e $y > 0$ representa o primeiro quadrante, onde os valores de x e y são sempre positivos. As únicas FCP's que estão nesse quadrante são as do olho esquerdo e sobranalha esquerda.

A Figura 3.9 mostra a divisão desses quadrantes.

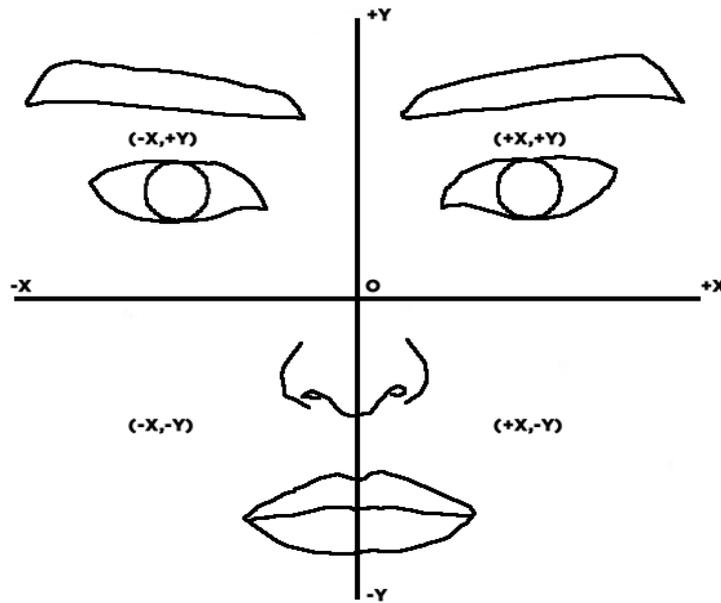


Figura 3.9: Divisão da face em quatro partes.

A Equação 3.3 mostra o cálculo necessário para a translação das coordenadas das FCP's em relação à origem.

$$X = X_{fcp} - X_o \mid Y = Y_{fcp} - Y_o (-1) , \quad (3.3)$$

onde (X,Y) são as coordenadas transladadas das FCP's e (X_{fcp}, Y_{fcp}) são as coordenadas originais das FCP's.

Para compensar as diferenças entre a distância da *webcam* com a face do usuário, os pontos devem passar por um processo de mudança de escala. Mais uma vez o valor da *base* será utilizado. Todas as coordenadas encontradas anteriormente passarão pela Equação 3.4, para aplicar a compensação da escala.

$$X_f = \frac{X}{base} \mid Y_f = \frac{Y}{base} \quad (3.4)$$

3.1.4 Classificação das AU's

Para a classificação das *Action Units*, que são movimentos musculares tabelados por Ekman e Friesen (1978 apud EKMAN; FRIESEN, 2005), a metodologia segue o padrão do FACS. A união de algumas FCP's pode formar uma deformação muscular. Os cálculos para

determinar essas deformações são mostrados na Tabela 3.2 e baseados no trabalho de Jongh (2002).

Tabela 3.2: Deformações musculares da face

Deformação	Equação	Exemplo
Altura das sobrancelhas <i>eb_height</i>	$\frac{((a17.y + a19.y + a21.y) / 3 + ((a18.y + a20.y + a22.y) / 3))}{2}$	
Franzimento das sobrancelhas <i>eb_frowned</i>	$\frac{(((a19.y - a21.y) + (a17.y - a21.y)) + ((a20.y - a22.y) + (a18.y - a22.y)))}{2}$	
Inclinação das sobrancelhas <i>eb_slanting</i>	$\frac{(((a17.y - a21.y) + (a21.y - a19.y)) + ((a18.y - a22.y) + (a22.y - a20.y)))}{2}$	
Distância entre sobrancelhas <i>eb_distance</i>	$a20.x - a19.x$	
Distância entre sobrancelhas e olhos <i>eeb_distance</i>	$\frac{((a17.y - a7.y) + (a8.y - a18.y))}{2}$	
Abertura dos olhos <i>e_openness</i>	$\frac{(((a7.y) - (a5.y)) + ((a8.y) - (a6.y)))}{2}$	
Inclinação dos olhos <i>e_slanting</i>	$\frac{((a3.y - a1.y) + (a4.y - a2.y))}{2}$	
Abertura da boca <i>m_openness</i>	$\frac{(((-(a25.y - a26.y) - (a27.y - a29.y)) / 2) + ((-a25.y - a26.y) - (a28.y - a30.y)) / 2))}{2}$	
Medida do sorriso <i>m_mos</i>	$\frac{((a23.y - (a26.y + ((a25.y - a26.y) / 2))) + (a24.y - (a26.y + ((a25.y - a26.y) / 2))))}{2}$	
Alargamento da boca <i>m_width</i>	$a24.x - a23.x$	
Altura do lábio superior <i>mul_height</i>	$(a29.y + a26.y + a30.y) / 3$	
Altura do lábio inferior <i>mll_height</i>	$(a27.y + a25.y + a28.y) / 3$	

Fonte:

Oliveira; Jaques, (2008).

