

Um Sistema de Reconhecimento de Expressões Faciais para Apoiar um Ambiente Virtual de Aprendizagem

Fábio Abrantes Diniz¹, Francisco Milton Mendes Neto¹, Francisco das Chagas Lima Júnior¹, Laysa Mabel de O. Fontes¹, Thiago de Aguiar Leal Domingues², Lucas Silva Figueiredo²

¹Programa de Pós-Graduação em Ciência da Computação – PPgCC
Universidade do Estado do Rio Grande do Norte – UERN
Universidade Federal Rural do Semiárido – UFERSA
BR 110 – Km 46 – Bairro Costa e Silva – Campus Central
59.625-620 Mossoró – RN, Brasil

²Centro de Informática – Universidade Federal de Pernambuco – UFPE
Cidade Universitária – 50740-560, Recife – PE, Brasil

{fabio.abrantes.diniz, thiago.a.l.d, lsf.cin}@gmail.com,
{miltonmendes, laysa}@ufersa.edu.br, limajunior@uern.br

Abstract. *Currently, the majority of Virtual Environments for Learning (VLEs) lack of efficient mechanisms that can identify affective phenomena throughout the student's interaction with the environment, since these phenomena deeply interfere in mental processes such as reasoning and, especially, motivation. To fill this gap, this paper proposes a system for recognizing facial expressions based on System Facial Action Coding System (FACS) and a method of classification rules based on emotion. This system, by means of a camera, investigates their performance in the recognition of emotion (happiness, sadness, anger and disgust) student, verifying his facial expression during the entire course in a VLE.*

Resumo. *Atualmente, a maioria dos Ambientes Virtuais de Aprendizagens (AVAs) não dispõem de mecanismos eficientes que possam identificar os fenômenos afetivos durante toda a interação do aluno com o ambiente, visto que esses fenômenos interferem profundamente nos processos mentais, como raciocínio e, principalmente, motivação. Para preencher esta lacuna, este trabalho propõe um sistema de reconhecimento de expressões faciais baseado no Sistema de Codificação de Ação Facial (FACS) e no método de classificação da emoção baseado em regras. Este sistema, por meio de uma câmera, investiga o seu desempenho no reconhecimento das emoções (alegria, tristeza, raiva e desgosto) do aluno, verificando sua expressão facial durante toda a realização do curso em um AVA.*

1. Introdução

O aumento da demanda pelos Ambientes Virtuais de Aprendizagens (AVAs) e os avanços das tecnologias de informação e comunicação têm gerado a necessidade de se melhorar estratégias na identificação dos estados afetivos do aluno durante todos os processos de aprendizagem e avaliativos do AVA. A maioria dos AVAs não dispõe de

recursos eficientes e automatizados que identificam os estados emocionais (por exemplo, as emoções de raiva, alegria, tristeza e desgosto,) do aluno durante toda sua interação com o AVA.

A verificação da afetividade do aluno é desafiante. É foco das pesquisas da área de Computação Afetiva [Sun; Lu, 2010]. A Computação Afetiva tenta fazer com que a emoção, existente na comunicação entre pessoas, esteja presente também durante a interação entre homem e computador. O computador deve entender as emoções humanas e/ou expressar afeto [Sun; Lu, 2010]. São várias as formas de identificar as emoções de um indivíduo. Dentre as tecnologias mais pesquisadas, os reconhecimentos de expressões faciais destacam-se por sua aceitabilidade por parte das pessoas e facilidade de coleta, mesmo de forma não colaborativa [Jain; Ross; Prabhakar, 2004].

Neste intuito, este trabalho projetou um sistema de reconhecimento de expressões faciais, chamado RedFace, que infere, em tempo real, as emoções alegria, tristeza, raiva e desgosto, utilizando imagens sequenciais capturadas pela câmera do Kinect [Microsoft Research, 2011]. A emoção reconhecida servirá de entrada para um AVA, que é composto por agentes de software que auxiliam na aplicação correta da teoria da Aprendizagem Baseada em Problema (*Problem-Based Learning* - PBL) [Fontes *et al.*, 2012], a fim de oferecer maior motivação e entusiasmo nos processos de aprendizagem do aluno.

Para isso, a aplicação desenvolvida neste trabalho utilizou a técnica de Viola-Jones [Viola; Jones, 2004] para a detecção da face, e a técnica de CANDIDE [Ahlberg, 2001] no processo de rastreamento dos pontos característicos da face. Na classificação da emoção, baseou-se no sistema psicológico de codificação facial FACS [Ekman *et al.*, 2002]. O método de classificação da emoção consiste em regras que buscam verificar a deformação geométrica da boca, dos olhos e das sobrancelhas. Os resultados parciais obtidos mostram que o sistema de reconhecimento de expressões faciais proposto pode ser aplicado com sucesso na identificação da emoção dos alunos durante todo o período de realização de um curso do AVA.

