

## FOURFACE: Uma ferramenta de reconhecimento de expressões faciais

Ailton Lopes de Sousa, Saulo William S. Costa, Yomara Pires, Fabiola Araújo

Laboratório de Desenvolvimento de Sistemas (LADES) – Faculdade de Computação –  
Universidade Federal do Pará – Campus de Castanhal, Avenida dos Universitários, Jaderlândia –  
Castanhal – PA – Brasil

ailtoonlopes@gmail.com, saulo.si.costa@gmail.com,  
yomara@ufpa.br, fpoliveira@ufpa.br

***Abstract.** Through studies of the areas of Affective Computing, this paper presents the results of the development of a computational tool capable of capturing the facial expressions of the user using the Kinect in order to carry out the development of scenarios to evaluate the affection of individuals by means of capturing the personality traits in facial expressions.*

***Resumo.** Através dos estudos das áreas de Computação Afetiva, este trabalho visa apresentar os resultados do desenvolvimento de uma ferramenta computacional capaz de capturar as expressões faciais do usuário utilizando-se do Kinect, a fim de realizar a elaboração de cenários para avaliação da afetividade de indivíduos por meio da captura dos traços de personalidade em expressões faciais.*

### 1. Introdução

Reconhecer emoções humanas por meio do computador é um dos desafios da Computação Afetiva. Dotar a máquina de emoções ou simplesmente identificá-las através de sistemas computacionais nem sempre é uma tarefa simples. Vários métodos de áreas diferentes como os da psicologia, antropologia, visão computacional entre outros, são necessários para que se possa alcançar, por meio de um software, aspectos sutis e inerentes à personalidade humana.

Levando em consideração o contexto acima, segundo Friedlander e Rahmilevtz [2010], expressões faciais são uma das mais poderosas, naturais e imediatas formas para os seres humanos comunicarem suas emoções e intenções. O reconhecimento facial consiste basicamente em identificar padrões em características faciais como formato de boca, do rosto, distância dos olhos, entre outros. Assim, um humano é capaz de reconhecer uma pessoa familiar mesmo com muitos obstáculos como distância, sombras ou apenas a visão parcial do rosto. Uma máquina, no entanto, precisa realizar inúmeros processos para detectar e reconhecer um conjunto de padrões específicos para rotular uma face como conhecida ou desconhecida [Silva; Cintra; 2015].

Desse modo, este estudo busca o desenvolvimento de uma ferramenta (FourFace) que seja capaz de fazer o reconhecimento biométrico e comportamental e que possam contribuir para o estudo da afetividade em ambientes computacionais, permitindo o uso de técnicas que vão além do campo computacional como a psicologia. Em consequência disso, haverá uma grande contribuição para o aumento da coerência, consistência e credibilidade das reações e respostas computacionais que derivam da interação humana via interface humano-computador.

## 2. Metodologia

Para poder tornar possível a identificação destes aspectos faciais e emocionais, foram utilizados, como referência, os estudos realizados a respeito do Sistema FACS: estudos da codificação das ações faciais que analisam individualmente cada pequena alteração descrita na face. Neste sistema, existem as AU's (Unidades de Ações) que são as peças usadas para ilustrar a face humana; os movimentos descritos nas AU's são causados pela atuação de músculos (contração e relaxamento) [Ekman et al., 2002]. Algumas das principais unidades de ações podem ser encontradas na Figura 1:

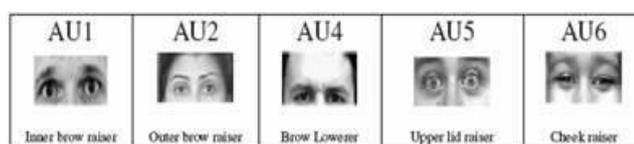


Figura 1. Principais unidades de animação – Sistema FACS. [Ekman et al., 2002]

A expressão facial é de fato a combinação de AU's. A princípio, esta abordagem possibilita que uma expressão facial seja entendida como uma combinação de unidades de ações faciais relevantes. Abaixo, listamos as AU's do Sistema FACS combinadas que tornam possível a ocorrência de uma expressão facial.

