

# SISTEMA VISUAL PARA INTERAÇÃO HUMANO-MÁQUINA

CAROLINA M. SALCEDO\*, JÉS DE JESUS CERQUEIRA\*, ANTONIO M. N. LIMA†

\*Programa de Pós-Graduação em Engenharia Elétrica

Escola Politécnica da Universidade Federal da Bahia

Rua Aristides Novis, 02, Federação, Salvador, Bahia, Brasil Telefone: +55-71-3283-9776

†Centro de Engenharia Elétrica e Informática

Departamento de Engenharia Elétrica da Universidade Federal de Campina Grande

Rua Aprígio Veloso, 882, Bodocongó, Campina Grande, Paraíba, Brasil Telefone: +55-83-310-1136

Emails: carolina.moreno@ufba.br, jes@ufba.br, amnlma@dee.ufcg.edu.br

**Abstract**— This paper presents preliminary results about the development of a real time vision system for human-machine interaction (**HMI**). The purpose of such vision system is to develop a tool that provides additional information for the caregiver (therapist, teacher, etc.), concerning the emotional behavior of people with limited social interaction. The vision system integrates the detection, tracking and facial expression recognition of human. The techniques for detection and tracking of faces are based on a combination of techniques which allows to obtain greater robustness of the system. The recognition of facial expressions is based on Active Shapes Models (**ASM**) and Facial Action Coding Systems (**FACS**). Experimental results are presented, and shown to be satisfactory, with a success rate of 82.25%.

**Keywords**— HMI, Detector, ASM, FACS, features.

**Resumo**— Este artigo apresenta resultados preliminares do desenvolvimento de um sistema de visão em tempo real para interação humano-máquina. O objetivo de tal sistema de visão é desenvolver uma ferramenta que forneça informação adicional ao cuidador (terapeuta, professor, etc.), acerca do comportamento emocional de pessoas com limitação na interação social. O sistema de visão contém: (i) detecção e rastreamento de faces humanas em tempo real; e (ii) reconhecimento de expressões faciais. As técnicas de detecção e rastreamento de faces são baseadas numa combinação de técnicas que permite obter maior robustez do sistema. O reconhecimento das expressões faciais é baseado no Modelo de Forma Ativa (**ASM**) e unidades de codificação de ação facial (**FACS**). Resultados experimentais são apresentados, e mostram-se satisfatórios, com uma taxa de acerto de 82.25%.

**Palavras-chave**— IHM, Detector, ASM, FACS, características

## 1 INTRODUÇÃO

Um dos principais objetivos na área de tecnologia da informação, é dotar o computador de um comportamento inteligente (um comportamento humano, por assim dizer), o que tem levado a pesquisas e desenvolvimento de inúmeros estudos que podem trazer melhorias significativas na área da interação humano-máquina, tornando-se ao mesmo tempo uma contribuição relevante para as ciências cognitivas, que vêm mostrando avanços significativos no auxílio de terapias para autismo; terapias para pessoas com deficiências físicas; em psicopedagogia e medicina.

Diversos pesquisadores argumentam que existe um conjunto de emoções denominadas básicas ou primárias, que foram preservadas durante processos de evolução, pois elas possibilitam a adaptação do comportamento dos organismos a diversas situações com as quais as criaturas se deparam diariamente. Entre as apresentadas na literatura como básicas, podem-se citar a raiva, a alegria, o medo, tristeza e a surpresa. Estas emoções servem a propósitos específicos e emergem em determinados contextos do ambiente, para preparar o organismo a responder apropriadamente (Almejrads, 2010; Ekman e Friesen, 1978).

A habilidade de imitar expressões faciais humanas estabelece as bases para uma comunicação não-verbal significativa entre humanos e máquinas. A análise de expressões faciais nasceu de Darwin, no século XIX onde ele propôs o conceito de expressões faciais universais em homens e animais.