Este trabalho está organizado da seguinte forma: a Seção 2 descreve o AVA baseado em agentes no qual o sistema de reconhecimento facial desenvolvido foi aplicado; a Seção 3 discute os trabalhos relacionados; a Seção 4 apresenta o sistema de reconhecimento de expressões faciais proposto, mostrando suas etapas de implementação; a Seção 5 apresenta uma análise dos resultados; e, por fim, na Seção 6, são apresentadas as considerações finais e as propostas de trabalhos futuros.

2. Ambiente Virtual de Aprendizagem Baseado em Agentes de Apoio à PBL

O AVA baseado em agentes de apoio à PBL [Fontes *et al.*, 2012], no qual o RedFace foi aplicado, está esquematizado na Figura 1. Ele utiliza a arquitetura cliente-servidor padrão do Moodle [França; Soares, 2011]. No lado cliente, são executadas as páginas em PHP e o Agente Pedagógico Animado (AgPA), os quais solicitam serviços ao servidor da aplicação, através da Internet [FONTES *et al.*, 2012]. No lado servidor localiza-se o sistema gerenciador de banco de dados (SGBD), onde são armazenados todos os dados referentes aos cursos, aos alunos cadastrados (inclusive suas imagens faciais), além de quatro tipos de agentes: Agente Detector de Problemas (AgDP),

Na educação, poucos trabalhos abordando reconhecimento de expressões faciais foram encontrados. Um dos trabalhos relacionados foi proposto por Changjun *et al* (2011). Eles fizeram um sistema de reconhecimento de expressões faciais em tempo real utilizando um classificador de máquina de vetores de suporte (*Support Vector Machines* - SVM), com o objetivo de estimar o estado afetivo dos alunos que utilizam um sistema de Ensino a Distância (EaD) baseado em computação afetiva. Um rastreador de faces reúne um conjunto de características de deslocamentos de movimento da face no sistema de EaD. Estas características são posteriormente utilizadas para treinar um classificador SVM para aprender a reconhecer expressões emocionais. Em seguida, o estado emocional é avaliado e compreendido.

Changjun *et al* (2011) fizeram treinamentos do algoritmo SVM com 100 imagens faciais de 15 alunos. Estas foram selecionadas aleatoriamente a partir do banco de dados, incluindo seis emoções básicas e neutras como um conjunto de treinamento. Depois utilizou 265 imagens faciais de 42 pessoas, como conjunto de teste. O modelo de sistema de EaD, proposto por Changjun *et al* (2011), com base em computação afetiva atingiu uma taxa de reconhecimento total de 84,55%.

No trabalho proposto, apresenta-se um sistema de reconhecimento de expressões faciais que inferi as emoções alegria, tristeza, raiva, e desgosto. Para isso, utilizou o sistema psicológico de codificação facial FACS, o qual rastreia todos os movimentos musculares faciais nas suas respectivas Unidades de Ações (UAs) [Ekman *et al.*, 2002]. Foram aplicadas regras que verificam a deformação geométrica da boca, olhos e sobrancelhas para reconhecer as UAs presentes nas expressões faciais das emoções inferidas. Para a detecção da face e rastreamentos dos pontos característicos da face foram utilizados a técnica Viola-Jones [Viola; Jones, 2004] e o modelo CANDIDE-3 [Ahlberg, 2001].

4. Sistema de Reconhecimento de Expressões Faciais

O sistema de reconhecimento de expressões faciais proposto foi desenvolvido no padrão Cliente/Servidor de modo a encapsular o acesso ao AVA proposto por Fontes (2012). Esse sistema consiste basicamente das seguintes etapas: (a) Capturação das imagens; (b) Rastreamento das características faciais; (c) Classificação da expressão facial; (d) Inferência da emoção. A arquitetura do sistema de reconhecimento da expressão facial é apresentada na Figura 2.

Este sistema infere algumas das emoções básicas (raiva, desgosto, alegria e tristeza) proposto por Ekman (2002) captadas pelas expressões faciais, usando imagens provenientes da câmera do Kinect. Essas emoções inferidas servirão como entrada para um AVA composto por um agente pedagógico projetado para proporcionar uma maior motivação e entusiasmo nos processos de ensino-aprendizagem [Fontes *et al.*, 2012].