Tabela 1. Associação das quatro expressões faciais com seus AU's e combinações [Diniz et al., 2013]

EXPRESSÃO	UNIDADES DE AÇÕES PRINCIPAIS					UNIDADES DE AÇÕES AUXILIARES			
	AU6	AU12				AU25	AU26	AU16	
Alegria	AU6	AU12				AU25	AU26	AU16	
Surpresa	AU5	AU26	AU27	AU1+2					
Tristeza	AU1	AU15	AU17			AU4	AU7	AU25	AU26
Raiva	AU2	AU4	AU7	AU23	UA24	AU17	AU16	AU25	AU26

Para a detecção das expressões, o SDK do Kinect na versão 1.8 fornece 6 (seis) unidades de animação (UA's), que tiram partido das unidades de ação dos estudos apresentados anteriormente [Ekman et al., 2002] para modelar facilmente qualquer expressão facial. Algumas das unidades de animação que o Kinect suporta são apresentadas na figura 2:

Valor e Nome da AU	Interpretação
<i>AU0 – Upper Lip Raiser</i> (CANDIDE-3 : AU10)	0 = neutro, a cobrir o lábio;
	1 = a mostrar os dentes completamente;
	-1 = lábio puxado para baixo, o máximo possível.
<i>AU1 – Jaw Lowerer</i> (CANDIDE-3 : AU26/27)	0 = maxilar fechado;
	1 = maxilar completamente aberto;
	-1 = maxilar fechado, o mesmo que 0.

**Figura 2. Unidades de animação rastreadas pelo FaceTracking SDK. [Correa, 2013]**

Cada UA (Unidade de Animação) é expressa como um valor numérico compreendido que varia entre -1 e 1. Mostra-se também, na Tabela acima, uma pequena explicação da semelhança dos valores de cada unidade de animação com o efeito que essa unidade causa.

Realizou-se uma comparação com os dados mais aproximados disponíveis com o *Face Tracking* SDK (Unidades de Animação – CANDIDE3), em versão C# com as AU'S do Sistema FACS, a fim de atribuir emoções para cada unidade de animação do SDK do Kinect. O exemplo desta comparação pode ser observado na Tabela 2.

**Tabela 2. Equivalência de AU's do FACS com as do FaceTracking  
(Expressão de Alegria)**

FACS		FACE TRACKING SDK		
AU	DESCRIÇÃO	UA	DESCRIÇÃO	
6	<i>Cheek Raiser</i>	0	<i>Upper Lip Raiser</i>	Alegre
12	<i>Lip Corner Puller</i>	4	<i>Lip Corner Depressor</i>	
1	<i>Inner Brow Raiser</i>			Surpresa
2	<i>Outer Brow Raiser</i>	5	<i>Outer Brow Raiser</i>	
5B	<i>(Slight) Upper Lid Raiser</i>			
26	<i>Jaw Drop</i>	1	<i>Jaw Lower</i>	
1	<i>Inner Brow Raiser</i>			Triste
4	<i>Brow Lowerer</i>	3	<i>Brow Lowerer</i>	
15	<i>Lip Corner Depressor</i>	0	<i>Upper Lip Raiser</i>	
17	<i>Chin Raiser</i>			
2	<i>Outer Brow Raiser</i>			Raiva
4	<i>Brow Lowerer</i>	3	<i>Brow Lowerer</i>	
7	<i>Lid Tightener</i>			
23	<i>Lip Tightener</i>	2	<i>Lip Stretcher</i>	
24	<i>Lip Pressor</i>			

Sabendo-se que o valor que cada Unidade de Animação se altera entre -1 e 1, de acordo com o mover dos músculos do rosto, é possível realizar uma análise desses valores a fim de poder inferir através dos valores a ocorrência de determinada expressão emocional.

O último passo deste processo é criar limites nos valores das unidades de

animação, de forma que, assim que eles forem ultrapassados, o sistema reconhecerá uma determinada expressão facial, tendo em conta os resultados dos valores das unidades de animação na expressão facial do utilizador. Estes limites foram definidos de forma experimental, assim foi permitido que por meio de comparações entre o valor da unidade animação e os limites atribuídos para cada expressão, o aplicativo possa verificar se as condições definidas para cada expressão são válidas e, dessa forma, permitir inferir a ocorrência da mesma. Os limites definidos para quatro emoções básicas (alegria, surpresa, tristeza e raiva) estão descritos na tabela 3.

