



**UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO
UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO**



FÁBIO ABRANTES DINIZ

**RedFace – Um Sistema de Reconhecimento de Expressões Faciais
para Apoiar um Ambiente Virtual de Aprendizagem.**

MOSSORÓ – RN

2013

FÁBIO ABRANTES DINIZ

**RedFace – Um Sistema de Reconhecimento de Expressões Faciais
para Apoiar um Ambiente Virtual de Aprendizagem.**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação – associação ampla entre a Universidade do Estado do Rio Grande do Norte e a Universidade Federal Rural do Semi-Árido, para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Francisco Milton Mendes Neto – UFERSA.

Co-orientador:
Prof. Dr. Francisco Chagas de Lima Junior - UERN

MOSSORÓ – RN

2013

Ficha catalográfica preparada pelo setor de classificação e catalogação da Biblioteca “Orlando Teixeira” da UFERSA

D585r Diniz, Fábio Abrantes.

RedFace – um sistema de reconhecimento de expressões faciais para apoiar um ambiente virtual de aprendizagem. / Fábio Abrantes Diniz. -- Mossoró, 2013.

133 f.: il.

Dissertação (Mestrado em Ciências da Computação) – Universidade Federal Rural do Semi-Árido.

Orientador: D.Sc. Francisco Milton Mendes Neto.

Co-orientador: D.Sc. Francisco Chagas De Lima Júnior.

1. Expressões Faciais. 2. PCA. 3. Ambiente Virtual de Aprendizagem e Afetividade. I. Título.

CDD: 004.678


FÁBIO ABRANTES DINIZ

**RedFace – Um Sistema de Reconhecimento de Expressões Faciais
para Apoiar um Ambiente Virtual de Aprendizagem.**

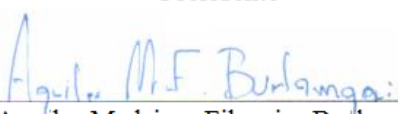
Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação para a obtenção do título de Mestre em Ciência da Computação.

APROVADA EM: 21 / 02 / 2013 .

BANCA EXAMINADORA



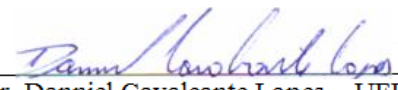
Prof. Dr. Francisco Milton Mendes Neto – UFERSA
Presidente



Prof. Dr. Aquiles Medeiros Filgueira Bu/lamaqui – UFRN
Membro Externo



Prof. Dr. Francisco Cuiagas de Lima Júnior – UERN
Membro Interno



Prof. Dr. Danniell Cavalcante Lopes – UFERSA
Membro Interno

Dedico este trabalho aos meus amados pais, José e Juranilda, a meus fiéis irmãos, Jean, José Emerson, Janílson e Jânio, e ao meu orientador Francisco Milton Mendes Neto. Todos me fortaleceram para que fosse possível concluir esta jornada em minha vida.

AGRADECIMENTOS

Agradeço a Deus, que me deu oportunidades em minha vida para que fosse possível chegar ao ponto que estou hoje. Agradeço aos meus pais, José Abrantes e Juranilda Diniz, que com muito custo e esforço me tornaram a pessoa que sou hoje. Devo tudo a eles.

A meus irmãos, Janílson, Jânio, José Emerson e Jean Carlos, que foram companheiros para todas as horas e nunca deixaram faltar nada na minha vida, desde bens materiais, alimentação a seus amores fraternais em todos meus momentos difíceis.

À minha namorada, Maria de Perpétuo de Alencar, que esteve sempre me dando forças, acompanhando-me neste momento e, ao mesmo tempo, é final e inicial em minha vida. Ela me deu amizade e amor em momentos que eu mais precisava.

Ao meu primo Darcio Abrantes, por me ajudar a encontrar uma boa moradia e pela virtude do companheirismo que foi fundamental nesta etapa da minha vida.

A todos os meus amigos, em especial Thiago Reis e Ithalo Moura que me divertem e sabem criticar-me quando necessário, e as pessoas importante para o trabalho como Thiago Domingues e Lucas Figueiredo que me ajudaram na implementação do protótipo desse trabalho.

A Francisco Chagas de Lima Júnior, co-orientador deste trabalho, pelo seu apoio no momento que eu mais precisava, pois eu estava ausente de co-orientador. Agradeço muito pela sua atenção e por ter me indicado materiais que foram úteis para eu explorar, e além de ter me ajudado nas correções dos meus artigos e da dissertação.

Ao meu orientador Francisco Milton Mendes Neto, pela sua paciência e confiança na minha capacidade e habilidades de desenvolver um trabalho. Agradeço muito pela sua atenção, leitura cuidadosa e conselhos que fizeram este trabalho melhor, e que, pela convivência, se tornou um grande amigo.

A todos os professores do programa de pós-graduação de ciência da computação da UERN/UFERSA que contribuíram direto ou indiretamente para realização deste trabalho de mestrado. Espero que todos continuem assim, pois acredito que esse programa vai ser um dos melhores do Estado do Rio Grande do Norte.

A CAPES, pelo apoio financeiro concebido no qual viabilizou a realização deste trabalho.

Obrigado a todos,

Fabio Abrantes Diniz

RESUMO

A maioria dos Ambientes Virtuais de Aprendizagens (AVAs) não dispõe de mecanismos eficientes que possam identificar os alunos tanto no acesso ao sistema quanto durante o desenvolvimento de suas atividades. Em geral, este tipo de ambiente adota como prática de segurança a utilização de mecanismo de autenticação por login e senha. O uso deste tipo simples de autenticação aumenta a vulnerabilidade a fraudes no aspecto relativo à identidade daquele que realiza o curso. Além disso, nos AVAs, a avaliação do processo de ensino e aprendizagem do aluno tem que levar em consideração os fenômenos afetivos durante toda a interação do aluno com o ambiente, visto que esses fenômenos interferem profundamente nos processos mentais, como memorização, raciocínio, atenção e, principalmente, motivação. Desta forma, o presente trabalho propõe melhorias no processo de aprendizagem e nas atividades avaliativas de um AVA composto por agentes inteligentes. O sistema proposto utiliza técnicas de Análise dos Componentes Principais (*Principal Component Analysis - PCA*), para reconhecimento de identificação facial e um Sistema de Codificação de Ação Facial (FACS), para rastreamento dos pontos característicos faciais, e um método de classificação da emoção baseado em regras para reconhecimento de expressões faciais. Este sistema, por meio de uma câmera, permite identificar o aluno durante toda a realização das avaliações pedagógicas do AVA, além reconhecer a emoção (alegria, tristeza, raiva e desgosto) do aluno pela sua expressão facial. A emoção servirá de entrada para um agente pedagógico animado, existente no AVA, que tentará motivá-lo à aprendizagem de forma mais eficaz.

Palavras-Chave: Expressões Faciais, PCA, Ambiente Virtual de Aprendizagem e Afetividade.

ABSTRACT

The most of Virtual Learning Environments (VLEs) has no efficient mechanisms for identifying students both in accessing a system and during the activities development. In general, this system uses security mechanisms for authentication by login and password. Using this simple type of authentication increases vulnerability to fraud in an aspect concerning the identity of who performs the course. In addition, in VLEs, the assessment of teaching-learning process of students must take into account the affective phenomena throughout the student's interaction with the environment, because they deeply affect the mental processes such as memory, reasoning, attention and mainly the motivation. Thus, this work proposes improvements in learning process and in the evaluative activities of a VLE composed by intelligent agents. The proposed system uses techniques of Principal Component Analysis (PCA) to face identification recognition and a Facial Action Coding System (FACS), for tracking facial features points, and an emotion classification method based on rules to facial expressions recognition. This system by means of a camera enables a student to be identified during a performance of VLE pedagogical assessments and recognizes student emotions (happiness, sadness, anger and disappointment) by facial expression. The emotion will serve as input for an animated pedagogical agent, which composes the VLE, in order to motivate the student to learn more effectively.

Keywords: Facial Expressions, PCA, Virtual Learning Environment and Affectivity.