Pesquisas recentes em terapia de Transtorno do espectro Autista - TEA, estão fazendo uso da Interação Humano-Máquina como auxílio às mesma, e têm-se apresentado avanços consideráveis. Os pesquisadores têm-se focado em desenvolver sistemas que permitam reconhecer e ensinar emoções a este tipo de pacientes. Mas o problema não se radica aí; é importante saber se o que o paciente está aprendendo é realmente o tipo de expressão adequada e, caso apareçam erros, ter outras ferramentas para corrigi-los.

Na literatura técnica existem propostos vários sistemas de visão para interação humano-máquina que reconhecem expressões faciais. Em Azcarate *et al.* (2005) por exemplo, construíram uma máscara que é deformada com os movimentos faciais realizados quando o usuário muda a expressão facial, utilizando os classificadores *Naive Bayes* (Precisão 64.3%). Bahonar e Charkari (2009), detectam regiões da face, tais como olhos, sobrancelhas, nariz e boca. Zhao *et al.* (2012), imple-

mentaram o reconhecimento da expressão facial usando modelos de Forma Ativa (**ASM**) e características de padrões binários locais (**LBP**) com um desempenho de 85%.

O objetivo deste artigo é apresentar os resultados preliminares do desenvolvimento de um sistema visão que permita fornecer dados adicionais do comportamento emocional. às pessoas com limitação na interação social (especificamente pessoas com Transtorno do Espectro Autista-**TEA**). As informações obtidas neste sistema de visão, vão permitir a identificação do estado emocional do autista, e assim, fornecer esta informação ao profissional (professor, terapeuta, ou mesmo a pessoa que cuida do autista) como auxílio nas terapias. Este sistema de visão, futuramente será combinado com **BCI** (Interface Cérebro-Computador), com a finalidade de obter um sistema de percepção mais robusto para detectar emoções.

A organização do artigo é como segue: na seção 2 são apresentados alguns conceitos acerca de interação humano-máquina no autismo; os métodos de detecção e rastreamento de faces utilizados são apresentados na seção 3.1; na seção 4 é apresentado o método de reconhecimento da expressão facial; os resultados experimentais são mostrados na seção 5; e seguidamente algumas considerações finais, e referências utilizadas.

## 2 Interação Humano-Máquina

Inúmeros são os problemas que envolvem o desenvolvimento das pessoas com autismo. Dentre eles, podemos citar as dificuldades em expressar os próprios sentimentos, de uma forma que os outros possam compreendê-los e aceitá-los. Outro problema básico refere-se à inadequação ou falta do uso da imaginação. A ausência de habilidade para perceber e compreender expressões emocionais em outras pessoas parece relacionar-se com a limitação, ou mesmo falta da capacidade de imaginar qualquer coisa.

Segundo o professor e psicólogo *Armindo Freitas-Magalhães*, director do *Facial Emotion Expression Lab (FEELab)* da Universidade Fernando Pessoa, em Portugal, dados preliminares de um estudo revelam que os deficientes mentais têm déficit cognitivo na identificação das emoções básicas – a alegria, a tristeza, a surpresa, o medo, a cólera e a aversão. Estas revelações podem ter implicações práticas na forma como se reeduca crianças com deficiências mentais, já que demonstram que estas podem interpretar mal a reação de um colega ou de um educador. Assim, é importante verificar até que ponto é possível alterar as práticas educacionais e ajudar um deficiente a ter outro tipo de atitude, interpretando bem as expressões faciais de quem trabalha com eles.

Assim, a necessidade de desenvolver ferramentas que ajudem a monitorar estados emocionais de pessoas com limitações na interação social (autistas, como um exemplo), permite auxiliar aos cuidadores nas suas atividades, por exemplo em uma escola, onde o professor tem vários alunos com este tipo de deficiência, fornecer informação comportamental emocional é uma grande ajuda.