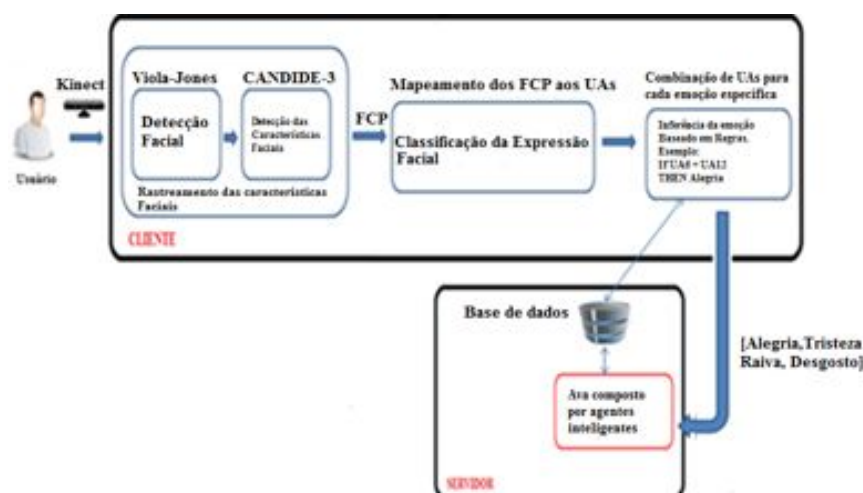


Figura 2. Arquitetura do sistema de reconhecimento de expressão facial.

O início do processo de reconhecimento das expressões faciais se dá a partir da aquisição da imagem. Utilizando a câmera do Kinect [Microsoft Research, 2011], o sistema captura os quadros (*frames*) do vídeo de um indivíduo, em tempos predefinidos (a cada 5 segundos), os quais serão utilizados como entrada para o módulo de rastreamento das características faciais.

No módulo de rastreamento das características faciais foi utilizada a técnica de detecção de face *Viola-Jones* [Viola; Jones, 2001] e a técnica de detecção das características faciais *CANDIDE-3* [Ahlberg, 2001], que encontra os Pontos Característicos da Face (FCPs) dos locais de interesse da face (boca, olhos, sobrancelhas e nariz). Ambas as técnicas foram implementadas, utilizando os recursos disponíveis na biblioteca do Kinect (Kinect SDK) [Microsoft Research, 2011].

A escolha do algoritmo *Viola-Jones* na detecção da face foi pelo motivo de possuir uma detecção facial bastante rápida em relação aos outros métodos existentes na literatura. Essa característica é essencial para aplicações que tem resposta imediata ao usuário, como é o caso da identificação dos alunos nos AVAs os quais demandam processamento próximo do tempo real. Já o modelo de *CANDIDE-3* por ser uma técnica de face parametrizada especificamente desenvolvida para codificação de faces humanas baseada em modelo, além de ser bastante usada em vários artigos [Dornaika; Ahlberg, 2004]. O *CANDIDE-3* é ajustado à face do indivíduo no primeiro quadro (*frame*) de vídeo e acompanhada até ao final do mesmo, enquanto a expressão facial vai sofrendo evoluções, pois é controlado pelas várias UAs. Mais detalhes da técnica *CANDIDE* encontra-se no trabalho proposto por Ahlberg (2001).

A combinação das duas técnicas (*CANDIDE-3* e *Viola-Jones*) atinge um nível aceitável de qualidade na detecção e rastreamento de imagens faciais em diferentes posições, atingindo resultados abaixo de 1 segundo com o uso do Kinect [Microsoft Research, 2011]. Logo, essa combinação é essencial para aplicações *web* que tem resposta imediata ao usuário, como é o caso dos AVAs os quais demandam processamento próximo do tempo real.

A Figura 3 mostra a execução do sistema de reconhecimento de expressões faciais, detectando e rastreando as características faciais de um indivíduo. Verificam as

execuções dos algoritmos Viola-Jones e CANDIDE-3 (tela do lado direito) atuando sobre um quadro do vídeo, detectando a face e as características da face (os FCPs) em tempo real. A tela do lado esquerdo mostra a extração das características da face com os FCPs do CANDIDE-3 interligados, entre si, por meio de arestas nos locais de interesse (boca, nariz, sobranalha e olhos).



Figura 3. Execução do rastreamento das características faciais.

Foram necessários somente 38 FCPs do modelo de CANDIDE-3 para o desenvolvimento do sistema de reconhecimento de expressões faciais deste trabalho. Esses FCPs (ilustrados na Figura 4) foram utilizados para fazer mapeamentos as suas respectivas UAs. Na próxima Subseção será detalhada a classificação das expressões faciais, descrevendo o processo do mapeamento dos FCPs as suas respectivas UAs.

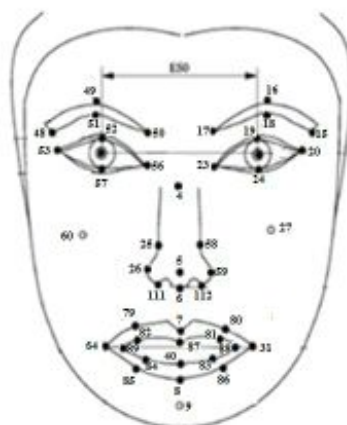


Figura 4. FCPs utilizados para o mapeamento das UAs.