LISTA DE TABELAS

Tabela 01 – Análise entre os principais métodos biométricos.....	30
Tabela 02 – Um comparativo das técnicas de detecção de característica.....	42
Tabela 03 – Algoritmos de reconhecimento de padrão no OpenCV.....	46
Tabela 04 – Exemplo de combinações de UAs formando algumas expressões de emoções....	67
Tabela 05 – Resumo de trabalhos de reconhecimento de emoções faciais.....	83
Tabela 06 – Testes <i>Random Forest</i>	98
Tabela 07 – Testes <i>K-Star</i>	98
Tabela 08 – Deslocamentos das deformações geométricas nas expressões faciais.....	107
Tabela 09 – Critérios de Reconhecimento dos UAs.....	108
Tabela 10 – Associação das seis expressões faciais com seus UAs e combinações.....	110
Tabela 11 – Resultado da precisão das emoções.....	114
Tabela 12 – Matriz de confusão.....	115
Tabela 13 – Exemplos de UAs do sistema FACS.....	125
Tabela 14 – Vértices do CANDIDE-3 com suas descrições.....	130

LISTA DE FIGURAS

Figura 01 – Exemplo de imagem digital monocromática.....	26
Figura 02 – Representação numérica de uma imagem digital monocromática.....	27
Figura 03 – Variação de resolução espacial de uma imagem com 256 níveis de cinzas.....	28
Figura 04 – Variação de profundidade de uma mesma imagem.....	28
Figura 05 – Alguns tipos de métodos de identificação existentes.....	31
Figura 06 – Arquitetura padrão de um sistema biométrico.....	33
Figura 07 – Características <i>Haar</i>	38
Figura 08 – Exemplo de cálculo de uma região retangular.....	39
Figura 09 – Cascata de classificadores com n fases.....	40
Figura 10 – Modelo CANDIDE.....	43
Figura 11 – Exemplo do CANDIDE-3 capturando alguns pontos característicos da face.....	43
Figura 12 – O modelo CANDIDE-3 refletindo várias formas e expressões faciais.....	44
Figura 13 – (A) Visão modelo comercial do Kinect. (B) Visão figurada do modelo Kinect...	48
Figura 14 – Resumo da arquitetura básica.....	49
Figura 15 – Articulações do esqueleto do homem	50
Figura 16 – Exemplo de face média (acima) e <i>Eigenfaces</i> mais importantes no treinamento..	56
Figura 17 – Arquitetura padrão do MOODLE.....	58
Figura 18 – Arquitetura do AVA baseado em agentes de apoio à PBL.....	60
Figura 19 – Animações referentes aos estados emocionais do AgPA.....	61
Figura 20 – Afetividade e manifestações do estado afetivo.....	64
Figura 21 – Seis expressões faciais emocionais básicas.....	65
Figura 22 – UA 43 do Sistema FACS	67
Figura 23 – Ramos da computação afetiva.....	69
Figura 24 – Mapeamentos dos aspectos afetivos.....	71
Figura 25 – Agente PAT.....	73
Figura 26 – Arquitetura do agente pedagógico Dóris.....	74
Figura 27 – Elementos envolvidos nas expressões faciais.....	76
Figura 28 – Metodologia básica para o reconhecimento de padrões.....	77
Figura 29 – Arquitetura do Sistema proposto por Rolim e Bezerra (2008).....	81
Figura 30 – Passos de reconhecimento de expressões faciais.....	85

Figura 31 - Interação entre os subsistemas do módulo de reconhecimento facial localizados no cliente e no servidor do Moodle.....	88
Figura 32 – Arquitetura do Módulo de Reconhecimento Facial.....	89
Figura 33 – Execução do algoritmo de detecção de face Viola-Jones em um vídeo.....	90
Figura 34 – Exemplo de funcionamento do K-NN.....	93
Figura 35 – Imagens de faces da base de dados.....	96
Figura 36 – Esquema de posicionamento da face.....	96
Figura 37 – Resultados da classificação para as imagens de dimensões 120x90, 32x24 e 12x9.....	97
Figura 38 – Interação do módulo de reconhecimento de expressões faciais com o AVA.....	101
Figura 39 – Arquitetura do Módulo de reconhecimento de expressão facial.....	103
Figura 40 – Execução do rastreamento das características faciais.....	105
Figura 41 – FCPs utilizados para o mapeamento das UAs.....	105
Figura 42 – (a) Sem translação. (b) Com translação. Fonte própria.....	106
Figura 43 – Pseudocódigo do algoritmo de inferência de emoções.....	111
Figura 44 – Expressão Facial para cada emoção específica: 1-Neutra, 2- Raiva, 3-Tristeza, 4-Alegria, 5-Desgosto.....	113
Figura 45 – Algoritmo <i>AdaBoost</i>	128
Figura 46 – Algoritmo de treinamento.....	129
Figura 47 – Código de transformação geométrica de translação e rotação de imagens 3D....	133
Figura 48 – Algoritmo de reconhecimento de emoções baseado em regras.....	134

LISTA DE ABREVIATURAS E SIGLAS

AgDP – Agente Detector de Problemas;
AgFG – Agente Formador de Grupos;
AgPA – Agente Pedagógico Animado;
AgR – Agente Recomendador;
API – *Application Programming Interface*;
AVA – Ambiente Virtual de Aprendizagem;
BDI – *Belief Desire Intention*;
Bits – *Binary digit*;
DCT – *Discrete Cosine Transform*;
DF – Agente *Directory Facilitator*;
EaD – Educação a Distância ;
FA – Falsa Aceitação;
FACS – Sistema de Codificação de Ação Facial;
FaD – *Face Detection*;
FCP – *Facial Characteristic Point*;
FeD – *Feature Detection*;
fps – *frames* por segundo;
FR – Falsa Rejeição;
GB – *Green Blue*;
IA – Inteligência Artificial;
IDE – *Integrated Development Environment*;
IHC – Interação Homem-Máquina;
JADE – *Java Agent Development Framework*;
LDA – *Linear Discriminant Analysis*;
MLL – *Machine Learning Library*;
MOODLE – *Modular Object-Oriented Dynamic Learning Environment*;
NUI – *Natural User Interface*;
OA – Objeto de Aprendizagem;
OOC – *Ortony, Clore e Collins*;
PBL – *Problem-Based Learning*;
PCA – *Principal Component Analysis*;

PHP – *Hypertext Preprocessor*;
PI – Processamento de Imagens;
RG – *Red Green*;
RGB – *Red, Green e Blue*;
RI – Raios Infravermelhos;
SGBD – Sistema de Gerenciamento de Banco de Dados;
SU – *Shape Unit*;
SVM – *Support Vector Machine*.
UA – Unidade de Ação;
UFERSA – Universidade Federal Rural do Semi-Árido;
USB – *Universal Serial Bus*;
VC – Visão Computacional;
XML – *Extensible Markup Language*;

SUMÁRIO

1 INTRODUÇÃO	17
1.1 OBJETIVO GERAL	19
1.1.1 OBJETIVOS ESPECÍFICOS	20
1.2 PUBLICAÇÕES RELACIONADAS	20
1.3 ORGANIZAÇÃO DA DISSERTAÇÃO	23
2 FUNDAMENTAÇÃO TEÓRICA.....	25
2.1 IMAGENS DIGITAIS	26
2.2 BIOMETRIA	29
2.2.1 IDENTIFICAÇÃO DE INDIVÍDUOS.....	31
2.2.2 SISTEMAS BIOMÉTRICOS	32
2.3 VISÃO COMPUTACIONAL.....	34
2.3.1 DETECÇÃO FACIAL.....	35
2.3.1.1 MÉTODO DE VIOLA E JONES	38
2.3.2 DETECÇÃO DE CARACTERÍSTICAS	40
2.3.2.1 MODELO CANDIDE.....	42
2.3.3 BIBLIOTECA OPENCV.....	45
2.3.4 BIBLIOTECA DO KINECT	47
2.4 RECONHECIMENTO DE FACES.....	51
2.4.1 ANÁLISE DOS COMPONENTES PRINCIPAIS E <i>EIGENFACES</i>	53
2.5 AMBIENTES VIRTUAIS DE APRENDIZAGEM	56
2.5.1 MODULAR <i>OBJECT-ORIENTED DYNAMIC LEARNING ENVIRONMENT</i> (MOODLE).....	57
2.5.2 AMBIENTE VIRTUAL DE APRENDIZAGEM BASEADO EM AGENTES DE APOIO À PBL	59
2.6 EMOÇÕES	63
2.7 MODELO PSICOLÓGICO DE CLASSIFICAÇÃO FACIAL – FACS.....	66
2.8 A COMPUTAÇÃO AFETIVA	68
2.8.1 MAPEAMENTO DE ASPECTOS AFETIVOS.....	70
2.8.2 USO DE COMPUTAÇÃO AFETIVA EM AMBIENTES VIRTUAIS DE APRENDIZAGEM.....	72
2.8.2.1 AGENTE PEDAGÓGICO PAT.....	72

2.8.2.2 AGENTE PEDAGÓGICO DÓRIS	74
2.9 RECONHECIMENTO DE EXPRESSÕES FACIAIS	75
3 TRABALHOS RELACIONADOS	79
3.1 RECONHECIMENTO DE FACES	79
3.2 RECONHECIMENTO DE EXPRESSÕES FACIAIS	82
4 REDEFACE: SISTEMA DE RECONHECIMENTO DE EXPRESSÕES FACIAIS DE APOIO A UM AVA	86
4.1 IMPLEMENTAÇÃO DO MÓDULO DE RECONHECIMENTO FACIAL.....	87
4.1.1 ARQUITETURA DO MÓDULO.....	87
4.1.1.1 CLASSIFICADOR <i>K-NEAREST NEIGHBORS</i>	92
4.1.1.2 CLASSIFICADOR <i>RANDOM FOREST</i>	93
4.1.1.3 CLASSIFICADOR <i>K-STAR</i>	94
4.1.2 EXPERIMENTOS E RESULTADOS.....	94
4.1.3 DISCUSSÕES DOS EXPERIMENTOS	99
4.2 IMPLEMENTAÇÃO DO MÓDULO DE RECONHECIMENTO DE EXPRESSÕES FACIAIS	100
4.2.1 ARQUITETURA DO MÓDULO.....	101
4.2.1.1 RASTREAMENTO DAS CARACTERÍSTICAS FACIAIS	103
4.2.1.2 CLASSIFICAÇÃO DAS EXPRESSÕES FACIAIS.....	106
4.2.1.3 INFERÊNCIA DAS EMOÇÕES.....	110
4.2.2 EXPERIMENTOS E RESULTADOS.....	112
5 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS.....	115
REFERÊNCIAS	117
APÊNDICE A – UAS DO SISTEMA FACS.....	125
APÊNDICE B – CÓDIGO <i>ADABOOST</i> E ALGORITMO DE TREINAMENTO	128
APÊNDICE C – VÉRTICES DO CANDIDE-3 COM SUAS DESCRIÇÕES.....	130
APÊNDICE D – CÓDIGO DE TRANSFORMAÇÃO GEOMÉTRICA DE TRANSLAÇÃO E ROTAÇÃO	133
APÊNDICE E – CÓDIGO DE RECONHECIMENTO DE EMOÇÕES BASEADO EM REGRAS	134

1 INTRODUÇÃO

O papel da educação na sociedade de hoje é indiscutível. A educação, seja ela presencial ou a distância, apresenta-se como elemento indispensável não apenas à manutenção da vida, mas à elevação de sua qualidade. A Educação a Distância (EaD), apresenta-se, neste contexto, como uma modalidade que oferece educação ao cidadão, favorecendo sua inclusão social e profissional (FIORESE e TAROUCO, 2006).