**Tabela 3 – Valores (condições) de cada UA para que as expressões sejam válidas.**

<b>EXPRESSÃO</b>	<b>CONDIÇÕES PARA DETECTAR EXPRESSÃO</b>
Alegria	$UA[0] > 0.2$ e $UA[4] < -0.2$
Surpresa	$UA[1] > 0.4$ e $UA[5] > 0.1$
Tristeza	$UA[0] < -0.1$ e $UA[3] > 0.1$
Raiva	$UA[3] > 0.2$ e $UA[2] < -0.2$

Para exemplificar o funcionamento da inferência da expressão facial baseada nos valores das UA's, é necessário que os valores das UA's respectivas de cada emoção estejam dentro dos limites estabelecidos, como na demonstração da seguinte expressão:

**Expressão facial alegre =  $UA0 > 0.7$  e  $UA4 < 0.3$**

### 3. Desenvolvimento

O modelo abaixo representa o processo que a ferramenta irá desempenhar para que o objetivo deste trabalho seja alcançado. Dessa forma, o sistema utilizará as imagens que serão capturadas pela câmera do Kinect [Microsoft Research, 2011], as quais serão usadas como entrada para o módulo de Rastreamento das características faciais, além disso, será feito uso de outras técnicas chamadas Viola-Jones [Viola and Jones, 2001] que faz a detecção do rosto do indivíduo e da técnica CANDIDE-3 [Ahlberg, 2001] que realiza a detecção das características faciais ao encontrar os pontos característicos do rosto de determinados locais de interesse (boca, olhos, sobrancelhas e nariz). Essas técnicas foram implementadas utilizando-se dos recursos da Biblioteca do Kinect – SDK.

Depois de realizada a captura de imagem e o rastreamento das características faciais por meio das técnicas acima apresentadas, tem-se como próxima etapa o módulo para a classificação das expressões que permitirá a inferência das emoções do usuário.

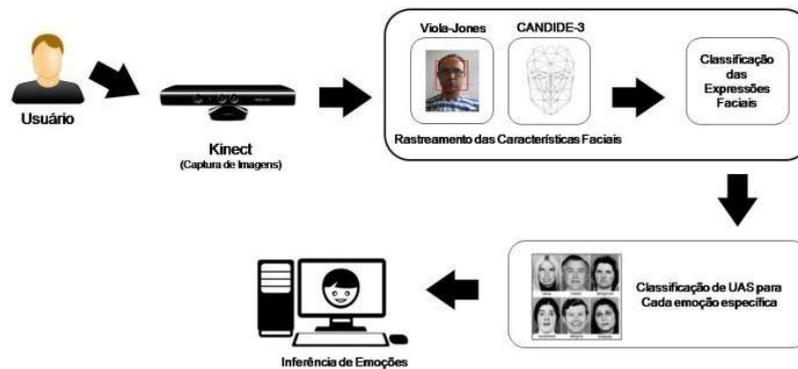


Figura 3. Processo de Funcionamento da Ferramenta

#### 4. Experimentos e Resultados

Diante disso, obteve-se um aplicativo de rastreamento de face para detectar quatro tipos de expressões (alegria, surpresa, tristeza e raiva) de acordo com os movimentos musculares associados às regiões locais de interesse. O funcionamento da aplicação é apresentado na Figura 4 e 5 onde se pode ver o software reconhecendo as expressões faciais desejadas, a partir das técnicas mencionadas neste trabalho e com o uso das unidades de animação do Face Tracking SDK – Kinect.