## 3 Detecção e Rastreamento de Faces

O detector proposto neste trabalho está inspirado no modelo de Viola e Jones (2004), que é um dos métodos mais usados e citados pela comunidade científica, e nas posteriores modificações feitas. Este detector é baseado em uma cascata de classificadores que explora toda a imagem em múltiplas escalas e localizações. Cada etapa da cascata é baseada no uso de características do tipo Haar, que são eficientemente computadas utilizando a imagem integral, selecionadas e combinadas por meio do método *AdaBoost* durante o treinamento (veja seção 3.2).

### 3.1 Descritores de Haar para Detecção de Faces

Há três contribuições fundamentais: a primeira é a introdução da integral da imagem, que permite que as características usadas pelo detector sejam processadas através de uma única passagem pela imagem. A segunda é um classificador simples e eficiente que é construído usando o *AdaBoost*, que seleciona um pequeno número de características visuais críticas de um conjunto muito grande de potenciais características. A terceira consiste num método para combinar os classificadores numa cascata, que permite descartar as regiões de fundo da imagem de forma rápida, focando-se assim apenas nas regiões de maior interesse.

### 3.2 Aprendizagem com Adaboost

O princípio do *Adaboost* é combinar um conjunto de classificação simples ou fracos<sup>1</sup> para formar uma função de classificação forte. A eficiência deste esquema reside no fato que os falsos negativos (ou componentes classificadas abaixo de um limiar) são eliminados progressivamente, de forma que as primeiras etapas eliminam um grande número deles com pouco processamento. Isto permite que as etapas finais tenham o tempo suficiente para classificar corretamente os casos mais difíceis (Salcedo *et al.*, 2012b).

Uma função de classificação fraca  $h_i(x, f, p, \theta)$  é constituída de um descritor  $f_i$ , de um limiar  $\theta_i$ ,

<sup>1</sup>A função de classificação fraca, só é capaz de reconhecer pelo menos duas classes.

e de uma paridade  $p_i$  que indica a direção do sinal de desigualdade na forma

$$h_i(x, f, p, \theta) = \begin{cases} 1 & \text{se } p_i f_i(x) < p_i \theta_i \\ 0 & \text{em outro caso} \end{cases}, \quad (1)$$

onde  $x$  é uma janela da imagem de entrada.

Os passos descritos para a implementação do algoritmo de *Adaboost* são a seguir apresentados (Freund e Schapire, 1995).

**Entrada:**  $S = \{(x_1, y_1), \dots, (x_m, y_m)\};$   
 $x_i \in X, y_i \in \{0, 1\}$

**0.1** Inicialize  $w_{1,i} = \frac{1}{2m} \forall (x_i, y_i) \in S$

**0.2** para  $t = 1, \dots, T$  faça

**0.3** Normalize os pesos  
 $w_{1,i} \leftarrow \frac{w_{t,j}}{\sum_{j=1}^n w_{t,j}}$

**0.4** Selecione o melhor classificador baseado no peso de erro  
 $\epsilon_t = \min_{f,p,\theta} \sum_i w_i |h(x_i, f, p, \theta) - y_i|$

**0.5** Defina  $h_t(x) = h(x, f_t, p_t, \theta_t)$

**0.6** Atualize os pesos  
 $w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$ , onde  $e_i = 0$  se a amostra  $x_i$  é classificada corretamente,  $e_i = 1$  caso contrario,  
e  $\beta_t = \frac{\epsilon_t}{1 - \epsilon_t}$

**0.7** O classificador robusto final é dado por:

$$C(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t, \\ 0 & \text{em outro caso,} \end{cases}$$

onde  $\alpha_t = \log \frac{1}{\beta_t}$

**1.8** **Algoritmo 1:** *Adaboost*, (Freund e Schapire, 1995)

O centro e o tamanho da face que se quer rastrear são encontrados através do *CamShift* operando na imagem de probabilidade de cores. O tamanho e a localização corrente da face são informados e usados para definir o tamanho e a localização da janela de busca da próxima imagem de vídeo.