Nos últimos anos têm-se notado um crescimento acelerado desta modalidade nos Ambientes Virtuais de Aprendizagens (AVAs), que apoiam para a EaD por meio da Internet (PENTEADO e MARANA, 2009). A proliferação dos AVAs é consequência das vantagens da educação baseada na Web (CANTONI, CELLARIO e PORTA, 2003), tais como: custos de distribuição reduzidos, aprendizado autodirigido, aprendizado geograficamente independente, atualização simples de materiais e gerenciamento simplificado de grandes grupos de estudantes.

Apesar dessas vantagens, o aumento da demanda pelos AVAs e os avanços das tecnologias de informação e comunicação têm gerado a necessidade de se melhorar estratégias tanto no acompanhamento e monitoramento dos alunos a distância como na identificação dos estados de ânimo do aluno durante os processos avaliativos e de aprendizado.

A maioria dos AVAs não dispõe de recursos eficientes e automatizados que garantam a participação dos usuários no desenvolvimento de suas atividades acadêmicas, incluindo seu processo de avaliação via Internet (RABUZIN *et al.*, 2006). Em geral, alguns destes ambientes adotam a utilização de mecanismos de identificação por *login* e senha como prática de segurança. O uso deste tipo simples de autenticação aumenta a vulnerabilidade a fraudes, tanto no acesso ao sistema quanto durante a participação do aluno nas atividades e nas avaliações do curso, pois outra pessoa pode substituir facilmente o indivíduo que deveria ser avaliado no AVA (RABUZIN, 2006). Logo, é importante que métodos de segurança mais eficazes sejam explorados, a fim de monitorar os alunos a distância durante todo o processo de aprendizagem no AVA.

Além disso, poucos AVAs têm mecanismos para identificar os estados afetivos (emoções) do aluno durante toda sua interação com o AVA. Isto se torna importante visto que os fenômenos afetivos interferem profundamente nos processos mentais do aluno, como memorização, raciocínio, atenção e, principalmente, a motivação (PICARD, 2004).

A busca por uma solução onde o AVA entenda e reaja às emoções dos alunos de forma personalizada, poucos resultados significativos têm sido alcançados de forma que contribua no processo de aprendizagem. No entanto, a verificação dos estados emocionais do aluno é desafiante e foco das pesquisas da área de Computação Afetiva (PICARD, 1997) citado por (SUN e LU, 2010). Este campo compreende pesquisas em diferentes linhas de estudos, tais como: PI, VC e Psicologia. O

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 (PICARD, 1997) citado por (SUN e LU, 2010). Ou seja, a máquina poderá ter uma capacidade de percepção dos estados emocionais do seu usuário e adaptar seu comportamento funcional (PICARD, 1997) citado por (ZHOU *et al.*, 2010).

São várias as formas de identificar uma pessoa e verificar suas emoções, e mais variados ainda são os métodos para captar e reconhecer tanto uma pessoa como a sua emoção transmitida. O principal meio é utilizando a tecnologia biométrica (DEWI *et al.*, 2011). O avanço da tecnologia biométrica, aliado à redução dos preços dos equipamentos e à evolução de sua potência, fez com que surgissem várias pesquisas da biometria em ambiente *Web* (DEWI *et al.*, 2011).

Dentre as tecnologias biométricas mais pesquisadas, os reconhecimentos de faces e de expressões faciais destacam-se por sua aceitabilidade por parte das pessoas e facilidade de coleta, mesmo de forma não colaborativa (JAIN, ROSS e PRABHAKAR, 2004). Com as câmeras mais baratas e potentes, aliadas a Internet cada vez mais veloz e a baixo custo, o reconhecimento das faces e expressões faciais nos AVAs tende a ser um dos vértices tecnológicos que possibilita o trabalho colaborativo (PENTEADO e MARANA, 2009).

Muitos projetos atualmente exploram a tecnologia da câmera em conjunto com a Internet na educação (PENTEADO e MARANA, 2009). As áreas de segurança, acessibilidade (portadores de necessidades especiais) e hipermídia adaptativa são exemplos de áreas que se beneficiam com pesquisas ligadas ao uso de uma câmera (WHITEHILL *et al.*, 2009). Logo, o uso de uma câmera e da Internet na perspectiva de apoio ao professor e aos processos pedagógicos é também primordial tanto na identificação do aluno durante os processos avaliativos como na verificação da motivação e outros aspectos afetivos no decorrer do aprendizado do aluno no AVA (WHITEHILL *et al.*, 2009).

Neste intuito, visando melhorar as relações entre homem e computador, este trabalho tem como meta identificar os alunos e reconhecer suas emoções básicas (raiva, alegria,

desgosto e tristeza), utilizando imagens provenientes de uma câmera. A emoção reconhecida servirá como 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). Esse AVA é composto por um agente pedagógico que usa expressões emocionais a fim de oferecer maior motivação e entusiasmo no processo de aprendizagem, gerando autonomia e proporcionando atividades desafiadoras aos aprendizes.

Para chegar a este objetivo, foi necessário realizar alguns passos intermediários utilizando métodos de VC (CHIU e RASKAR, 2009), como os métodos de detecção de faces (*Face Detection* - FaD) e detecção de características faciais (*Feature Detection* - FeD) (YANG *et al.*, 2002) citado por (BAGHERIAN e RAHMAT, 2008; PATIL *et al.*, 2010). Com base nestas detecções, posteriormente, foi realizado o reconhecimento das faces e das classificações de emoções presentes em imagens, utilizando algoritmos de aprendizado de máquina da Inteligência Artificial (IA) e os esquemas do Sistema de Codificação de Ação Facial (FACS) (EKMAN *et al.*, 2002) citado por (ABDAT *et al.*, 2011).

1.1 OBJETIVO GERAL

Tendo em vista a problemática apresentada, o presente trabalho tem como objetivo geral aplicar métodos existentes em Visão Computacional, Aprendizagem de Máquina, Processamento de Imagens e na Psicologia, a fim de implementar um sistema de reconhecimento de faces e de emoções faciais, denominado REDFACE. O sistema proposto é capaz de detectar uma face, reconhecê-la e classificar algumas de suas emoções (alegria, desgosto, raiva e tristeza) das expressões faciais de indivíduos (alunos), as quais servirão de entrada para um AVA, composto por um agente inteligente pedagógico animado com expressões emocionais desenvolvido por Fontes *et al* (2012) .

1.1.1 OBJETIVOS ESPECÍFICOS

Para atingir o objetivo principal desse trabalho, que é desenvolver o sistema REDFACE para um AVA composto por agentes inteligentes (FONTES *et al.*, 2012), também foram alcançados os seguintes objetivos específico:

- Entender o uso do conhecimento de Computação Afetiva, IHC e VC;
- Agregar conhecimento sobre técnicas de detecção e identificação de faces e de expressões faciais para reconhecer a área do rosto de um indivíduo em uma imagem e, depois, identificar tanto a pessoa com suas emoções;
- Estudar técnicas e métodos de detecção de características existentes nas faces;
- Estudar bibliotecas de desenvolvimento de software existente na área de VC e PI, tais como: OpenCV (INTEL, 2007) e Kinect SDK (MICROSOFT RESEARCH, 2011);
- Implementar os sistemas de reconhecimento de faces e de expressões faciais;
- Adquirir bases de dados de imagens de faces e de expressões faciais para fazer os treinamentos e testes da aplicação do módulo REDFACE, a fim de avaliar o sistema;
- Identificar metodologias de avaliação dos sistemas de reconhecimento de faces e de expressões faciais com o propósito de avaliar os processos de reconhecimento das faces e das expressões faciais;
- Analisar os resultados obtidos durante a implementação das amostras na etapa do teste;
- Sugerir propostas para evolução e continuidade deste trabalho, ponto de partida de um sistema mais elaborado para identificação de emoções.