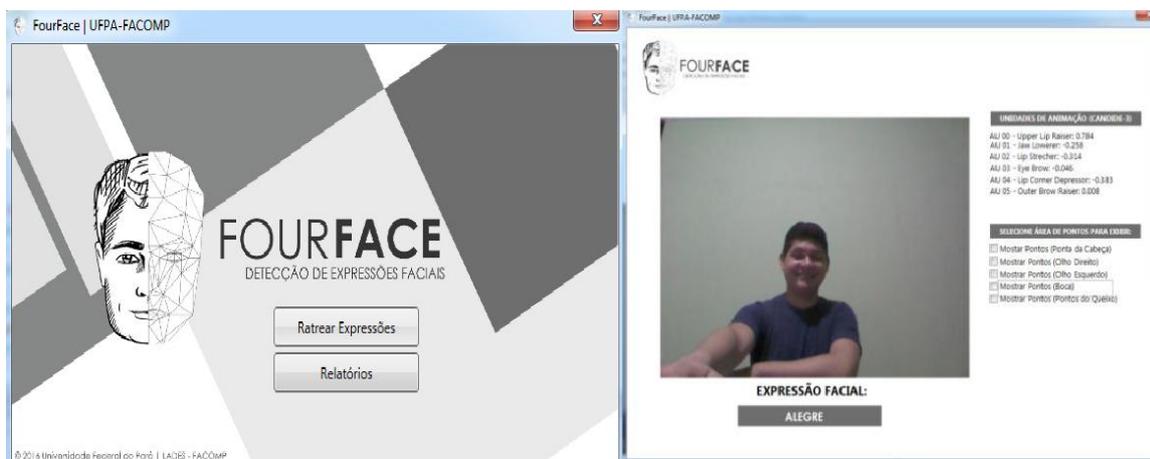


Figura 4. Interfaces Gráficas da ferramenta.

Para realização dos experimentos referentes ao *FourFace*, o seguinte cenário: seleção de 20 voluntários com idade, sexo e cor de pele distinta, um sensor Kinect para captura das faces, o uso de imagens que represente as emoções faciais desejadas para mostrar ao voluntário qual a simulação da expressão que deseja analisar.

Em tempo real, cada voluntário simulou cerca de 20 emoções (5 de alegria, 5 de raiva, 5 de tristeza, 5 de desgosto) sendo um total de 400 emoções simulada pelos voluntários. A análise dessas imagens foi realizada sobre variações de luzes e alguns indivíduos possuíam barbas, bigode ou usavam óculos. Os indivíduos eram de ambos os sexos com idades na faixa de 7 a 55 anos.

Como desempenho total da ferramenta proposta neste trabalho, obteve-se um resultado de 67,9% no rastreamento de todas as 400 expressões analisadas no teste, conforme Tabela 4. Abaixo é mostrada uma análise individual de cada expressão

rastreada e as dificuldades na fase de teste.

**Tabela 4 – Resultado de precisão das expressões faciais.**

<b>EXPRESSÃO</b>	<b>TAXA DE ACERTO</b>
Alegria	81.37%
Surpresa	60.54%
Tristeza	60.24%
Raiva	69.52%

## 5. Análise de Resultados

Na Tabela 4 é apresentado o desempenho de reconhecimento de cada expressão facial, do qual é possível fazer as seguintes análises: a) a expressão alegria foi a que mais obteve acerto em comparação com as demais expressões, uma taxa de acerto de 81,37%. b) para as expressões raiva e tristeza já se esperava um resultado aproximado entre as duas, pois estas possuem similaridades entre si, como a UA 03 (Brow Lowerer) e aproximando-se de uma taxa de acerto de 69,52 e 60,24 respectivamente. c) a expressão facial surpresa obteve uma taxa de acerto de 60,54, valor muito baixo do esperado, de acordo com testes preliminares onde a taxa de acerto foi bem maior.

Durante os experimentos na ferramenta, pessoas distintas foram selecionadas. Foi identificado que o Kinect possui dificuldades em reconhecer pessoas com a pele mais escuras, levando um tempo maior para a detecção da própria face, assim, também, como o uso de acessórios (óculos, por exemplo) e barba. O posicionamento para a câmera do aparelho deve sempre permanecer frontalmente, o deslocamento para outras posições inibe o reconhecimento dos algoritmos VIOLA-JONES e CANDIDE-3. Outra característica que implica no reconhecimento são as variações de luzes que dificultam diretamente no reconhecimento da expressão. Quanto ao comportamento dos voluntários, foi percebido que uma boa parcela não conseguia representar uma determinada expressão, como no caso da expressão surpresa, onde um grande número de voluntários não conseguiu atingir os limites definidos na tabela 3, isso se justifica pelo fato de nem todos voluntários se sentirem confortável em expressar determinada emoção, que exigia o posicionamento da boca aberta. Algumas com deslocamentos mais marcado e comuns, como a alegria, foi facilmente representado, pois a precisão com as UAs, e por conta da fidelidade e da facilidade de representação da emoção, é maior.

Durante o teste cada expressão foi simulada 5 vezes por cada voluntário, obtendo-se 100 expressões simuladas para cada expressão facial. Como o reconhecimento é feito em tempo real, ao tentar simular o deslocamento para determinada expressão, características de outras expressões também são detectadas quando suas condições são verdadeiras.