Como deformações podem ser observadas em imagens estáticas, elas são procuradas em cada *frame* do vídeo.

A deformação é um valor numérico que representa a presença da mesma. Por exemplo, quanto maior o valor da variável $m_openness$ (abertura da boca), mais a boca está aberta.

Uma AU representa um movimento referente às variáveis de deformações. Para saber se uma AU ocorreu, é necessário comparar os valores das variáveis de deformações com os seus valores na mesma face com a emoção neutra. Para isso a metodologia necessita saber manualmente o estado neutro do rosto analisado.

Depois de obter a imagem com a face do usuário com a emoção neutra, é possível analisar a ocorrência da AU, baseado no sistema FACS. A seguir a Tabela 3.3 mostra cada AU com sua regra de ocorrência.

Tabela 3.3: Critérios de classificação das AU's

AU	Critério
1	$eb_height > (eb_height_n + 0.1)$
2	$(eb_frowned < (eb_frowned_n + 0.05)) \wedge (eb_height > (eb_height_n + 0.2))$
4	$(eb_height < (eb_height_n - 0.06)) \wedge (eb_distance < (eb_distance_n + 0.24))$
5	$e_openness \geq (e_openness_n - 0.05)$
6	$(m_mos \geq (m_mos_n + 0.15)) \wedge ((e_openness \geq (e_openness_n - 0.2)) \wedge (e_openness < (e_openness_n - 0.05)))$
7	$(e_openness < (e_openness_n - 0.2)) \wedge (eb_distance < (eb_distance_n + 0.1))$
9	$(m_width < (m_width_n + 0.3)) \wedge (e_openness < (e_openness_n)) \wedge (eb_height < (eb_height_n - 0.05))$
10	$(mul_height > mul_height_n - 0.3) \wedge (mll_height \geq (mll_height_n - 0.3)) \wedge (m_openness > (m_openness_n + 0.3))$
12	$(m_mos \geq (m_mos_n + 0.2)) \wedge (m_width > (m_width_n + 0.2))$
15	$m_mos \leq m_mos_n$
16	$(mul_height \leq (mul_height_n - 0.2)) \wedge (mll_height < (mll_height_n - 0.5)) \wedge (m_openness > (m_openness_n + 0.2))$
17	$(m_openness < (m_openness_n + 0.17)) \wedge (mll_height < (mll_height_n - 0.15))$
20	$(m_width > (m_width_n + 0.25)) \wedge ((m_mos < (m_mos_n + 0.1)) \wedge ((m_mos + 0.1) > 0))$
22	$(m_width < (m_width_n - 0.8)) \wedge (m_openness > (m_mos_n + 0.4))$
23	$m_openness < (m_openness_n - 0.2)$
24	$(mul_height < mul_height_n) \&\& (mll_height > mll_height_n)$
25	$m_openness \geq (m_openness_n + 0.15)$
26	$(m_openness \geq (m_openness_n + 0.3)) \wedge (m_openness \leq (m_openness_n + 0.7))$
27	$m_openness \geq (m_openness_n + 0.7)$

Fonte: Oliveira; Jaques, (2008).

3.1.5 Classificação da emoção

AU Emoção		Ocorrência das AUs																		
		1	2	4	5	6	7	9	10	12	15	16	17	20	22	23	24	25	26	27
Tristeza	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
	1	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
	1	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0
	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0
Repulsa	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	1	0
	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	1	0
	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0

Fonte: Oliveira; Jaques (2008).

Ela é suficiente para determinar uma emoção. Porém, além da combinação das AU's, é necessário saber o nível de intensidade que uma AU ocorre. Por exemplo, um sorriso fraco pode identificar uma AU 12 e já indicar a emoção da alegria. Porém, a alegria pode ocorrer com a combinação da AU 12 + AU 6 (Sobrancelhas abaixadas e juntas), que seria uma expressão mais completa da alegria. Além disso, a própria AU 12 tem uma intensidade em si mesma, que é o quanto estão afastados os cantos do lábio. Em outras palavras, o tanto que o usuário está “sorrindo”.

A metodologia propõe uma forma de inferir as emoções da alegria e tristeza utilizando uma intensidade, que é mostrada como uma barra de porcentagem. Quanto mais características da emoção ocorrer, maior a chance dessa emoção está sendo inferida corretamente. Uma lógica baseada em *Fuzzy* (RIZOL; MESQUITA; SAOTOME, 2011).