4.1. Classificação das Expressões Faciais

Depois da identificação dos FCPs, a próxima etapa do sistema de reconhecimento de expressões faciais é feito pelo módulo de classificação das expressões faciais (Figura 2). Nesse módulo é realizado o mapeamento dos FCPs para suas respectivas UAs definidas no FACS [Ekman *et al.*, 2002]. Para encontrar as UAs foram necessários, primeiramente, fazer algumas transformações geométricas nas coordenadas dos FCPs para corrigir as distorções provocadas pelas mudanças em orientação, tamanho e formato das imagens 3D (terceira dimensão) capturadas da câmera do kinect. As principais transformações geométricas (escala, rotação e translação) foram aplicadas nas coordenadas dos FCPs, e depois submetidas a critérios que interpretam aos seus deslocamentos geométricos.

Com os pontos FCPs transformados a um novo sistema de coordenada, são feitos os cálculos dos deslocamentos de deformações geométricas nas expressões faciais, mostrados na Tabela 1. Esses deslocamentos são as ações faciais que submetidas as regras definidas das FACs [Ekman *et al.*, 2002], resultam nas suas UAs específicas.

Tabela 1. Deslocamentos das deformações geométricas nas expressões faciais.

Distância	Deformação	Equação	Distância	Deformação	Equação
D2	Mídia vertical da parte interna das sobrancelhas esquerda e direita até o canto interno dos olhos.	$(abs(pontos[17]y - pontos[23]y) + abs(pontos[50]y - pontos[56]y))/2$	D14	Abertura vertical dos olhos	$(abs(pontos[19]y - pontos[24]y) + abs(pontos[52]y - pontos[57]y))/2$
D3	Deslocamento horizontal dos cantos esquerdo e direito do lábio até o centro do nariz.	$(abs(pontos[31]x - pontos[33]x) + abs(pontos[64]x - pontos[65]x))/2$	D15	Deslocamento vertical entre o canto externo da narina e canto interno da sobrancelha.	$(abs(pontos[26]y - pontos[37]y) + abs(pontos[50]y - pontos[59]y))/2$
D4	Abertura vertical da boca.	$abs(pontos[7]y - pontos[3]y)$	D16	Deslocamento horizontal entre os cantos internos das sobrancelhas	$abs(pontos[17]x - pontos[56]x)$
D5	Deslocamento vertical entre os cantos dos lábios e os olhos.	$abs(pontos[31]y - (pontos[20]y + pontos[23]y)/2) + abs(pontos[64]y - (pontos[33]y + pontos[36]y)/2)/2$	D17	Deslocamento vertical entre o centro da pálpebra superior e o centro superior da sobrancelha.	$(abs(pontos[19]y - pontos[49]y) + abs(pontos[52]y - pontos[16]y))/2$
D6	Deslocamento vertical entre os cantos externos das sobrancelhas e os cantos externos dos olhos.	$(abs(pontos[15]y - pontos[20]y) + abs(pontos[48]y - pontos[53]y))/2$	D18	Deslocamento vertical entre o canto externo da sobrancelha e o canto interno do olho.	$(abs(pontos[15]y - pontos[23]y) + abs(pontos[48]y - pontos[26]y))/2$
D8	Deslocamento vertical das pálpebras inferiores em relação aos centros dos olhos.	$abs(pontos[24]y - (pontos[20]y + pontos[23]y)/2) + abs(pontos[57]y - (pontos[33]y + pontos[36]y)/2)/2$	D19	Deslocamento vertical entre o centro do olho e o centro inferior da sobrancelha.	$(abs(pontos[18]y - (pontos[20]y + pontos[23]y)/2) + abs(pontos[31]y - (pontos[36]y + pontos[33]y)/2))/2$
D9	Deslocamento vertical entre os cantos da boca e o centro do nariz.	$(abs(pontos[7]y - pontos[5]y) + abs(pontos[64]y - pontos[5]y))/2$	D20	Deslocamento vertical entre o centro do olho e o centro do lábio externo superior.	$(abs(pontos[7]y - (pontos[26]y + pontos[23]y)/2) + abs(pontos[7]y - (pontos[36]y + pontos[33]y)/2))/2$
D10	Deslocamento horizontal entre os cantos dos lábios.	$abs(pontos[31]x - pontos[64]x)$	D21	Deslocamento vertical entre o canto interno do olho e o canto externo da boca.	$(abs(pontos[23]y - pontos[31]y) + abs(pontos[16]y - pontos[64]y))/2$
D11	Deslocamento vertical entre o centro externo do lábio superior e o centro do nariz.	$abs(pontos[7]y - pontos[6]y)$	D23	Deslocamento vertical entre o centro do lábio externo inferior e o centro dos cantos internos das narinas.	$abs(pontos[8]y - pontos[6]y)$
D12	Deslocamento vertical entre o centro externo do lábio inferior e o centro do nariz.	$abs(pontos[8]y - pontos[6]y)$	D24	Deslocamento vertical entre o centro do olho e a bochecha.	$(abs(pontos[50]y - pontos[20]y) + abs(pontos[51]y - pontos[33]y))/2$
D13	Deslocamento vertical entre os pontos imediatamente acima da narina para o canto interno do olho.	$(abs(pontos[17]y - pontos[23]y) + abs(pontos[50]y - pontos[56]y))/2$			

Depois de calculados os deslocamentos de deformações geométricas, são executados os mapeamentos desses deslocamentos aos seus respectivos UAs, utilizando a Tabela 2 que apresenta os critérios que definem as UAs. O processo compara os deslocamentos de deformação (Tabela 1) de uma face realizando uma expressão facial com deslocamentos de deformação da mesma face no estado neutro para a identificar as UAs. O resultado desse processo indica se os deslocamentos de uma face, que realiza uma expressão, aumentam ou diminuem em relação à face neutra.