## 6. Conclusão

Apresentamos neste trabalho, os resultados do desenvolvimento de um software para monitoramento de expressões do rosto, levando em consideração a detecção de quatro expressões faciais (alegria, raiva, surpresa e tristeza). Os resultados gerados através da aplicação podem ser usados como métrica para aperfeiçoamento de qualquer

ferramenta, na qual o usuário tem interação direta com o computador. Após testes experimentais a ferramenta FourFace obteve uma taxa total de acurácia de 67,9 %, resultado este considerado satisfatório para o emprego da técnica utilizada que reconhece expressões faciais em tempo real.

A proposta de fazer o reconhecimento de 4 emoções faciais utilizando técnicas de reconhecimento facial, foi provada através dos experimentos, a qual se mostrou eficaz na comprovação dos resultados. Sabe-se, no entanto, que somente o uso das técnicas apresentadas neste trabalho não são suficiente para que se tenha uma ferramenta de reconhecimento de expressões faciais tão precisa quanto o reconhecimento humano. O uso de métodos mais avançados que utilizem comparações de padrões, por exemplo, podem otimizar os resultados até então encontrados.

Como trabalhos futuros e com a necessidade de aperfeiçoar a ferramenta, deseje-se o emprego de uma abordagem utilizando métodos híbridos, onde poderão fazer a classificação das emoções, como o método de Análise de Componentes Principais-PCA, Análise de Componentes Independentes – ICA e/ou Redes Neurais, além da construção de uma base dados maior que as usadas nos experimentos vistos até aqui. Também serão feitas uma análise sobre o emprego das técnicas de reconhecimento de expressões por meio das novas versões do Kinect ou até mesmo de uma câmera convencional, saindo da dependência do sensor.

## Referências

- Ahlberg, J. C. (2001) “CANDIDE-3 – an updated parameterized face”. Report No. LiTH-ISY- R2326, Dept. of Electrical Engineering, Linköping University, Sweden.
- Costa, Saulo William S., Ailton Lopes de Sousa, and Yomara Pires. (2015) "Computação Afetiva: Uma ferramenta para avaliar aspectos afetivos em aplicações computacionais".
- Correia, Miguel Medeiros.(2013) “Reconhecimento de elementos da língua gestual portuguesa com Kinect”.
- Diniz, Fábio Abrantes, et al. (2013) "Um sistema de reconhecimento facial aplicado a um ambiente virtual de aprendizagem composto por agentes pedagógicos."ProceedingsofInternationalConferenceonEngineeringand Computer Education. Vol. 8.
- Ekman, P.; Friesen, W. V.; Hager, J. C. (2002) “Facial Action Coding System: Investigator’s guide”. Research Nexus division of Network Information Research Corporation, Salt Lake City, Estados Unidos .
- Friedlander, M.; Rahmilevitz, G. Software Para Análise de Emoções na Face. Trabalho de conclusão de Curso, Escola Politécnica da Universidade de São Paulo, 2010.
- Microsoft Research. (2011) “Programming Guide: Getting Started with the Kinect for Windows SDK Beta”. Disponível em [http://research.microsoft.com/en-us/um/redmond/projects/kinectsdk/docs/ProgrammingGuide\\_KinectSDK.docx](http://research.microsoft.com/en-us/um/redmond/projects/kinectsdk/docs/ProgrammingGuide_KinectSDK.docx). Acesso em 30 de Setembro de 2016.
- Picard, R. W. (1997) “Affective Computing”. Cambridge, EUA: The M.I.T. Press.

Prates, R. O. and Barbosa, S. D. J. (2016) “Avaliação de Interfaces de Usuário Conceitos e Métodos”. Disponível em [http://homepages.dcc.ufmg.br/~rprates/ge\\_vis/cap6\\_vfinal.pdf](http://homepages.dcc.ufmg.br/~rprates/ge_vis/cap6_vfinal.pdf). Acesso em 23 de Setembro de 2016.

Silva, Alex Lima, and Marcos Evandro Cintra. "Reconhecimento de padrões faciais: Um estudo."

Sousa, Ailton Lopes de et al. Reconhecimento de expressões faciais e emocionais como método avaliativo de aplicações computacionais. Encontro Regional de Computação e Sistemas de Informação-ENCOSIS, 2016.

Viola, P. and JONES, M. (2001) “Robust real-time object detection”. Technical report, University of Cambridge.