O Quadro 3.1 mostra a combinação das AU's necessárias para a inferência da alegria e da tristeza.

Quadro 3.1: Combinação das AU's necessárias para inferir as emoções da alegria e da tristeza.

	Combinação de AU's					
Alegria	12	12 + 6				
Tristeza	1 + 4 + 15	1 + 4 + 15 + 25	6 + 15	6 + 15 + 25	1 + 4 + 15 + 17	17 + 25
	1 + 4 + 15 + 17 + 26	1 + 4 + 15 + 26	15	6 + 15 + 26	1 + 4 + 15 + 17 + 25	17 + 26

Para saber se a probabilidade da emoção está corretamente inferida, cada AU identificada é associada com um peso. Esse peso é determinado pela quantidade de AU's que podem estar presentes na emoção. Por exemplo, para inferir a emoção da Alegria, é necessário a presença da AU 12 sozinha, ou da AU 12 e AU 6 juntas. Como só existem duas AU's (AU 6 e AU 12), cada AU presente recebe o peso de 50%. A ocorrência da AU 6, mesmo sem a presença da AU 12 significa 50% de chance de ocorrência da Alegria. Já a AU 12 sozinha não recebe 100% de peso, pois pela regra, ela por si só representa a Alegria.

A Tabela 3.5 mostra o peso de cada AU e o peso da combinação delas.

Tabela 3.5: Pesos das AU's e das combinações

ALEGRIA		TRISTEZA	
COMBINAÇÃO (AU)	PESOS (%)	COMBINAÇÃO	PESO (%)
6	50%	1, 4, 6, 17, 25 ou 26	14,3%
12 ou 12 + 6	100%	1 + 4, 1 + 17, 1 + 25, 1 + 26, 4 + 17, 4 + 25, 4 + 26, 6 + 25 ou 6 + 26	28,6%
		1 + 4 + 25 ou 1 + 4 + 26	42,9%
		1 + 4 + 15, 1 + 4 + 15 + 17, 17 + 26, 15, 1 + 4 + 15 + 17 + 25, 1 + 4 + 15 + 25, 6 + 15, 6 + 15 + 25, 1 + 4 + 15 + 17 + 26,	100%

		1 + 4 + 15 + 26, 15 ou 6 + 15 + 26	
--	--	------------------------------------	--

A Figura 3.10 mostra um fluxograma que representa a inferência da Alegria com base no sistema de pesos proposto pela metodologia.

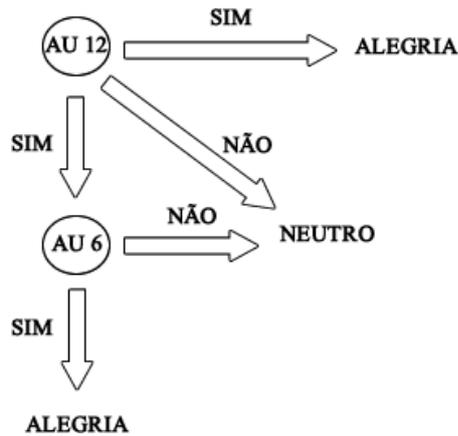


Figura 3.10: Fluxograma de inferência da Alegria.

De forma análoga, a inferência da Tristeza conta com o mesmo sistema de pesos, com a diferença que são sete AU's distintas e doze combinações diferentes.

A emoção Neutra também é analisada. Mas ela é mostrada somente na ausência total das emoções da Alegria e da Tristeza.

4 AVALIAÇÃO DOS RESULTADOS

Este capítulo mostra a avaliação da metodologia proposta, assim como os resultados observados na execução da mesma.

4.1 Ferramentas utilizadas

Para demonstrar os resultados, um estudo de caso foi desenvolvido utilizando as seguintes ferramentas:

- Notebook HP com processador de dois núcleos e 4GB de memória RAM;
- Webcam HP TrueVision com 1.3 megapixels de resolução;

O teste é feito em um ambiente controlado, com pouco movimento externo e iluminação artificial, por apresentarem resultados menos variáveis.

4.2 Estudo de caso

Para iniciar o estudo de caso, o usuário fica em frente a *webcam* do notebook. A proximidade do usuário a *webcam* é de livre arbítrio, pois como explicado na Seção 3.1.3 do Capítulo 3, o sistema se adequa a essa distância de forma automática.

O usuário fica sempre em uma posição horizontal em relação a *webcam*, pois a metodologia não trata a rotação dos pontos a serem trabalhados. A rotação da face, por exemplo, pode ocasionar possíveis erros nos resultados apresentados.