O custo computacional deste algoritmo é relativamente baixo e é bastante preciso. Após esta fase é necessário que se faça uma conversão da imagem para uma escala de cinza, facilitando os processamentos posteriores. O uso de *CamShift* acentua a capacidade do uso software em ambientes de baixa iluminação (Salcedo *et al.*, 2012a). O algoritmo *CamShift*, pode ser resumido nos seguintes passos:

**1.1** Eleição de um tamanho de janela de procuração.

**1.2** Eleição da localização inicial dessa janela.

**1.3** Calcular *MeanShift* (uma ou varias interações em função da convergência). Se armazena o momento de ordem zero (área da distribuição de probabilidade).

**1.4** Tomar como novo tamanho da janela 1 de tal forma que seja função do momento de ordem zero (área).

**1.5** Repetir os dois passos anteriores até chegar à convergência (a localização média está separada do centro atual da janela num valor que não supera um umbral definido).

**2.6** **Algoritmo 2:** *Camshift*, (Bradski, 1998)

### 3.3 Rastreamento de Faces Usando CamShift

A técnica de segmentação aqui usada é baseada é chamada de *CamShift* (*Continuously Adaptive Mean Shift*). É uma segmentação de cor de imagens apresentado por Bradski (1998). É baseado numa técnica estatística onde se busca o pico entre distribuições de probabilidade em gradientes de densidade. Esta técnica é chamada de “média por deslocamento” (*mean shift*) e foi adaptada no *Camshift* para tratar a mudança dinâmica das distribuições de probabilidade das cores numa sequência de vídeo (Bradski, 1998).

Para cada quadro, a imagem é convertida para outra de distribuição de probabilidade de cor através de um modelo de histograma da cor da pele.

## 4 Extração de Características Faciais

Em Ekman e Friesen (1978), é feita uma classificação de seis emoções básicas que, segundo os autores, são universais para todos os seres humanos: angry (raiva), disgust (desgosto), fear (medo), happy (feliz), sad (triste), and surprise (surpreso). Os mesmos autores, desenvolveram o sistema de ação de codificação facial (FACS) para analisar e sintetizar a expressão facial com base em unidades de ação muscular (Action Units- AU), que descrevem movimentos faciais básicos. Uma vez que as unidades de ação muscular são independentes de qualquer interpretação, elas podem ser usadas em conjunto com outros descritores, o que inclui a sua associação ao reconhecimento de emoções básicas. Este método é um dos mais usados na literatura

para o desenvolvimento de sistemas de reconhecimento de expressões faciais.

Cada unidade de ação (**AU**) tem como base estudos relacionados à anatomia dos músculos faciais. Muitos pesquisadores se inspiraram nessas **AU** fazendo uso de processamento de imagem e vídeo para automaticamente rastrear características faciais e então utilizá-las para categorizar as diferentes expressões. Por meio da utilização de combinações diferentes destas unidades de ação, é possível definir um largo espectro de expressões faciais.

#### 4.1 Modelo de Forma Ativa (**ASM**)

O Modelo de Forma Ativa (**ASM**), representa um modelo paramétrico deformável, onde é construído um modelo estatístico da variação da forma global de um conjunto de treinamento. Este modelo, chamado de Modelo Pontual de Distribuição (**PDM**), é utilizado para ajustar um modelo de ocorrências invisíveis do objeto anteriormente anotado no conjunto de treinamento (Zhao *et al.*, 2012). A variabilidade da forma é aprendida off-line usando Análise de Componentes Principais (**PCA**).

A abordagem **ASM** trabalha com formas dos objetos mapeados no plano  $2D$ , logo qualquer forma de um objeto pode ser descrita também como um vetor de  $m$  pares de pontos  $(x, y)$  que contornam essa forma, como pode ser visto na equação (2).

$$x = (x_1, \dots, x_m, y_1, \dots, y_m)^T. \quad (2)$$

Esta equação reduz a relação dos pares de coordenadas  $(x, y)$  para apenas uma forma. Se ao contrário, houver mais exemplos do objeto de estudo, haverá então um vetor  $x_i$  para cada diferente forma ( $i = 1, 2, \dots, N$ ), formando assim uma matriz de dados  $X$ . Se é considerado que um determinado ponto  $(x, y)$  de uma imagem  $I(x, y)$  está correlacionado com os outros pares  $(x, y)$  de todas as imagens da matriz de formas  $X = (x_1, x_2, \dots, x_N)$ , então poderá se aplicar a análise de componentes principais (**PCA**) para determinar a variância do conjunto de formas em  $X$ .