1.2 PUBLICAÇÕES RELACIONADAS

Os resultados das pesquisas desenvolvidas durante este trabalho, até o momento, resultaram nas seguintes publicações:

- Em periódicos:
 - ✓ Diniz, F. A.; Mendes Neto, F. M.; Lima Júnior, F. C.; Fontes, L. M. O. RedFace: Um Sistema de Reconhecimento Facial para Identificação de Estudantes em um Ambiente Virtual de Aprendizagem. *RENOTE. Revista Novas Tecnologias na Educação*, v. 10, p. 1-11, 2012.
 - ✓ Diniz, F. A.; Mendes Neto, F. M.; Lima Júnior, F. C.; Fontes, L. M. O. RedFace: Um Sistema de Reconhecimento Facial Baseado em Técnicas de Análise de Componentes Principais e Autofaces. *Revista Brasileira de Computação Aplicada*, 2013.
 - ✓ Diniz, F. A.; Mendes Neto, F. M.; Lima Júnior, F. C.; Fontes, L. M. O. *A Facial Recognition System Based on Techniques of Principal Component Analysis and Autofaces with K-NN, K-Star and Random Forest Classifiers. Research Notes in Information Science (RNIS)*, 2013.
 - ✓ Fontes, L. M. O.; Mendes Neto, F. M.; Diniz, F. A.; Carlos, D. G.; Jácome Júnior, L.; Silva, L. C. N. *An Animated Pedagogical Agent to Support Problem-Based Learning. IEEE-RITA*, 2013.
 - ✓ Fontes, L. M. O.; Mendes Neto, F. M.; Diniz, F. A.; Carlos, D. G.; Jacome Júnior, L.; Silva, L. C. N. Um Agente Pedagógico Animado de Apoio à Aprendizagem Baseada em Problema. *IEEE-RITA*, v. 7, p. 181-188, 2012.
- Em Eventos:
 - ✓ Diniz, F. A.; Mendes Neto, F. M.; Lima Junior, F. C.; Fontes, L. M. O. *A Facial Recognition System Based on Techniques of Principal Component Analysis and Autofaces with K-NN, K-Star and Random Forest Classifiers. In: 7th International Conference on Computer Sciences and Convergence Information Technology (ICCIIT), Seoul, IEEEExplore Digital Library*, 2012.
 - ✓ Diniz, F. A.; Mendes Neto, F. M., Fontes, L. M. O. RedFace: Um Sistema de Reconhecimento Facial Baseado em Técnicas de Análise de Componentes Principais e Autofaces *In: IV Congreso Internacional de Computación Y Telecomunicaciones (COMTEL)*, Lima, 2012.
 - ✓ Diniz, F. A.; Mendes Neto, F. M.; Lima Junior, F. C.; Fontes, L. M. O. RedFace: Um Sistema de Reconhecimento Facial para Identificação de Estudantes em um Ambiente Virtual de Aprendizagem *In: XX Ciclo de Palestras Novas Tecnologias na Educação*, Porto Alegre, 2012.

- ✓ Diniz, F. A.; Mendes Neto, F. M.; Lima Junior, F. C.; Fontes, L. M. O. Um Sistema de Reconhecimento Facial Aplicado a um Ambiente Virtual de Aprendizagem Compostos por Agentes Pedagógicos. *In: VIII International Conference on Engineering and Computer Education 2013 (ICECE2013), Proceedings of VIII International Conference on Engineering and Computer Education*, Luanda: UniPiaget, 2013.
- ✓ Diniz, F. A.; Mendes Neto, F. M.; Lima Júnior, F. C.; Fontes, L. M. O. A Facial Recognition System for Students Identification in a Virtual Learning Environment. *In: 2013 International Conference on Social Science and Health (ICSSH 2013)*, 2013, Los Angeles, CA, USA. *Advances in Education Research*. Los Angeles: CPCI-SSH (ISSHP), 2013.
- ✓ Diniz, F. A.; Mendes Neto, F. M.; Lima Júnior, F. C.; Fontes, L. M. O. A Facial Recognition System for Students Identification in a Virtual Learning Environment Composed by Pedagogical Agents. *In: 2013 World Conference on Information Systems and Technologies (WorldCIST'13)*, 2013, Algarve. *Proceedings of 2013 World Conference on Information Systems and Technologies*, 2013.
- ✓ Fontes, L. M. O.; Mendes Neto, F. M.; Diniz, F. A.; Carlos, D. G.; Jácome Júnior, L.; Silva, L. C. N. Uma Arquitetura Multiagente de Apoio ao Processo de Ensino e Aprendizagem no Moodle Baseado na PBL. *In: VIII International Conference on Engineering and Computer Education - ICECE 2013, 2012, Luanda. Proceedings of VIII International Conference on Engineering and Computer Education - ICECE 2013*. Luanda: COPEC, 2013.
- ✓ Fontes, L. M. O.; Mendes Neto, F. M.; Diniz, F. A.; Carlos, D. G.; Jácome Júnior, L.; Silva, L. C. N. 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, 2012, Rio de Janeiro. Anais do XXIII Simpósio Brasileiro de Informática na Educação*. Rio de Janeiro: UFRJ, 2012.
- ✓ Diniz, F. A.; Mendes Neto, F. M.; Lima Júnior, F. C.; Fontes, L. M. O. A Facial Recognition System Online for Students Identification in a Virtual Learning Environment. *In: 5th International Conference on Computer*

Supported Education - CSED 2013, 2013, Aachen, Germany. Proceedings of 5th International Conference on Computer Supported Education. Aachen, Germany: CSEDU, 2013.

- ✓ Diniz, F. A. ; Mendes Neto, F.M.; Lima Junior, F. C. ; Fontes, L. M. O. ; Domingues, T. A. L. ; Figueiredo, L. S. . A Facial Expression Recognition System to Support a Virtual Learning Environment. *In: ICIIIP2013: 2013 8th International Conference on Intelligent Information Processing*. Seoul: ACM Digital Library, 2013.

1.3 ORGANIZAÇÃO DA DISSERTAÇÃO

Nesta seção, ao apresentar a organização desta dissertação, pretende-se direcionar o leitor sobre as linhas seguidas ao longo do seu desenvolvimento. Esta dissertação é dividida em capítulos conforme apresentado a seguir:

- **O Capítulo 2 – Fundamentação Teórica.** Neste capítulo são descritos os principais conceitos necessários ao entendimento e implementação dos objetivos traçados, com um estudo sobre os principais aspectos da biometria, suas características gerais, os sistemas biométricos, as limitações e métricas de desempenho e algumas aplicações de sistemas biométricos. Neste capítulo também se detalha estudos de VC. É apresentado um estudo sobre o estado-da-arte de abordagens de detecção e reconhecimento facial. São estabelecidos alguns entendimentos básicos sobre AVAs e uma visão geral de um tipo de AVA baseado em agentes de apoio à PBL, no qual o sistema proposto foi aplicado. Por fim, dá-se ênfase nos estudos dos conceitos de emoção, do modelo de classificação das expressões faciais, Computação Afetiva, seguido de uma breve revisão bibliográfica sobre reconhecimento computacional de emoções por face.
- **O Capítulo 3 – Trabalhos Relacionados.** Neste capítulo são apresentados os trabalhos relacionados tanto com sistemas de reconhecimento faciais como com reconhecimento de expressões faciais.
- **O Capítulo 4 – Trabalho Proposto.** Descreve os passos e as técnicas adotadas no desenvolvimento do sistema. São apresentadas as arquiteturas usadas para a implementação do módulo de reconhecimento facial e do módulo de

reconhecimento expressões faciais. São detalhados os métodos usados, os resultados experimentais e as discussões dos resultados.

- **O Capítulo 5 – Considerações Finais e Trabalhos Futuros.** São apresentadas as contribuições e conclusões das atividades desenvolvidas durante o projeto, bem como perspectivas de trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo descreve aspectos relacionados ao sistema proposto.

A Seção 2.1 traz uma discussão sobre as definições de imagens digitais.

A Seção 2.2 define os conceitos de biometria, mostrando suas principais características, seus principais métodos biométricos, tipos de processos de identificação de um indivíduo e os principais sistemas biométricos existentes na literatura com sua arquitetura padrão e avaliações de seus desempenhos.

A Seção 2.3 define a área Visão Computacional (VC), detalhando suas características mais importantes no reconhecimento de imagens de um objeto específico. Mostra as principais etapas durante o processo de reconhecimento das faces, detalhando os principais métodos existentes tanto na detecção da face como na detecção de características. E, por fim, descreve as duas principais bibliotecas existentes e mais utilizadas pela comunidade científica que automatizam o processo de VC.

A Seção 2.4 detalha os sistemas de reconhecimento facial. Descrevendo suas características, os fatores que atrapalham o reconhecimento facial e as principais etapas do projeto de um sistema de reconhecimento facial.

A Seção 2.5 detalha o que são AVAs, descrevendo suas principais características. Apresenta e descreve o Moodle, um dos mais importantes AVA usado na literatura. E, por fim, detalha um AVA baseado em agentes de apoio à PBL, no qual o sistema proposto neste trabalho está integrado.

A Seção 2.6 aborda os conceitos de emoções, resumindo suas definições e características. Mostra as diferentes maneiras de comunicar as emoções. Detalha a face como medida de avaliação da expressão emocional. E, por fim, descreve as teorias existentes que diferenciam as emoções.

A Seção 2.7 descreve as principais características do FACS para inferência de emoções, detalhando o que são UAs e o mapeamento dos UAs para a classificação das emoções.

A Seção 2.8 resume os conceitos de afetividade e sua importância nos sistemas educativos. Define o que é afetividade e suas características mais importantes para dar aos computadores as habilidades emocionais. Especifica os principais recursos que fazem os sistemas educativos serem afetivos. Mostra as três principais linhas de pesquisas, nos quais os

trabalhos existentes estão abordando sobre a afetividade. Por fim, detalha as principais formas de mapeamento de reconhecimento das emoções de uma pessoa e analisa o uso de Computação Afetiva no AVA.