O sistema inicia e executa as três primeiras etapas da metodologia. Na primeira, face é detectada. Na próxima etapa, o olho esquerdo, o olho direito, a sobrancelha esquerda, a sobrancelha direita e a boca são detectados, na ordem citada e de forma distinta, utilizando cada técnica proposta pela metodologia. Após essa etapa, são detectadas as FCP's e o ponto Origem.

A Figura 4.1 mostra as FCP's e do ponto Origem já detectados e desenhados sobre a face do usuário.



Figura 4.1: FCP's e ponto Origem desenhadas sobre a face do usuário. Os pontos azuis pequenos representam as FCP's das sobrancelhas e o ponto azul grande representa o ponto Origem. Os pontos brancos representam as FCP's dos olhos. Os pontos verdes representam as FCP's da boca.

Depois de detectadas as FCP's, o sistema continua detectando as mesmas em cada *frame* do vídeo, até o usuário interagir com a janela.

Existem duas formas de interação. A primeira é com o botão esquerdo do *mouse* sobre a janela, que executa a próxima etapa da metodologia, a de inferir as AU's.

Cada AU é devidamente calculada. Porém, mais um passo é necessário para concluir essa etapa. É necessário guardar os valores das AU's do rosto em estado Neutro. Esse passo é importante para a etapa posterior, pois ele guarda as variáveis de deformação neutras, que serão usadas para serem comparadas com as deformações atuais para a inferência da emoção.

A segunda forma de interação com a janela é com o botão direito do *mouse* sobre a janela. Com o *click* do botão, o usuário determina manualmente o estado Neutro da sua face.

A vantagem de tornar este passo manual é que o usuário pode corrigir sempre que desejar a sua face em estado Neutro, caso não esteja satisfeito com os resultados obtidos.

Quando este passo é executado, a última etapa é iniciada - a classificação das emoções. Cada AU detectada é utilizada dentro das regras da classificação das emoções da Alegria e da Tristeza demonstradas na Seção 3.1.5 do Capítulo 3.

De forma automática, o sistema mostra barras de probabilidade nas emoções de estado Alegre, Neutro e Triste.

A Figura 4.2 mostra o resultado da inferência com a face do usuário em estado Neutro.



Figura 4.2: Face do usuário em estado Neutro. A barra amarela mostra que o estado está 100% neutro.

A Figura 4.3 mostra a face do usuário demonstrando a emoção de Alegria.

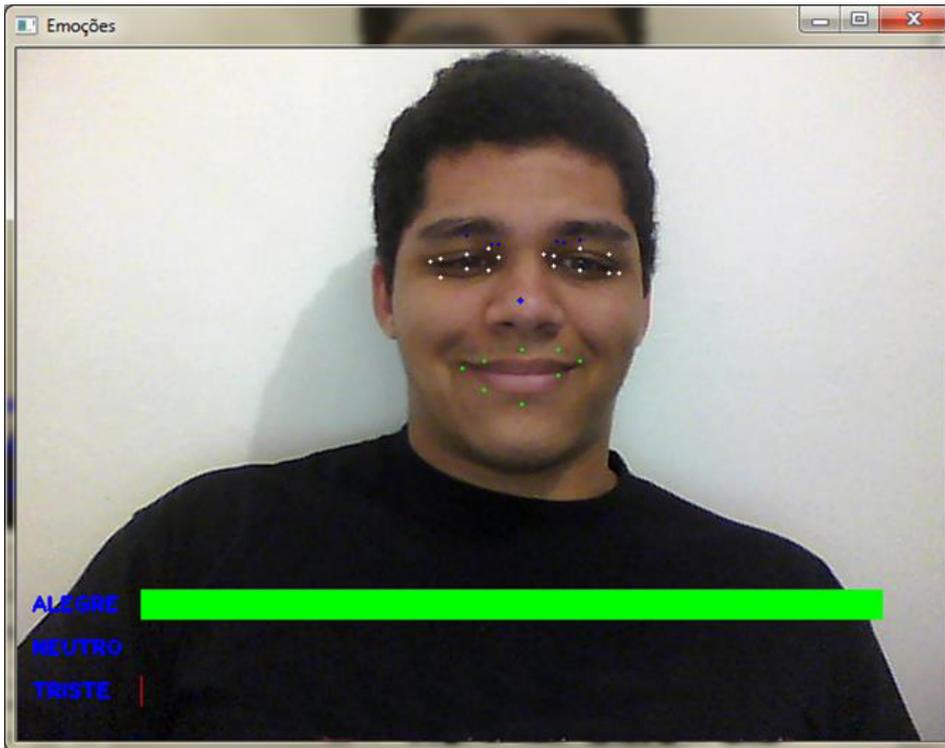


Figura 4.3: Ocorrência da emoção da Alegria na face do usuário. A barra verde demonstra que a Alegria está acontecendo com a probabilidade de 100%.