Porém, antes da matriz  $X$  ser calculada pelo **PCA**, é necessário um alinhamento das formas em  $t_x, t_y$  (translação),  $s$  (escala) e  $\theta$  (rotação), isto porque o que se deseja conhecer é como se distribuem os pontos em  $\mathbb{R}^n$  em relação a uma forma base.

$$x = T_{t_x, t_y, s, \theta}(X) \quad (3)$$

A forma média neste domínio alinhado é dada como:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^n x_i \quad (4)$$

E o desvio de cada forma, a partir da forma média  $x_i - \bar{x}$ . A estimativa da matriz de covariância  $\Sigma$  da matriz  $X$  é dada por

$$\Sigma = \frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T, \quad (5)$$

Para calcular os autovalores e autovetores de  $\Sigma$ , deve-se resolver a equação  $\Sigma\Phi = \lambda_i\Phi$ , onde  $\lambda_i$  e o  $i$ -ésimo valor próprio de  $\Sigma$  ( $\lambda_i \geq \lambda_{i+1}$ ) e  $\Phi^T\Phi$ . Sabe-se que os vetores próprios da matriz de covariâncias,  $\Sigma$ , correspondentes aos valores próprios mais elevados descrevem a maior parte das variações admissíveis, e a proporção da variância total explicada por cada vector próprio é igual ao valor próprio correspondente. Assim, a maior parte da variância da forma pode ser explicada por um pequeno número de vetores próprios  $t$ , chamados modos de variação. Desta forma, cada objeto do conjunto de treino pode ser descrito pela forma média e pela combinação dos primeiros  $t$  vetores próprios obtidos:  $\bar{x} = x + \Phi_f b_f$ , onde  $\Phi_f = (\Phi_{f1}\Phi_{f2}\dots\Phi_{ft})$  representa a matriz dos primeiros  $t$  autovetores das formas e  $b_f = (b_{f1}b_{f2}\dots b_{ft})$  é o vetor de pesos de cada autovetor. Os autovetores são ortogonais e portanto  $\Phi_f^T\Phi_f = I$  e  $b_f = \Phi_f^T(x - \bar{x})$  (Cootes e Taylor, 2004).

A figura 1 mostra exemplos de formas faciais. Qualquer nova forma  $x$  é uma deformação da forma média  $\bar{x}$ , dado pela combinação linear dos autovetores  $\Phi_f$  e o vetor de pesos  $b_f$ . Variando-se os valores do vetor  $b_f$ , pode-se modelar linearmente outras formas diferentes em relação ao conjunto de treinamento.

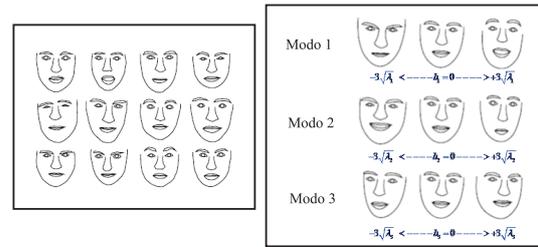


Figura 1: Exemplos de formas faciais do conjunto de treinamento. Adaptado de Cootes e Taylor (2004).

Como a variância de  $b_f$  sobre o conjunto de treino é dada por  $\lambda_i$ , os seus limites estão no intervalo  $-3\sqrt{\lambda_i} \leq b_{fi} \leq 3\sqrt{\lambda_i}$ , dado que, a população está concentrada entre três desvios padrões da média (Cootes *et al.*, 1995).

## 5 Resultados

O programa para o sistema de visão proposto aqui foi desenvolvido em linguagem *C++* e testado em

um computador com sistema operacional *Windows 7*.