A Seção 2.9 descreve as características dos sistemas automatizados de reconhecimento de emoções de um indivíduo. Detalha a importância do reconhecimento de expressões faciais na descoberta das emoções de uma pessoa. Verifica os principais elementos que compõem a expressão facial. Discute alguns dos principais fatores que influenciam no reconhecimento das emoções faciais. E, por fim, resume as principais etapas envolvidas no reconhecimento da expressão facial, detalhando algumas técnicas para o reconhecimento da expressão.

2.1 IMAGENS DIGITAIS

Uma imagem é uma função de intensidade de luz bidimensional denotada por $f(x,y)$, onde x e y denotam coordenadas espaciais e o valor de f no ponto (x, y) é proporcional a intensidade luminosa (brilho) neste ponto. A Figura 1 destaca um ponto no olho da mulher com intensidade luminosa $f(x,y)$.



Figura 1 – Exemplo de imagem Digital monocromática (Falcão; Leite, 2003).

A fim de processar uma imagem digital no computador, é preciso realizar uma amostragem (uniforme) de $f(x, y)$ nas coordenadas espaciais, junto com o valor da intensidade luminosa, gerando uma matriz de $N \times M$ amostras, com cada elemento sendo uma quantidade discreta (GONZALEZ *et al.*, 2003) citado por (ROLIM e BEZERRA, 2008). Nesta matriz, cada elemento $p(x,y)$, $x = 0,1,\dots, M-1$ e $y = 0,1,\dots, N-1$, é chamado de *pixel* (*picture element*), o qual é o menor elemento de uma imagem digital. Os *pixels* podem assumir valores que variam

de 0 a 254 numa escala de 255 tons de cinza. Uma imagem digital em escala de cinza é considerada como uma matriz onde seus elementos representam apenas a intensidade luminosa naquele ponto. A Figura 2 exemplifica uma representação de uma imagem digital monocromática de tamanho 10x10 *pixels*, com cada *pixel* tendo seus respectivos valores numéricos de intensidade luminosa.

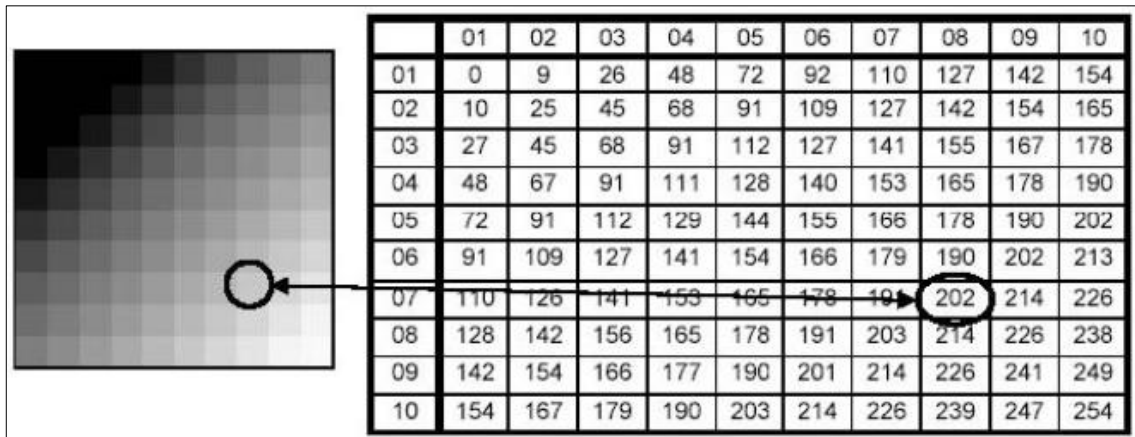


Figura 2 – Representação numérica de uma imagem digital monocromática (ROLIM e BEZERRA, 2008).

Outros conceitos interessantes de se observar em uma imagem são a resolução espacial e a profundidade da imagem. Ambos indicam a quantidade de amostras numa matriz $N \times M$ e quantidade de níveis de cinza L . Ambos são necessários para gerar uma boa imagem digital.

Por exemplo, a **resolução espacial** da imagem é 200x300 *pixels*, quando uma imagem $f(x,y)$ contida numa região retangular de 40cm em x por 20cm em y . Se tivermos amostras com espaços uniformes a cada 1 mm em x e em y , ou seja, com *pixel* de dimensões 1mm x 1mm, teremos uma matriz com $N=200 \times M=300$ amostras, totalizando 60000 *pixels*. Já para dizer, por exemplo, que a **profundidade da imagem** é 8 bits (*binary digit*) por *pixel* (1 byte por *pixel*), quando o nível de cinza é $L= 256$ (valor da escala de cinza variando entre 0 e 255), o computador requer no máximo 8 bits para ser armazenado na memória do computador. Numa imagem de resolução 200x300 *pixels* necessitam de oitenta mil *kilobytes*.

Logo, quanto maior a densidade do espaço da imagem, ou seja, menor a distância entre os *pixels*, maior o grau de detalhamento possível na imagem, e quanto maiores os valores possíveis assumidos pelos *pixels*, maior a qualidade da imagem. No entanto, outros fatores como os ruídos, por exemplo, interferem na qualidade da imagem (ROLIM e BEZERRA, 2008).

As Figuras 3 e 4 exemplificam variações de resolução espacial e profundidade, respectivamente, em uma imagem. Na Figura 3, o *pixel* começa a ser perceptível, conseqüentemente, nota-se uma degradação sofrida pelas imagens por causa da perda da resolução espacial.



Figura 3 – Variação de resolução espacial de uma imagem com 256 níveis de cinzas (FALCÃO e LEITE, 2003).

Já na Figura 4, os detalhes da imagem original cedem a regiões homogêneas de níveis de cinza.



Figura 4 – Variação de profundidade de uma mesma imagem (FALCÃO e LEITE, 2003).

Uma imagem colorida pode ser representada em diferentes espaços de cores. O mais comum é o espaço de cor chamado de RGB (Iniciais de *Red*, *Green* e *Blue*). Por exemplo, a imagem é decomposta nas cores vermelho, verde e azul, transformando a imagem em uma matriz tridimensional, estabelecendo nos *pixels* uma intensidade para cada uma das três cores, a soma dos três valores forma a cor final (FALCÃO e LEITE, 2003) citado por (ROLIM e BEZERRA, 2008). Logo, quanto maiores as intensidades luminosas, maior será o número de cores da imagem, melhorando a qualidade da imagem e, conseqüentemente, maior será o seu tamanho.

A seção seguinte define os conceitos de biometria, mostrando suas principais características, seus principais métodos biométricos, tipos de processos de identificação de um indivíduo e os principais sistemas biométricos existentes na literatura com suas arquiteturas padrão e avaliações de seus desempenhos.

2.2 BIOMETRIA

A biometria pode ser definida como a análise das características fisiológicas ou psicológicas para a identificação automática de um indivíduo, para um determinado propósito (HERNÁNDEZ, 2008). Proporciona uma medida da vida, propondo uma descrição mais minuciosa de como as características fisiológicas ou psicológicas são mensuráveis e de como podem ser registradas para verificar identidade de uma pessoa.

Na biometria, as partes do corpo da pessoa é a sua própria chave e não há como perdê-la. Esta técnica deve tornar obsoletos os sistemas de verificação de identidade por números, deixando os novos sistemas mais seguros (HERNÁNDEZ, 2008). A tecnologia de biometria resolve problemas que não podem ser solucionados pelos sistemas atuais, como senhas esquecidas, perdidas ou furtadas. Além disso, é possível verificar se o indivíduo a ser identificado é realmente aquele que está fisicamente presente ao local da identificação.

Existem diversos requisitos que decidem a adequação de um traço biométrico ao desenvolvimento de uma determinada aplicação. Segundo Jain *et al* (2004), os principais recursos desejáveis a uma determinada característica biométrica são:

- **Singularidade:** a característica deve ser a única, distinta para cada pessoa;
- **Permanência:** a característica não deve se alterar ao longo do tempo;
- **Universalidade:** Todos os indivíduos precisam possuir a informação a ser verificada;
- **Desempenho:** Analisa a precisão e velocidade do reconhecimento, as características necessárias para trazer um reconhecimento com precisão e velocidade aceitáveis e os fatores operacionais e ambientais que afetam a acurácia e velocidade do reconhecimento;
- **Mensuralidade:** A informação tem que ser passível de medição;
- **Aceitabilidade:** Verifica o grau de aceitação das pessoas no sistema biométrico;

- **Proteção:** Dificuldade ou facilidade de burlar o sistema por meio de métodos fraudulentos.

Nenhuma particularidade biométrica possui todos esses requisitos. Consequentemente, não há característica biométrica ótima.

Neste trabalho, como dito no capítulo da Introdução, foi explorado o reconhecimento biométrico de pessoas e de suas emoções baseado em suas faces. Dentre as vantagens do reconhecimento biométrico pela face destacam-se: (i) traço biométrico mais comumente usado por indivíduos no reconhecimento de pessoas; (ii) permite identificação a distância, sem contato com a câmera e, até mesmo, sem conhecimento do indivíduo; e (iii) existência da universalidade das emoções das pessoas por meio das expressões faciais (JAIN *et al.*, 2004).