A Figura 4.4 mostra a emoção da Tristeza detectada.



Figura 4.4: Ocorrência da emoção da Tristeza na face do usuário. A barra vermelha mostra que a Tristeza está ocorrendo com a probabilidade de 100%.

É possível ainda ser detectado a presença de possibilidade das duas emoções ao mesmo tempo, como mostrado na Figura 4.5.

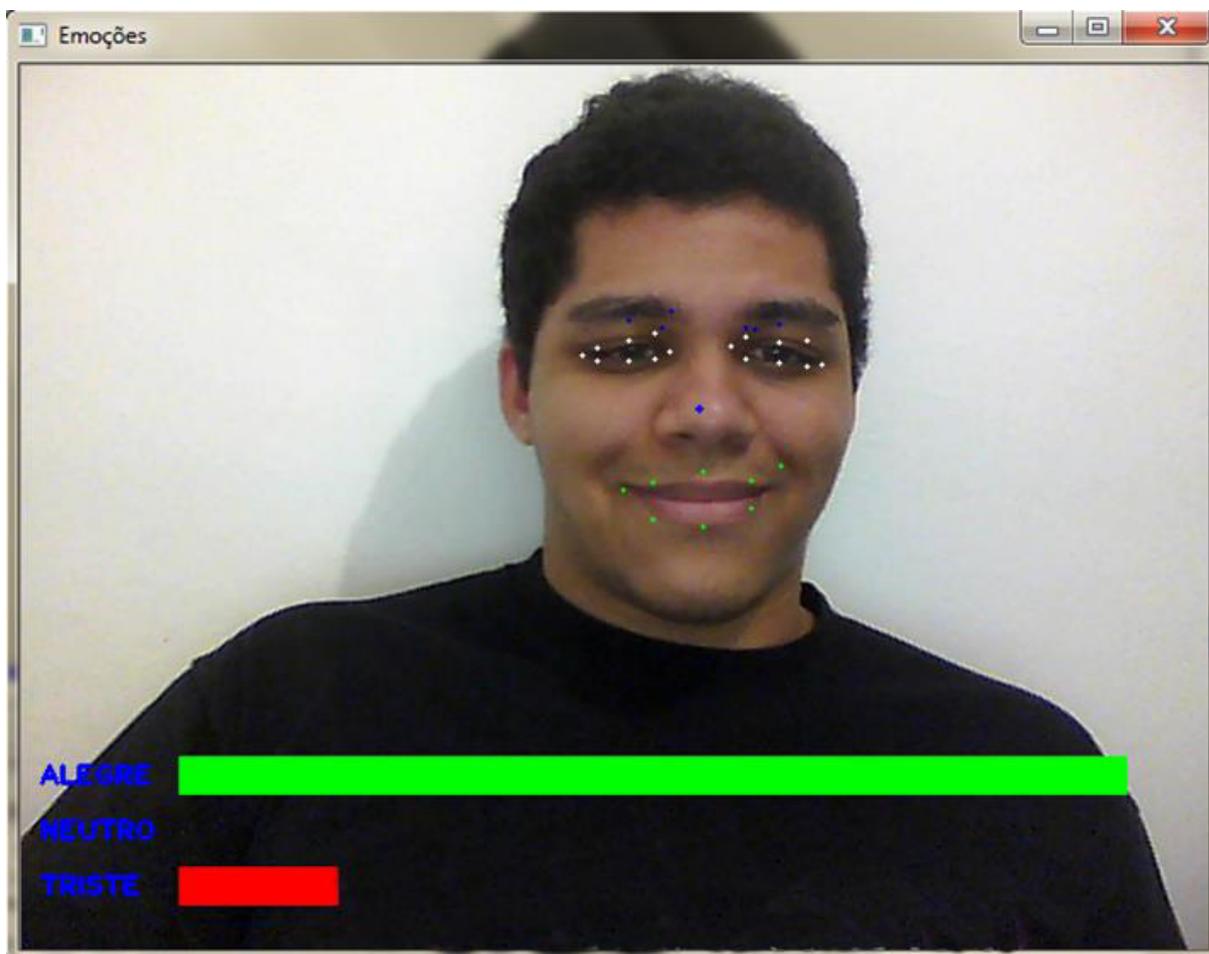


Figura 4.5: Ocorrência das duas emoções simultaneamente. A barra verde está 100% cheia e a barra vermelha parcialmente cheia.