Primeiramente foi implementado um detector de faces e o seu rastreamento em tempo real usando uma combinação dos métodos de *Haar Cascades* e *Camshift*.

Na Figura 2 é apresentado o resultado da detecção de faces utilizando *características de Haar* em combinação com o método de *Camshift*. A téc-



Figura 2: Detecção e rastreamento de face.

nica para a extração de expressões envolve a criação de modelos **ASM** treinados em imagens de faces (45 pessoas, entre homens e mulheres para o treinamento) tomadas de um banco de dados de disponibilidade pública apresentada em Lucey *et al.* (2010), que contém 123 pessoas, cada uma apresentando sete expressões diferentes, cada imagem é marcada com a emoção correspondente. Para algumas pessoas no banco de dados, nem todas as emoções estão disponíveis. Nesta etapa de treinamento são marcados os pontos que definem o contorno da face, para isto foi usada a livreria *asmlibrary*. A figura 3 mostra o processo de alocação dos pontos fiduciais formando o **ASM**.

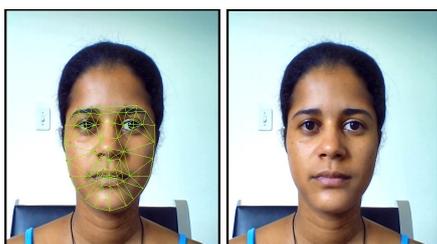


Figura 3: Alocação de pontos fiduciais usando o modelo **ASM**.

Uma vez alocados os pontos na face, identificam-se as unidades de ação facial propostas por Ekman e Friesen (1978) para representar cada expressão utilizando a livreria *OpenCV*. Em Cootes *et al.* (1995), é mostrada a numeração na alocação destes pontos, que pode ser vista na figura 4.

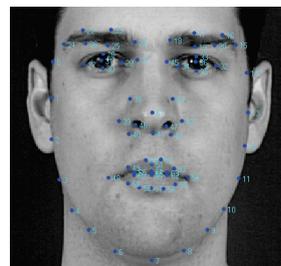


Figura 4: Alocação de pontos fiduciais enumerados usando **ASM**.

As figuras 5 e 6 mostram o resultado da detecção e o reconhecimento das expressões faciais anteriormente nomeadas. Foram realizados experimentos durante a implementação e na fase de testes, utilizando uma amostra (35 pessoas) da base de dados nomeada anteriormente. O método proposto obteve uma taxa média de detecção de 82.25%, calculada usando o as taxas de Falsa Aceitação (FAR) e Falsa Rejeição (FRR) calculadas como apresentadas em Ribeiro *et al.* (2010), na tabela 1 podem ser vistas as taxas de reconhecimento para cada expressão.

Tabela 1: Resultados da classificação por expressão facial.

expressão facial	Taxa de Reconhecimento
Confused	74%
Surprised	76%
Happy	88%
Neutral	91%

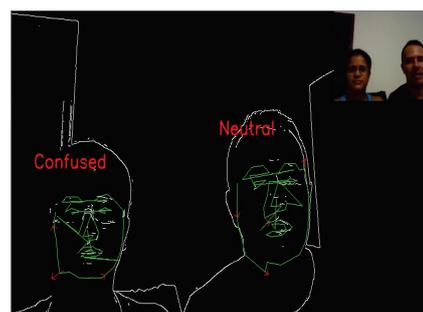


Figura 5: Resultados do reconhecimento de expressões faciais.

## 6 Comentários Finais

Neste artigo são apresentados os resultados preliminares do desenvolvimento de um sistema de percepção visual para interação humano-máquina que permite fornecer informação comportamental (emocional) no auxílio a cuidadores de pessoas

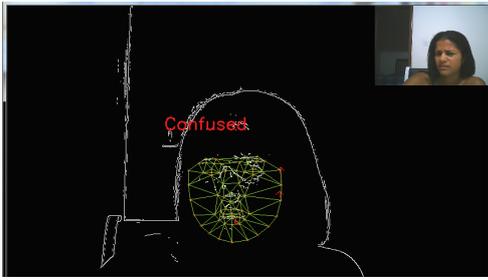


Figura 6: Resultados do reconhecimento de expressões faciais.

com autismo, este sistema inclui a detecção, rastreamento de faces usando os métodos de *Viola-Jones* e *Camshift*, e reconhecimento de expressões faciais usando modelos de forma ativa (**ASM**), em tempo real.