Com a descoberta das UAs, a próxima etapa (módulo de Inferência das Emoções) é mapear a combinação das UAs na sua respectiva emoção. Para realizar esse processo, foi utilizado os estudos de Ekman *et al* (2002) resumidos na Tabela 3. Nesta Tabela 3 estão agrupadas as UAs que mais comumente são encontradas nas expressões faciais das emoções básicas (alegria, tristeza, desgosto, raiva, surpresa e medo) [Yongmian; Qiang, 2005]. Tais UAs foram desenvolvidas por Ekman *et al* (2002) nas suas FACS. Para cada UA codificada, é baseada nos envolvimento dos músculos faciais [Ekman *et al.*, 2002].

Tabela 2. Critério de Reconhecimento dos UAs.

AU	Critério de Reconhecimento	AU	Critério de Reconhecimento
1	Aumento do deslocamento D2 em relação ao D2_n da imagem Neutra.	16	Diminuição dos deslocamentos D4, D12 e D23 em relação ao D4_n, D12_n e D23_n da imagem Neutra.
2	Aumento dos deslocamentos D6 e D18 em relação ao D6_n e D18_n da imagem Neutra.	17	Diminuição do deslocamento D11 em relação ao D11_n da imagem Neutra. Aumento do deslocamento D12 em relação ao D12_n da imagem Neutra.
4	Diminuição dos deslocamentos D7, D16 e D17 em relação ao D7_n, D16_n e D17_n da imagem Neutra.	20	Aumento do deslocamento D10 em relação ao D10_n da imagem Neutra. Diminuição do deslocamento D12 em relação ao D12_n da imagem Neutra.
5	Aumento do D14 e D1 em relação ao D14_n e D1_n da imagem Neutra.	23	Diminuição do deslocamento D10 em relação ao D10_n da imagem Neutra.
6	Diminuição dos deslocamentos D14 e D14 em relação ao D24_n e D14_n da imagem Neutra.	24	Aumento do deslocamento D11 em relação ao D11_n da imagem Neutra.
7	Diminuição do deslocamento D8 e aumento dos deslocamentos D14 e D19 em relação ao D8_n, D14_n e D9_n da imagem Neutra.	25	Aumento dos deslocamentos D4 e D12 em relação ao D4_n e D12_n da imagem Neutra. Diminuição do deslocamento D10 em relação ao D10_n da imagem Neutra.
9	Diminuição dos deslocamentos D2, D13, D14 e D15 em relação ao D2_n, D13_n, D14_n e D15_n da imagem Neutra.	26	Aumento do deslocamento D4 e D12 em relação ao D4_n, D12_n da imagem Neutra. Diminuição do deslocamento D11 em relação ao D11_n da imagem Neutra.
10	Diminuição dos deslocamentos D11, D12 e D20 em relação ao D11_n, D12_n, D20_n da imagem Neutra. Aumento do deslocamento D21 em relação ao D21_n da imagem Neutra.	27	Aumento do deslocamento D4 e D12 em relação ao D4_n, D12_n da imagem Neutra. Diminuição do deslocamento D11 em relação ao D11_n da imagem Neutra.
12	Aumento dos deslocamentos D3 e D10 em relação ao D3_n, D10_n da imagem Neutra. Diminuição do deslocamento D5 em relação ao D5_n da imagem Neutra.		
15	Aumento do deslocamento D9 em relação ao D9_n da imagem Neutra.		

De acordo com a Tabela 3, as UAs estão agrupadas em UAs principais e UAs auxiliares das expressões faciais. As UAs principais são aquelas combinações de UAs ou UA que podem ser classificadas sem ambiguidades, ou seja, são fortemente pertinentes para uma das seis expressões faciais.

Tabela 3. Associação das seis expressões faciais com seus UAs e combinações. Adaptado de Yongmian e Qiang (2005).