A Tabela 1 apresenta uma análise entre os principais métodos biométricos em termos dos requisitos desejáveis, ilustrando seus pontos fortes e suas limitações (JAIN *et al.*, 2004). O valor de cada campo da Tabela 1 é o nível com o qual uma técnica biométrica atende o requisito: baixo, médio e alto.

Tabela 1 – Análise entre os principais métodos biométricos (JAIN *et al.*, 2004).

Método	Universalidade	Permanência	Mensuralidade	Desempenho	Aceitabilidade	Proteção
Biométrico						
Impressão Digital	Médio	Alto	Médio	Alto	Médio	Médio
Face	Alto	Médio	Alto	Médio	Alto	Médio
Íris	Alto	Alto	Médio	Alto	Baixo	Alto
Geometria da mão	Médio	Médio	Alto	Médio	Médio	Baixo
Voz	Médio	Baixo	Médio	Baixo	Alto	Baixo
Assinatura	Baixo	Baixo	Alto	Baixo	Alto	Baixo

Ao cumprir esses requisitos, a escolha do método biométrico tende a ser mais confiável. Além disso, a combinação de mais uma biometria, dependendo da aplicação, torna os sistemas de identificação ainda mais robustos. Um sistema de identificação automático de pessoas baseado somente em impressões digitais ou em reconhecimento facial, dependendo de suas características, não vai conseguir satisfazer todas as exigências de funcionalidade e desempenho.

O reconhecimento facial é natural e não intrusivo, mas não é genuíno, enquanto que a verificação de impressão digital é autêntico, mas é intrusiva e pode causar resistência nos usuários, dependendo da aplicação.

2.2.1 IDENTIFICAÇÃO DE INDIVÍDUOS

O reconhecimento de indivíduos tem sido o suporte para possibilitar a interação de seres humanos na sociedade. Com o aumento na quantidade de indivíduos e a mobilidade geográfica entre grupos de pessoas, surgiu a necessidade de se criar meios para comprovar se uma pessoa realmente é quem informa ser.

Verificar se uma pessoa pode frequentar determinada instalação, se ela é procurada pelo governo, se lhe é permitido acessar dados em um computador são exemplos de aplicações atuais da identificação para segurança.

De acordo com Penteado e Marana (2009), o processo de identificação de pessoas pode ser classificado em três tipos, como mostrado na Figura 5:

- **Posse:** Algo que um indivíduo tem posse. Por exemplo, um cartão de crédito;
- **Conhecimento:** Uma informação conhecida por uma pessoa. Por exemplo, os caracteres da senha.
- **Características Biométricas:** Uma informação anatômica, por exemplo, a face, impressão digital e íris. Uma informação fisiológica, por exemplo, calor do rosto. E uma informação comportamental, por exemplo, a voz e assinatura do indivíduo.

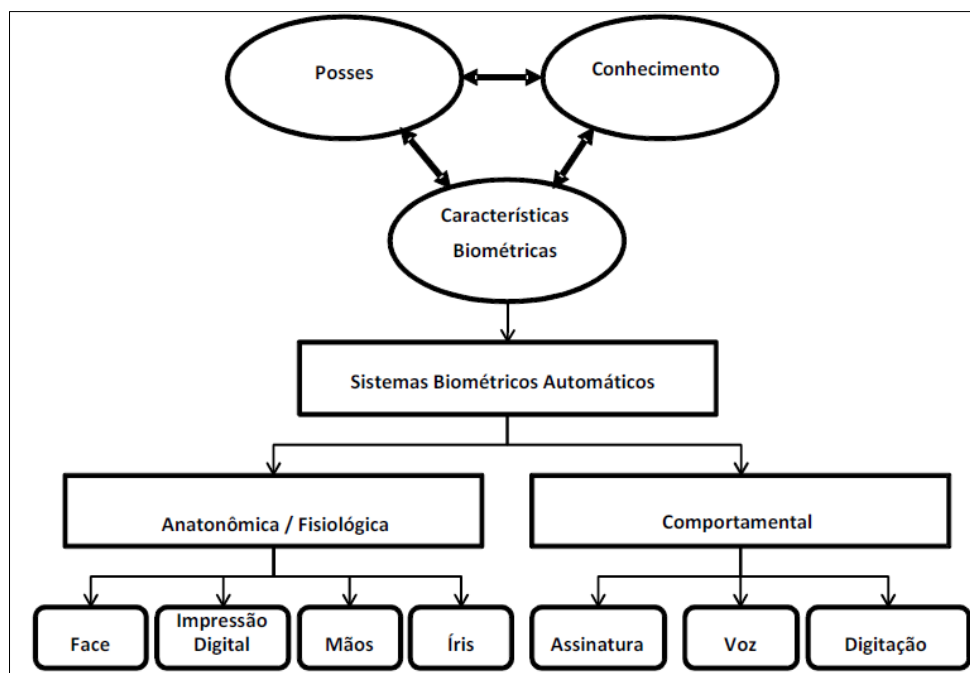


Figura 5 – Alguns tipos de métodos de identificação existentes (PENTEADO e MARANA, 2009)

Portanto, sistemas automatizados, como a autenticação de usuários baseada em cartões (posse) ou senhas (conhecimento) podem facilmente ser perdidos ou esquecidos. Esses problemas podem ser contornados pelo uso de autenticação biométrica, que faz com que a interação homem-máquina para autenticação de usuários seja mais transparente, aumentando a segurança e a sua usabilidade. Um exemplo é o reconhecimento facial, porque satisfaz a maioria dos critérios na escolha da solução de biometria ideal: é fácil de executar, rápido, moderadamente conveniente e, principalmente, não é intrusivo para o usuário.

2.2.2 SISTEMAS BIOMÉTRICOS

Um sistema de biometria é, por essência, um sistema de reconhecimento de padrão que faz uma identificação particular, determinando a autenticidade de uma característica fisiológica ou psicológica específica de um usuário (JAIN *et al.*, 2004). Uma questão importante ao se projetar um sistema é determinar como um indivíduo será identificado. Dependendo do contexto, um sistema de biometria ou pode ser um sistema de verificação (autenticação) ou um sistema de identificação.

Em um sistema biométrico, é possível identificar a seguinte arquitetura padrão com os seguintes módulos, mostrados na Figura 6 (JAIN *et al.*, 2004):

- **Módulo Aquisição de Dados:** um leitor captura o dado biométrico do usuário. Os dados do usuário são armazenados pela primeira vez no sistema.
- **Módulo de Autenticação e Identificação:** os dados coletados são comparados aos *templates* na base de dados.
- **Módulo de extração de características:** o dado adquirido pelo leitor é avaliado, gerando um conjunto de dados que representa a característica capturada do indivíduo (*template*).
- **Módulo de comparação:** as características extraídas são comparadas com os *templates* armazenados no banco de dados de *templates*. O módulo valida a identidade ou fornece a classificação do indivíduo dentro dos *templates* cadastrados previamente.
- **Banco de dados de *templates*:** repositório da informação biométrica extraída dos indivíduos.

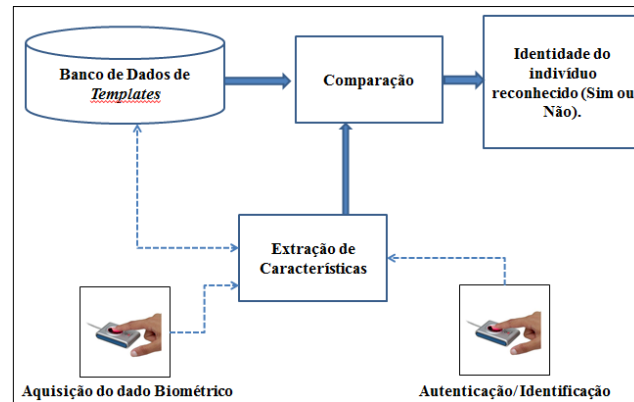


Figura 6 – Arquitetura padrão de um sistema biométrico. Adaptado de JAIN *et al.* (2004).

Os sistemas biométricos apresentam diversas limitações de desempenho relacionadas às amostras adquiridas entre os indivíduos. Dentre elas, podem-se destacar a pose, a iluminação, a escala e a qualidade da imagem, os quais podem levar a grandes variações nos *templates* gerados para o mesmo indivíduo (JAIN *et al.*, 2004). Outras limitações são as amostras biométricas de diferentes pessoas que podem apresentar variações muito pequenas entre si, como, por exemplo, a face de irmãos gêmeos.

Devido a essas limitações, os sistemas biométricos são avaliados na sua precisão de identificar os indivíduos (HERNÁNDEZ, 2008). Descobrir taxa de acurácia é uma forma de avaliar o quanto o sistema biométrico é preciso na identificação. Quanto maior a taxa de acurácia, maior é o grau de semelhança existente entre o dado de entrada e o *template* armazenado. Outras formas de avaliar os sistemas biométricos é descobrir: (i) *Falsa Aceitação* (FA): reconhecimento de uma amostra biométrica como sendo verdadeira, quando na verdade é falsa; (ii) *Falsa Rejeição* (FR): reconhecimento de uma amostra biométrica como sendo falsa, quando na verdade ela pertence ao indivíduo em questão.