O sistema completo de reconhecimento e detecção desenvolvido neste trabalho mostrou-se eficiente, conseguindo trabalhar em tempo real com uma definição de  $640 \times 480$  pixels. Também é importante anotar que sistemas de interação baseados em Visão Computacional podem ser implementados em computadores convencionais utilizando como dispositivos de captura e interação câmeras **USB** convencionais e de baixo custo.

Como trabalhos futuros estão adicionar mais expressões faciais, combinar os **ASM** com um classificador como máquinas de vetores de suporte - **SVM**, por exemplo, e usar métodos de detecção de novidades com a classificação dos padrões.

### Agradecimentos

Os autores deste artigo gostariam de agradecer às instituições a seguir o apoio dado ao desenvolvimento desta pesquisa: **FAPESB** - Fundação de Amparo à Pesquisa do Estado da Bahia; **CNPq** - Conselho Nacional de Desenvolvimento científico e Tecnológico; e **CAPES** - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

### Referências

- Almejrada, A. S. (2010). Human emotions detection using brain wave signals: A challenging. *European Journal of Scientific Research* **44**(4), 640–659.
- Azcarate, A., F. Hageloh, K. V. D. Sande e R. Valentini (2005). Automatic facial emotion recognition. *Universiteit van Amsterdam* p. 110.
- Bahonar, H. e M. A. Charkari (2009). Facial feature detection and extraction using symmetry and region-based deformable template matching.. Em: *14th International Computer Conference*. Tehran, Iran. pp. 664–669.
- Bradski, G. R. (1998). Computer video face tracking for use in a perceptual user interface. *Intel Technology Journal Q2* **28**, 115.
- Cootes, T. F., C. J. Taylor, D. H. Cooper e J. Graham (1995). Active shape modelstheir training and application. *Computer Vision and Image Understanding* **61**(1), 38–59.
- Cootes, T. F. e C. J. Taylor (2004). Statistical models of appearance for computer vision.
- Ekman, P. e W.V. Friesen (1978). *Facial Action Coding System: a Technique for the Measurement of Facial Movement*. Consulting Psychologists Press. Palo Alto.
- Freund, Y. e R. E. Schapire (1995). A decision-theoretic generalization of on-line learning and an application to boosting.
- Lucey, P., J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar e I. Matthews (2010). The extended cohn-kande dataset (*ck+*): A complete facial expression dataset for action unit and emotion-specified expression. Em: *Third IEEE Workshop on CVPR for Human Communicative Behavior Analysis*. San Francisco, CA, USA.
- Ribeiro, I., G. Chiachia e A. N. Marana (2010). Reconhecimento de faces utilizando análise de componentes principais e a transformada census. Em: *VI Workshop de Visão Computacional*. pp. 25–30.
- Salcedo, C. M., C. A. Peña, J. J. Cerqueira e A. M. N. Lima (2012a). Designing a real time artificial vision system for human interaction with an omnidirectional mobile platform. Em: *Robotics Symposium and Latin American Robotics Symposium SBR/LARS*. pp. 21–26.
- Salcedo, C. M., J. J. Cerqueira e A. M. N. Lima (2012b). Sistema de visão para interação homem-robô. Em: *XIX Congresso Brasileiro de Automática*. pp. 4873–4880.
- Viola, P. e M. Jones (2004). Robust real-time face detection. *Computer Vision* **57**(2), 137–154.
- Zhao, X., H. Zhang e Z. Xu (2012). Expression recognition by extracting facial features of shapes and textures. *Journal of Computational Information Systems* **8**(8), 3377–3384.