Emoção	Indicações Visuais Principais					Indicações Visuais Auxiliares				
Alegria	UA6	UA12				UA25	UA26	UA16		
Tristeza	UA1	UA15	UA17			UA4	UA7	UA25	UA26	
Desgosto	UA9	UA10				UA17	UA25	UA26		
Surpresa	UA5	UA26	UA27	UA1+2						
Raiva	UA2	UA4	UA7	UA23	UA24	UA17	UA16	UA25	UA26	
Medo		UA20	UA1+5	UA5+7		UA4	UA5	UA26	UA25	UA7

Em contraste, as UAs auxiliares são as que podem ser apenas aditivamente combinadas com as UAs principais a fim de proporcionar um apoio suplementar para a classificação de expressão facial. Consequentemente, uma expressão facial pode conter uma UA principal e uma UA auxiliar. Por exemplo, UA9 (nariz enrugado) pode ser diretamente associado com uma expressão de desgosto, enquanto se torna ambígua se associar somente uma única UA17 (levantador do queixo), para a expressão de desgosto. Quando UA9 e UA17 aparecem simultaneamente, a classificação desta combinação de UAs torna-se mais certa para classificar a expressão de desgosto. Assim, UA9 é uma UA principal de desgosto, enquanto UA17 é uma UA auxiliar de desgosto.

Combinando as UAs principais pertencentes à mesma emoção, aumenta o grau de confiança na classificação para essa emoção. Entretanto, combinando UAs principais ao longo de emoções diferentes pode resultar em: (1) emoção diferente, por exemplo, a combinação de UA1 (levantador da sobrancelha interna), uma UA principal para tristeza e UA5 (levantador da pálpebra superior), uma UA principal para surpresa, gera uma combinação UA principal de medo, e (2) aumento da ambiguidade, por exemplo, quando UA26 (boca aberta), uma UA principal para surpresa, combina com UA1, uma UA principal para a tristeza, o grau de crença em surpresa é reduzido e a ambiguidade de classificação é aumentada.

Em princípio, a nossa abordagem permite que uma expressão facial seja uma combinação de UAs faciais mais relevantes, utilizando para isto regras de produção

(cadeias de “ifs”), para realizar as classificações das UAs as suas respectivas emoções. Ou seja, o algoritmo classifica a uma respectiva emoção se esta tiver a maior quantidade de UAs associadas à respectiva expressão facial analisada. Atualmente, o sistema de reconhecimento de expressões faciais proposto neste trabalho só está inferindo as emoções alegria, tristeza, desgosto e raiva, pois o agente pedagógico do AVA baseado em PBL [Fontes *et al.*, 2012], atualmente só está projetado para entender somente essas emoções. Ficando as emoções básicas de surpresa e medo para trabalhos futuros.

5. Experimentos e Resultados

Esta seção apresenta os resultados experimentais que demonstram a precisão da abordagem proposta. Para a captura das imagens foi utilizado o hardware (câmera) Kinect. Na utilização dos métodos de detecção de faces (detecção pelo método Viola-Jones) e extração das características (detecção dos FCPs pelo CANDIDE) da etapa do rastreamento das características foram implementadas utilizando os recursos disponíveis da biblioteca do Kinect. Esta biblioteca contém algoritmos e métodos da área de Visão Computacional escrito nativamente em C++ [Microsoft Research, 2011].

O ambiente integrado de desenvolvimento do sistema implementado utilizado foi a IDE *Microsoft Visual Studio Express 2010* com a linguagem C++ padrão, para extensões da biblioteca SDK Kinect e C++, com suas bibliotecas de acesso a dados, interface com o usuário e comunicação entre processos. O MySQL foi o banco de dados escolhido para armazenar as faces neutras, o qual foi utilizado pelo AVA proposto por Fontes *et al* (2012). Todos os experimentos foram realizados em um *notebook* modelo *Samsung RV411*, composto por um processador Intel Core i5-480M, 2.66 GHz, 4 GB de memória RAM e 640 GB de HD.

Os experimentos foram conduzidos com a mudança e adição de parâmetros. O algoritmo de reconhecimento da emoção (Baseado em Regras) foi executado de forma automática e exaustiva, de modo que foi analisado somente o melhor resultado da taxa de acerto. Foram realizados experimentos durante o desenvolvimento e na fase de testes. Durante o desenvolvimento foi utilizados pequenas amostras de imagens de expressões faciais às quais foram essenciais para os primeiros ajustes (ajustes grossos) na aplicação. Depois de obter valores considerados satisfatórios durante o desenvolvimento, foram realizados testes mais abrangentes, utilizando uma amostra de dados maior para ajustes finais (ajustes finos).

Os testes foram feitos em tempo real, observando 10 alunos do curso de medicina da Universidade Estadual do Rio Grande do Norte (UERN). Cada indivíduo simularam 20 emoções (5 emoções de tristeza, 5 emoções de alegria, 5 emoções de raiva e 5 emoções de desgosto), totalizando 200 emoções realizadas por esses indivíduos. A aquisição de imagens foi realizada sobre variações de luzes e alguns indivíduos possuíam barbas, bigode ou usavam óculos e chapéus. Os indivíduos eram de ambos os sexos (6 homens e 4 mulheres) com idades na faixa de 21 a 45 anos. Nestes testes, o rastreamento das características faciais foi feita de modo automatizado, utilizando a técnicas de Viola-Jones e o modelo CANDIDE para detectar os pontos característicos faciais (FCPs).