Essa mistura de detecções ocorre em casos em que uma AU detectada em uma emoção é compartilhada com a outra. No exemplo da Figura 4.5, ocorreu a detecção das AU's 6 e 12 representando a Alegria. A AU 6 também está presente no escopo da detecção da Tristeza, explicando assim sua aparição simultânea com a Alegria.

Esse passo é executado em cada *frame* do vídeo de forma contínua até a finalização do sistema por intermédio do usuário.

4.3 Análise dos resultados

Foi realizado o recolhimento de dados estatísticos referente aos testes feitos no estudo de caso. Os experimentos foram realizados com três pessoas diferentes. Em cada experimento, seis baterias de testes contendo cada uma 100 (cem) *frames* consecutivos foram realizados, sendo que cada intervalo entre elas o programa é calibrado manualmente pelo usuário, através do botão direito do *mouse* – que atualiza os dados das FCP's relacionadas à emoção Neutra a cada clique.

Cada experimento terá duas tabelas de resultados. A primeira mostra os resultados obtidos em três baterias de teste focadas na emoção da Alegria e a segunda mostra os resultados obtidos em três baterias de teste focadas na emoção da Tristeza. Além disso, cada teste contabilizará somente as emoções em sua probabilidade máxima.

Nas tabelas, o campo Alegre (%) mostra a porcentagem de inferências da emoção da Alegria. O campo Triste (%) mostra a quantidade de inferências da emoção da Tristeza. E o campo Neutro ou Emoção Parcial (%) indica a porcentagem da inferência da emoção Neutra ou da emoção da Alegria ou Tristeza sem o peso de 100%.

4.3.1 Experimentos

O Experimento I é feito em um ambiente com iluminação artificial através de uma lâmpada fluorescente de 34 *watts* incidindo diretamente sobre indivíduo. O fundo é uma parede branca. A câmera da *webcam* está a 60 centímetros do usuário.

O Experimento II é realizado em um local com iluminação artificial com uma lâmpada fluorescente de 30 *watts*, incidindo por trás da pessoa utilizada para os testes. O fundo é composto de uma estante com diversos objetos.

O Experimento III é realizado em um local com iluminação artificial com uma lâmpada fluorescente de 30 *watts*, incidindo sobre a pessoa utilizada para os testes. O fundo é composto de uma parede branca.

A Tabela 4.1 mostra os resultados obtidos nas três primeiras baterias de teste com os três usuário. A emoção que é testada primeiramente é a Alegria.

Tabela 4.1: Experimentos com os usuários focados na Alegria

	ALEGRE (%)			NEUTRO OU EMOÇÃO PARCIAL (%)			TRISTE (%)		
	Teste 1	Teste 2	Teste 3	Teste 1	Teste 2	Teste 3	Teste 1	Teste 2	Teste 3
Experimento I	59	56	81	1	0	4	40	44	15
Experimento II	43	43	56	24	30	8	33	27	36
Experimento III	56	62	60	22	20	18	22	18	22

A Tabela 4.2 os resultados obtidos nas três baterias de teste com os três usuário focados na emoção da Tristeza.

Tabela 4.2: Experimentos com os usuários focados na Tristeza

	ALEGRE (%)			NEUTRO OU EMOÇÃO PARCIAL (%)			TRISTE (%)		
	Teste 1	Teste 2	Teste 3	Teste 1	Teste 2	Teste 3	Teste 1	Teste 2	Teste 3
Experimento I	1	0	0	12	7	9	87	93	91
Experimento II	24	6	39	15	10	0	61	84	61
Experimento III	18	9	1	44	56	50	38	35	49

No Experimento I observa-se uma média aritmética de 65,3% de acerto na emoção da Alegria e de 90,3% de acerto na emoção da Tristeza. No Experimento II, a média aritmética de acerto na emoção da Alegria é de 47,3% e na emoção da Tristeza é de 68,3%. No Experimento III a média aritmética de acertos observada é de 59,3% na emoção da Alegria e 40,6% na emoção da Tristeza.

A quantidade elevada de acertos na emoção da Tristeza se dá em conta do número maior de AU's envolvidas na inferência da mesma. Essa quantidade é facilitada pelos erros de detecção das FCP's, resultando em um maior número de falsos positivos.

Essa explicação serve, de forma análoga, para justificar a quantidade menor de acertos da emoção da Alegria, assim como a quantidade elevada de erros de inferência da mesma.