As taxas de ocorrência de tais erros são críticas para se avaliar a eficiência do sistema. À vista disso, a avaliação de desempenho de um sistema biométrico leva em consideração os seguintes fatores (JAIN *et al.*, 2004):

- **Taxa de Falsa Aceitação (FA) ou taxa de falsos positivos:** probabilidade de que um indivíduo não autorizado seja autenticado. Pode ser estimada da seguinte maneira: $FAR = \text{número de falsas aceitações} / \text{número de tentativas de impostores}$;
- **Taxa de Falsa Rejeição (FR) ou taxa de falsos negativos:** probabilidade de que um indivíduo autorizado seja incorretamente rejeitado. Pode ser estimada da seguinte maneira: $FR = \text{número de falsas rejeições} / \text{número de tentativas de genuínos}$.

A próxima seção define a área VC, detalhando suas características mais importantes no reconhecimento de imagens de um objeto específico. Mostra suas principais etapas durante o processo de reconhecimento da face, detalhando os principais métodos existentes tanto na detecção da face como na detecção de características. E, por fim, descreve as duas principais bibliotecas existentes e mais usadas pela comunidade científica que automatizam o processo de VC.

2.3 VISÃO COMPUTACIONAL

Para que uma face e a expressão facial sejam reconhecidas, é preciso seguir uma metodologia coerente. As aplicações da área de VC são as mais adotadas como metodologias para a detecção da face e das expressões faciais pelo computador. Nesta área de pesquisa, utilizam-se métodos computacionais que tornam possível realizar a detecção, a extração e a interpretação das informações presentes nas imagens.

A VC é uma ciência responsável pela extração de dados significativos a partir de imagens capturadas por câmeras de vídeo, sensores, *scanners*, entre outros dispositivos. Estes dados permitem reconhecer, manipular e pensar sobre os objetos que compõem uma imagem (CHIU e RASKAR, 2009). Dessa forma, a VC fornece ao computador uma infinidade de informações precisas a partir de imagens e vídeos, de forma que o computador consiga executar tarefas inteligentes baseado nesses dados extraídos das imagens pelos processos da VC.

Dentre as várias etapas encontradas no processo de VC, as mais importantes para a realização desta pesquisa são Detecção de Face e Detecção de Características. A detecção de face busca reconhecer em uma imagem a área referente ao rosto de uma pessoa. Já a detecção de características tenta encontrar determinadas informações características das imagens. A palavra “características”, em VC, refere-se a um determinado elemento informativo que se busca encontrar ou identificar em uma imagem. Por exemplo, as características de um rosto (como olhos, boca, sobrancelhas etc.), que serão buscadas neste trabalho, podem ser rastreadas em uma imagem e ter a sua área delimitada.

Assim, são apresentadas, nas próximas subseções, as principais características que envolvem a detecção e extração de uma imagem facial. Nas duas últimas subseções, apresentar-se-ão duas bibliotecas usadas em detecção de objetos, conhecidas como OpenCV e

Kinect. A primeira pertence à Intel (INTEL, 2007) e a segunda pertence à Microsoft (MICROSOFT RESEARCH, 2011).

2.3.1 DETECÇÃO FACIAL

A fase da detecção de faces revela os primeiros obstáculos a serem transpostos no que diz respeito à VC. Segundo Yang *et al* (2002) citado por (PATIL *et al.*, 2010), a tarefa básica de algoritmos desta etapa é determinar se há ou não qualquer rosto na imagem e, se estiver presente, retornar a localização e dimensão da face.

A localização precisa da face em um cenário é uma atividade essencial, pois influenciará no desempenho geral do sistema, visto que isso permitirá a extração futura de características que estão relacionadas somente à face em questão, eliminando as informações que não representam as características relacionadas à mesma.

No entanto, inúmeros agentes, externos e internos, afetam a eficiência dos algoritmos responsáveis pela constatação da presença ou não de uma face humana na imagem ou sequência de imagens analisadas. A detecção da face, portanto, torna-se desafiadora devido a fatores como: distinção na pose da face (frontal, perfil, inclinado), presença de elementos na face (barba, óculos), oclusão da face, orientação da imagem (rotação do eixo óptico da câmera) e condições da imagem (intensidade da iluminação e características da câmera).

Para construir sistemas totalmente automatizados que analisam as informações contidas em imagens de face, técnicas de detecção facial robustas e eficientes são obrigatórias. Numerosas técnicas têm sido desenvolvidas para detectar rostos em imagens estáticas e dinâmicas, de intensidades únicas ou coloridas.

Resumidamente, segundo Yang *et al* (2002) citado por (PATIL *et al.*, 2010), os métodos de detecção de face em imagens se classificam dentro de quatro categorias: métodos baseados em conhecimento, métodos baseados em características invariantes, métodos de comparação de padrões e métodos baseados em aparência.

- **Métodos baseados em conhecimento**

São métodos que usam alguma base de regras estabelecida a partir de um conhecimento pré-estabelecido sobre o problema, ou seja, esses métodos possuem regras que definem o que é uma face, de acordo com o conhecimento do pesquisador. Por exemplo, sabemos que as pessoas, em condições normais, possuem determinadas

características imutáveis, tais como: todas as faces teoricamente possuem dois olhos, um nariz e uma boca, distribuídos de maneira específica sobre a face. Com este conhecimento é possível estabelecer regras que reconhecem uma face humana (YANG et al., 2002) citado por (PATIL *et al.*, 2010).

Este método pode sofrer algumas desvantagens vindas da formação do conjunto. Se as regras são muito gerais, o sistema pode classificar regiões não face como face. Por outro lado, se o conjunto de regras for muito específico, o sistema pode não ser eficaz ao tentar reconhecer faces se estas não satisfizerem todas as regras, diminuindo a precisão da detecção (YANG et al., 2002) citado por (PATIL *et al.*, 2010). Outra desvantagem é a definição de faces em poses variadas. A principal vantagem desses métodos é o bom desempenho obtido na busca por faces em fundos de cenas não complexos, por exemplo, o fundo de um objeto que não possui muitos elementos em sua constituição, como uma parede branca.

- **Métodos baseados em características invariantes**

Em contraste com o método baseado em conhecimento, este método não utiliza regras já pré-definidas que dizem o que é uma face, ou seja, não existe um conhecimento prévio sobre o que é uma face. Segundo Yang *et al* (2002) citado por (PATIL *et al.*, 2010), esta técnica encontra primeiramente as características invariantes da face (cor da pele, textura, nariz, olhos, sobrancelhas, boca e entre outras) e, posteriormente, modelos estatístico são aplicados para confirmar se a imagem é uma face ou não. A principal desvantagem de tal abordagem é que tais características podem ser corrompidas devido às condições de iluminação ou algum tipo de ruído. A grande vantagem apresentada pelos métodos desta abordagem é que as características invariantes se mantêm e são localizadas independente de pose ou orientação (YANG *et al.*, 2002) citado por (PATIL *et al.*, 2010).

As características invariantes mais utilizadas em métodos que utilizam essa abordagem são: características faciais, em que alguns métodos desta abordagem tentam localizar características faciais que mantêm certas propriedades em diferentes faces (caso dos dois olhos, da boca, do nariz); textura que usa a característica da pele e do cabelo presente em faces que podem diferenciar faces humanas de outros objetos; cor de pele, a qual se pode encontrar a região de uma face.

- **Métodos de comparação de padrões**

É uma técnica clássica de detectar objetos. O objeto é procurado em uma imagem através de um modelo padrão que é definido manualmente ou parametrizado

por uma função. Logo, nessa abordagem, alguns padrões ou modelos são usados. Estes modelos formados podem ser o rosto todo ou partes do rosto, como nariz, olhos, lábios, ou uma combinação deles ou podem ser figuras geométricas, como triângulo, quadrado, círculo, e outras figuras.

Os dois principais métodos de comparação de modelos utilizados para detectar faces são: modelos pré-definidos e modelos deformáveis. Nos modelos pré-definidos, as técnicas normalmente tentam, em uma primeira etapa, extrair as bordas e contrastes de imagens que podem estar relacionadas a contornos e a características de uma face. Os dados extraídos são posteriormente comparados com os modelos de faces que detectam ou não uma face, conforme suas aderências (YANG *et al.*, 2002) citado por (PATIL *et al.*, 2010). Nos modelos deformáveis, os métodos tentam localizar os contornos utilizando técnicas de identificação de bordas. Os métodos da abordagem de comparação de modelos apresentam como vantagem, a simplicidade de implementação, flexibilidade e eficiência, porém, existem dificuldades para definição dos modelos devido às variações de pose, de escala e de forma de faces.

- **Métodos baseados na aparência**

Em contraste com os métodos de comparação de modelos, os quais modelos são pré-definidos por especialistas, os métodos baseados em aparência passam por um processo de aprendizado utilizando exemplos de imagens. Em geral, os métodos baseados na aparência dependem de técnicas de análise estatística e de aprendizagem de máquina para encontrar as características relevantes da face e não face da imagem. Desta forma, as informações necessárias para realizar a detecção das faces são retiradas do conjunto de imagens (YANG *et al.*, 2002) citado por (PATIL *et al.*, 2010).

Os métodos baseados na aparência são muito aplicados devido a sua eficiência e robustez em comparação aos demais. Estes métodos também proporcionam uma elevada taxa de acerto e rapidez na detecção da face com poses e orientações variadas. No entanto, é necessária uma grande base de dados de exemplos de faces e não faces para o treinamento dos classificadores e realização de buscas sobre toda a imagem em escalas variadas.

Na Subseção 2.3.1.1, é apresentada uma breve descrição de um método utilizado neste trabalho para a detecção de faces, o qual utiliza um método baseado em aparência, realizando a detecção de faces frontais de forma rápida e robusta em tempo real.