Depois de rastrear esses FCPs, foram calculados os deslocamentos geométricos das emoções (Tabela 1). Em seguida foi utilizado o método baseado em regras para

encontrar as UAs de acordo com as variações dos deslocamentos geométricos (Tabela 2) para, em fim, inferir as emoções relacionadas com suas específicas combinações de UAs (Tabela 3). A Tabela 4, mostra que o resultado da taxa de reconhecimento total é de 85,5%. O desempenho geral da classificação automática emocional de expressões faciais executadas pelo sistema foi testado em um conjunto de 200 imagens de face de 10 indivíduos. Essas imagens são usadas para validar as regras da Tabela 2 da Subseção 4.1.

O desempenho de cada emoção resultante do algoritmo de Reconhecimento de emoção é mostrado na Tabela 4, e pode ser visto que alegria e raiva executam bem em comparação com as outras emoções, ambas atingindo uma taxa de acerto de 90%. Este resultado é esperado, pois, essas são emoções muito distintas, causando uma série de deformação dentro do rosto. As UAs associadas com essas emoções são detectadas na máscara do CANDIDE-3, portanto, o movimento dessas áreas é facilmente detectado pelo nosso sistema. Por outro lado, as outras emoções (desgosto e tristeza) não são bem detectadas na máscara em relação a alegria e raiva. A emoção desgosto tem uma boa precisão (80% de acerto), no entanto, é a mais baixa em relação às outras emoções. Uma explicação para isso pode ser o fato de que essa emoção é bastante sutil, e é facilmente confundida com a emoção raiva e com a expressão neutra. Essa confusão também aconteceu com a emoção tristeza, a qual foi confundida pela emoção raiva e neutra. Essa confusão pode ser vista na Tabela 5 que mostra a matriz de confusão resultante do algoritmo de reconhecimento de expressões faciais baseado em regra proposto neste trabalho.

Tabela 4. Resultado da precisão das emoções

Emoção	Taxa de acerto
Alegria	90%
Tristeza	82%
Desgosto	80%
Raiva	90%

Como pode ser observada na Tabela 5, a expressão mais ambígua facial foi desgosto, apresentando a menor taxa de acerto, uma vez que foi classificada incorretamente na maioria das vezes com a expressão raiva e, em seguida, neutro. As expressões faciais que se seguem, são a raiva e alegria, com uma taxa de acerto de classificação similar (90%). Os resultados das emoções, no geral, apresentam-se satisfatório para sua aplicabilidade ao AVA baseado em agentes, pois todas as emoções possuíram taxa de acerto acima de 80%. Um valor comparável a vários sistemas de reconhecimento de emoções existentes [Sebe *et al.*, 2004]. Além disso, segundo Sebe *et al* (2004), a média do reconhecimento de emoções faciais feitos por humanos é de aproximadamente 85% de precisão, e que alguns algoritmos computacionais obtém sucesso entre 74% e 98%. No entanto, no caso de o sistema reconhecer as emoções erroneamente, o AVA receberá a emoção falsa e se responsabilizará se trata da mesma forma que uma emoção verdadeira, de acordo com regras pré-estabelecidas pelo AVA [Fontes *et al.*, 2012].

Comparando os resultados do sistema proposto com os resultados apresentados por Changjun *et al* (2011), que propuseram um sistema de reconhecimento de emoções baseado no algoritmo SVM para um AVA, este último foi testado em 100 imagens

faciais em 2D (duas dimensões) de 15 alunos, atingindo uma taxa de reconhecimento total é de 84,55%. Enquanto o sistema proposto, baseado em regras, foi testado em 200 imagens faciais em 3D (três dimensões) de 10 indivíduos, atingindo um reconhecimento total de 85,5%.

Tabela 5. Matriz de confusão resultante dos testes do Sistema proposto.

Emoções	Tristeza	Alegria	Raiva	Desgosto	Neutro
Tristeza	0.82	0.00	0.06	0.00	0.12
Alegria	0.00	0.90	0.0	0.0	0.10
Raiva	0.00	0.00	0.90	0.05	0.05
Desgosto	0.00	0.00	0.15	0.80	0.05
Neutro	0.15	0.00	0.25	0.00	0.6