O estado Neutro é pouco provável de ocorrer, já que um bom número de AU's é trabalhado na detecção da Alegria e da Tristeza juntas. É pouco provável uma face não conter essas AU's. Porém, algumas pessoas não são muito suscetíveis a mudanças drásticas de emoções aparentes. Ou seja, se expressam de forma menos acentuada, como o caso do Experimento III, que tem uma média de inferência da emoção Neutra ou de emoções parciais em 20% no Teste da Alegria e de 50% no Teste da Tristeza.

O crescimento nos acertos das inferências na sequência dos testes é explicado pela calibragem feita antes de cada teste. O programa se adequa ao ambiente em cada intervalo. Porém, isso pode servir também para a diminuição dos acertos, caso não seja utilizado de maneira correta.

Devido as configurações do *notebook* utilizado para os testes, o sistema roda com uma taxa de 1 frame por segundo sem variações.

Os resultados observados na avaliação do estudo de caso são satisfatórios no contexto trabalhado. A ocorrência de duas (ou mais) emoções simultaneamente pode não ser prática e pode confundir os resultados, caso sejam utilizadas mais emoções além da Alegria e da Tristeza.

5 CONCLUSÃO

Uma metodologia de inferência da emoção da Alegria e da Tristeza foi o objetivo deste trabalho, e foi demonstrada cada etapa necessária para desenvolvê-la.

Criar um método de inferência de emoções não é uma tarefa simples. Vários obstáculos são encontrados no caminho, como limitação tecnológica – computadores com pouca capacidade de processamento, câmeras de resolução baixa, etc. – ou mesmo um modelo prático a ser trabalhado.

Apesar dos diversos trabalhos relacionados à inferência de emoções, poucos têm resultados próximos à perfeição almejada pelo seu desenvolvedor. Para chegar perto dessa perfeição, é necessário escolher um caminho consistente a ser seguido.

De acordo com as referências escolhidas durante o levantamento dos conhecimentos básicos para o desenvolvimento de uma metodologia com o objetivo de inferir as emoções da Alegria e da Tristeza, a biblioteca de Visão Computacional OpenCV foi escolhida por conter diversos métodos estudados e também escolhidos para o desenvolvimento da metodologia.

O método criado para a identificação das FCP's é o ponto forte do trabalho, pois em geral, a identificação delas não é feita de forma automática.

A interação do usuário com o sistema desenvolvido para a demonstração da metodologia é simples e prática. Os resultados são bons e consistentes, considerando que são observados em tempo real.

Apesar disso, toda metodologia é sujeita a falhas e a melhorias. Uma sugestão de melhoria futura seria a utilização de *Active Shape Model* para a detecção das FCP's, visto que esse processo tem uma acurácia muito alta para detecção de pontos faciais (HSIEH, 2011).

Podem ser utilizadas mais emoções para a inferência futuramente, além de alguma forma de detectar o sentimento Neutro de forma automática, onde o próprio sistema se adequa e corrige o mesmo em tempo real.

REFERÊNCIAS

- ADELSON, E. H. **Checkershadown Proof**. 1995. Disponível em: <http://web.mit.edu/persci/people/adelson/checkershadown_proof.html>. Acesso em: 10 jan. 2013
- BOOTH, Paul A. *An introduction to human-computer interaction*. Lawrence Erlbaum. Mahwah, 1989.
- CAMPOS, F. M. S. de. **Detecção e rastreamento de faces em vídeos – Como detectar faces em vídeos?** 2011. Disponível em: <<http://www.bitabit.eng.br/2011/02/21/como-detectar-faces-em-videos/>>. Acesso em: 10 jan. 2013.
- EKMAN, Paul; FRIESEN, Wallace V. *Facial Action Coding System, Consulting Psychologist Press*. Palo Alto, CA, 1978.
- EKMAN, Paul; FRIESEN, Wallace V. **EMFACS-7: Emotional Facial Action Coding System**. 1983.
Disponível em: <<http://www.face-and-emotion.com/dataface/facs/emfacs.jsp>>
Acesso em 18 dez. 2012.
- EKMAN, Paul; FRIESEN, Wallace V.; HAGER, Joseph C. *Facial Action Coding System: The Manual*. A Human Face. Estados Unidos, 2002.
- EKMAN, P.; ROSENBERG, Erika L. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Segunda Edição. OXFORD UNIVERSITY PRESS, 2005.
- FREUND, Y.; SCHAPIRE, R. E. *A decision-theoretic generalization of on-line learning and an application to boosting*. J. Computer and System Sciences, 1995.
- FRIESEN, W.V.; EKMAN, P. **EMFACS-7: Emotional Facial Action Coding System**. *Unpublished manuscript*, University of California at San Francisco, 1983.
- HANSON, A. R.; RISEMAN, E. M. *Computer Vision Systems*. Academic Press, New York, 1978.
- HSIEH, Chen-Chiung. *A Facial Expression Classification System Based on Active Shape Model and Support Vector Machine*. *International Symposium on Computer Science and Society (ISCCS)*, 2011.
- INTEL. *Open Computer Vision Library Reference Manual*. Intel Corporation, USA, 2001.
- DOCUMENTAÇÃO do OpenCV. Disponível em: <<http://docs.opencv.org>>. Acesso em: 10 jan. 2013
- IZARD, C. E. *Basic Emotions, Relations Among Emotions, and Emotion-Cognition Relations*. *Psychological Review*, Vol. 99, No. 3, 561-565, 1992.

JONGH, E. J. de. **FED: An online facial expression dictionary as a first step in the creation of a complete nonverbal dictionary**. Dissertação (Mestrado) — *Delft University of Technology*, 2002.

KOBAYASHI, H.; HARA, F. **The recognition of basic facial expressions by neural network**. *Proceedings of International Joint Conference on Neural Network*. p. 460–466. 1991

KOBAYASHI, H.; HARA, F. **A Basic Study of Dynamic Recognition of Human Facial Expressions**. *Proceedings of IEEE International Workshop on Robot and Human Communication*, pp.271-275, 1993-11.

KOBAYASHI, H.; HARA, F, TANGE A. **Real-Time recognition of six basic facial expressions**. *Proceedings of IEEE International Workshop on Robot and Human Communication*, pp.179-186, 1995.

LIENHART, R.; MAYDT, J. **An extended set of Haar-like features for rapid object detection**. *IEEE ICIP 2002*, Vol. 1, pp. 900-903, Sep. 2002.

MARR, D. **Vision: A Computational Investigation into the Human Representation and Processing of Visual Information**. W. H. Freeman, San Francisco, 1982.

MATHEUS, A, S. **Câncer de Pâncreas**. Disponível em: <<http://www.asmatheus.com.br/cancerpancreas.html>>. Acesso em: 10 jan. 2013.

MENEZES, P.; BARRETO, J.C.; DIAS, J. **Face tracking based on Haar-like features and eigenfaces**. *5th IFAC Symposium on Intelligent Autonomous Vehicles*, Lisboa, Portugal, 2004.

MOTOKAWA, Koiti. **Field of retinal induction and optical illusion**. *Journal of Neurophysiology*, Vol 13, páginas 413-426. 1950.

OLIVEIRA, Eduardo de; JAQUES, Patrícia Augustin. **Inferindo emoções do usuário pela face através de um sistema psicológico de codificação facial**. Porto Alegre: PIPCA, 2008.

OLIVEIRA, Rogéria Bernardo de. **Benefícios da ergonomia física, cognitiva e organizacional para as empresas**. 2010.
Disponível em: <<http://www.administradores.com.br/informe-se/artigos/beneficios-da-ergonomia-fisica-cognitiva-e-organizacional-para-as-empresas/48442/>>.
Acesso em: 12 de Dezembro de 2012.

POTÚČEK, Igor. **Projects**. Disponível em: <<http://www.potucek.net/projects.html>>. Acesso em: 11 jan. 2013.

PICARD, Rosalind. **Affective Computing**. *MIT Press*, Cambridge, 1995.

RIZOL, P. M. S. Rocha; MESQUITA, L.; SAOTOME, O. **Lógica Fuzzy Tipo-2**. *Revista SODEBRAS*, v. 6, p. 27-46, 2011.

ROBERTS, L. G. *Machine perception of three-dimensional solids*. *Optical and Electro-Optical Information Processing*, chapter 9, páginas 159–197, MIT Press, Cambridge, Massachusetts, 1965.

RUSSEL, J. A. *Is There Universal Recognition of Emotion From Facial Expression? A Review of the Cross-Cultural Studies*. *Psychological Bulletin*, Vol. 115, No. 1, 102-141, 1994.

SNAVELY, N., Seitz, S. M.; SZELISKI, R. *Photo tourism: Exploring photo collections in 3D*. *ACM Transactions on Graphics*, 25(3), 835–846, 2006.

SZELISKI, R. *Computer Vision: Algorithms and Applications*. Springer, 2008.

TOSCANO, Roberto. *Integral Image*. 2011. Disponível em:
<<http://imaginebits.wordpress.com/2011/08/17/integral-image/>>. Acesso em: 10 jan. 2013.

VIOLA, P.; JONES, M. *Rapid object detection using boosted cascade of simple features*. *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

WILSON, Phillip Ian. *Facial feature detection using haar classifiers*. *Consortium for Computing Sciences in Colleges*, 2006.

WINSTON, P. H. *The Psychology of Computer Vision*. McGraw-Hill, New York, 1975.