Por fim, está sendo realizado um estudo de caso do sistema de reconhecimento de expressões faciais integrado ao AVA como forma de validar a eficácia dos resultados apresentados neste trabalho. Esse estudo de caso está integrado ao projeto intitulado “Um Agente Pedagógico 3D de Apoio a Estudantes de Medicina na Resolução de Problemas na Área da Oncologia Utilizando a Aprendizagem Baseada em Problema” aprovado pelo COHM (Centro de Oncologia e Hematologia). O estudo de caso está sendo realizado com estudantes do curso de medicina, sendo essa validação apoiada pelo COHM. Nessa ocasião, os estudantes estão utilizando esse sistema de reconhecimento de emoções integrado com o AVA baseado em agentes de apoio a PBL, a fim de resolverem problemas relacionados à área de Oncologia. De acordo com os resultados preliminares, obtidos de simulações com 10 alunos de medicina, o sistema reconhecimento de emoções atingiu uma taxa de reconhecimento de aproximadamente 85% e um tempo de resposta ficando em média de 3 a 5 segundos, utilizando a Internet com velocidade variando de 2 Mbps (*Megabits* por segundo) a 512 kbps (*Kilobits* por segundo). Estes resultados são considerados equivalentes aos testes descritos neste trabalho. Além disso, foi o tempo de respostas foi considerado eficaz, pois foi classificado como baixo e aceitável pelos estudantes do curso de medicina na resolução de problemas da área de oncologia.

6. Considerações Finais e Trabalhos Futuros

Neste trabalho, são tratados problemas de inferência de emoções de indivíduos que interagem em um AVA, utilizando o reconhecimento de expressões faciais. A abordagem, utilizando o rastreamento de características faciais por meios das técnicas Viola-Jones e CANDIDE-3, atingiu uma taxa total de acurácia de reconhecimento de emoção de 85,5 %. Um resultado alto e satisfatório para aplicação no AVA baseado em agentes. Foi proposta uma nova abordagem de reconhecimento de expressão facial baseado em regras a qual se mostrou eficaz de acordo com os resultados dos experimentos. E, por fim, serão buscadas alternativas que sejam independentes da tecnologia Kinect, permitindo que apenas uma *câmera* convencional seja utilizada, diminuindo os custos para os alunos de um AVA. No futuro, com base nas características faciais encontradas, outros métodos de classificação de emoção poderão ser usados. Também serão estudados outros módulos de reconhecimento de emoções, tais como: recursos de áudio e gestos.

Referências

- Ahlberg, J. CANDIDE-3 -- an updated parameterized face. Report No. LiTH-ISY-R-2326, Dept. of Electrical Engineering, Linköping University, Sweden, 2001.
- Aquino Jr., Plinio T. Personas Alunos e Papéis do Docente no Atendimento da Diversidade de Perfis. In: SBSI 2012 - VIII Simpósio Brasileiro de Sistemas de Informação, 2012, São Paulo. Anais do VIII Simpósio Brasileiro de Sistemas de Informação (SBSI 2012). Porto Alegre: Sociedade Brasileira de Computação - SBC, 2012. v. 01. p. 767-778.
- Changjun, Z.; Shen, P.; Chen, X. Research on algorithm of state recognition of students based on facial expression. Electronic and Mechanical Engineering and Information Technology (EMEIT), International Conference on, vol.2, no., pp.626-630, 12-14 Aug. 2011.
- Dornaika, F.; Ahlberg, J. Fast and reliable active appearance model search for 3-D face tracking. IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 34, no. 4, pp. 1838-1853, Aug. 2004.
- Ekman, P.; Friesen, W.V.; Hager, J.C. Facial Action Coding System: Investigator's guide. Research Nexus division of Network Information Research Corporation, Salt Lake City, Estados Unidos, 2002.
- Fontes, L. M. O.; Mendes Neto, F. M.; Diniz, F. A.; Jácome Junior, L.; Silva, L. C. N.; Carlos, D. G. Um Agente Pedagógico Animado de Apoio à Aprendizagem Baseada em Problema Utilizando o Moodle. In: Workshop sobre Avaliação e Acompanhamento da Aprendizagem em Ambientes Virtuais, XXIII Simpósio Brasileiro de Informática na Educação. Rio de Janeiro, 2012.
- Jain, A. K.; Ross, A.; Prabhakar, S. An Introduction to Biometric Recognition, IEEE Transactions on Circuits and Systems for Video Technology Special Issue on Image and Video-Based Biometrics. v. 14, n. 1, p. 4-20, 2004.
- Microsoft Research. Programming Guide: Getting Started with the Kinect for Windows SDK Beta, 2011. Disponível em:<http://research.microsoft.com/en-us/um/redmond/projects/kinectsdk/docs/ProgrammingGuide_KinectSDK.docx>. Acesso em: 13 Setembro 2011.
- Sebe, N.; Lew, M. S.; Huang, T. S. The State-of-the-Art in Human-Computer Interaction. In: European Conference on Computer Vision 2004 (ECCV 2004): International Workshop on Human Computer Interaction (HCI'04), p 1-6, Praga, Republica Tcheca, 2004.
- Sun Duo; Lu Xue Song. Research on E-learning system based on affective computing. Information Management and Engineering (ICIME), 2010 The 2nd IEEE International Conference on, vol., no., pp.697-699, 16-18, 2010.
- Viola, P.; Jones, M. Robust real-time object detection. Technical report, University of Cambridge, 2001.
- Yongmian, Z.; Qiang, Ji. Active and dynamic information fusion for facial expression understanding from image sequences. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.27, no.5, pp.699-714, 2005.