2.3.1.1 MÉTODO DE VIOLA E JONES

A técnica de Viola-Jones, desenvolvida por Viola e Jones (2001) citado por (EPHRAIM *et al.*, 2009), é uma das técnicas de detecção de faces mais utilizada no meio acadêmico e na indústria. Sua grande vantagem é a baixa complexidade computacional, podendo realizar a localização de face em uma imagem de dimensão 320x240 em tempo real.

Este método não usa características como o movimento de um vídeo nem da cor para aumentar o processamento da imagem. Ele utiliza a ideia de “Janela de Processamento”, a qual extrai características da imagem, para depois aplicar uma árvore de decisão, que iterativamente irá informar a existência ou não de uma face nesta janela.

Segundo Viola e Jones (2001) citado por (EPHRAIM *et al.*, 2009), existem três módulos essenciais na formação do algoritmo Viola-Jones: (1) a Integral da Imagem, que permite uma rápida avaliação da característica de cada janela, sem trabalhar com a luminosidade do *pixel*, (2) a utilização do algoritmo *Adaboost*, o qual seleciona o mínimo de características (*Haar-like*) a serem treinadas e (3) a criação de uma estrutura em árvore, chamada de cascata de classificadores.

A extração de características da face é simples, pois não utiliza diretamente a intensidade do *pixel*, o que torna a atividade de aprendizagem mais fácil, diminuindo seu custo computacional. As características utilizadas são as do tipo *Haar* (*Haar-like features*), as quais são representações retangulares baseadas em *Haar wavelets* (Viola e Jones, 2001) citado por (EPHRAIM *et al.*, 2009). No modelo de Viola-Jones, a característica é a soma de todos os *pixels* da região branca subtraída da soma dos *pixels* da região preta. Este resultado representará o valor encontrado pela característica para uma região específica da face (Figura 7).

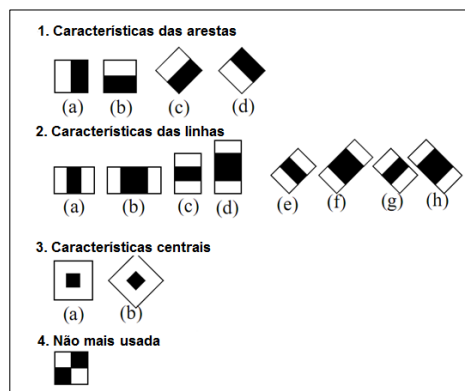


Figura 7 – Características *Haar*. Adaptado de Lienhart e Maydt (2002).

Viola e Jones (2001) citado por (EPHRAIM *et al.*, 2009) originaram quatro tipos de características *Haar* mostradas na Figura 7 (subfiguras 1a, 1b, 2a e 4). Lienhart e Maydt (2002) adicionaram mais características, sendo acrescentadas mais algumas verificadas na Figura 7 (subfiguras 1c, 1d, 2b, 2c, 2d, 2e, 2f, 2g, 2h, 3a e 3b) e retirada uma da Figura 7 (subfigura 4), a qual foi substituída pela combinação das subfiguras 2g e 2e. Nota-se que algumas características aparecem inclinadas a 45 graus, posição adotada por Lienhart e Maydt (2002) para aumentar o desempenho da proposta original.

Um método convencional para calcular a característica *Haar* da Figura 7 seria a utilização de dois laços aninhados para realizar a soma de cada uma das regiões de cada retângulo e, em seguida, realizar a subtração de cada um. Utilizando a Figura 8 como exemplo, o ponto x contém o somatório da área do retângulo A. O ponto z corresponde ao somatório dos retângulos A e B, da mesma forma que o ponto t é igual a $A + C$ e o ponto y , $A + B + C + D$. A soma dos *pixels* na área do retângulo D é obtido pelo cálculo entre os pontos: $y + x - (z + t)$. Os processos seguintes realizados pelo modelo serão executados com base na imagem integral obtida.

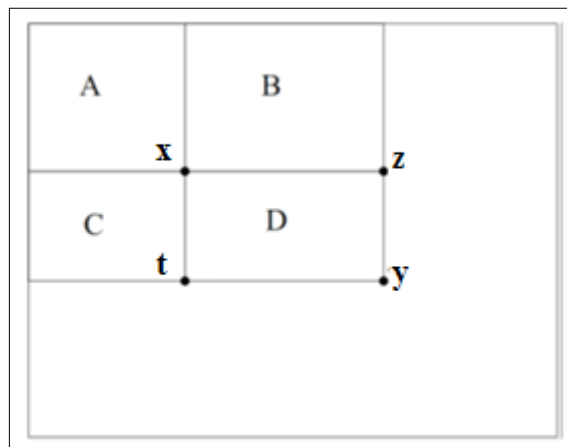


Figura 8 - Exemplo de cálculo de uma região retangular (VIOLA e JONES, 2001).

A fim de selecionar características *Haar* específicas, o método Viola e Jones utiliza um método de aprendizagem de máquina chamado *AdaBoost*. O *AdaBoost* combina muitos classificadores "fracos" para criar um classificador "forte". O termo "fraco" significa que o classificador só obtém a resposta certa um pouco mais frequentemente do que o faria se adivinhasse aleatoriamente (VIOLA e JONES, 2001) citado por citado por (EPHRAIM *et al.*, 2009). No entanto, se houver um conjunto inteiro desses classificadores fracos, cada um obtendo resposta na direção certa, pode-se ter um classificador forte, combinando as precisões

de cada classificador fraco, a fim de obter a solução correta. *AdaBoost* combina um conjunto de classificadores fracos e atribui um peso para cada. Esta combinação ponderada é o classificador forte (VIOLA e JONES, 2001) citado por (EPHRAIM *et al.*, 2009).

O método Viola-Jones combina conjuntos de classificadores *AdaBoost* em cadeia, mostrados na Figura 9. Cada classificador C_i , $i = 1, 2, \dots, N$, é um distinto classificador *AdaBoost* com um número relativamente pequeno de classificadores fracos.

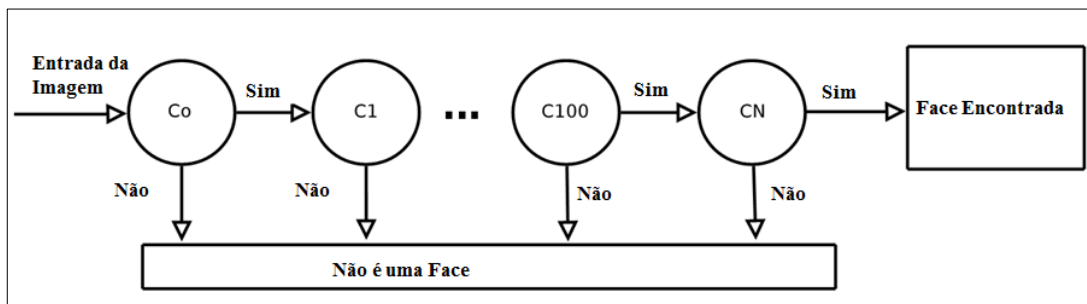


Figura 9 - Cascata de classificadores com n fases. Adaptada de Ma (2007).

O limiar de aceitação em cada nível é fixado como baixo o suficiente para passar todas, ou quase todas as faces de exemplo no conjunto treinado. Os classificadores são treinados em cada nível para classificar imagens de treinamento que passaram por todas as etapas anteriores. Segundo Viola e Jones (2001) citado por (EPHRAIM *et al.*, 2009), esses classificadores foram treinados por um conjunto de treinamento contendo milhares de faces. Durante a utilização, se qualquer um destes classificadores não classificar uma região de imagem, essa região é imediatamente classificada como "não-face". Quando um classificador passa uma região de imagem, ela vai para o próximo filtro na cadeia. Regiões de imagem que passam por todos os classificadores na cadeia são classificadas como "face". Essa abordagem é chamada de cascata. No Apêndice B é apresentado o algoritmo *AdaBoost* e o algoritmo de Treinamento utilizado na construção da cascata, ambos abordado no trabalho de Rolim (ROLIM e BEZERRA, 2008).

2.3.2 DETECÇÃO DE CARACTERÍSTICAS

Mesmo existindo técnicas que podem realizar a identificação e a extração de características independente de localização prévia do rosto, realizando a etapa de detecção

facial primeiro, consegue-se otimizar a tarefa de FeD, pois nestes casos, aplicados em vários algoritmos, a região onde se encontram as características reduz consideravelmente.

De modo geral, os métodos que são utilizados para a detecção facial são os mesmos para a detecção de características. Como os diversos elementos da face possuem características diferenciadas, geralmente são aplicadas técnicas distintas a cada segmento (vindos da segmentação da etapa de pré-processamento) para realizar a tarefa de extração. Segundo Bagherian e Rahmat (2008), as técnicas de extração de características mais usadas na detecção das características são classificadas em: baseado em geometria, baseado em modelo (*template*), baseado em aparência e técnicas de segmentação de cores.

- **Baseado em geometria:** as características são extraídas utilizando informações geométricas, tais como: as posições relativas e os tamanhos dos componentes da face. Estas técnicas exigem uma medida de limite (*threshold*) que pode afetar adversamente o desempenho alcançado.
- **Baseado em modelo:** esta técnica combina os componentes faciais aos modelos previamente projetados utilizando uma função de energia apropriada. Por conseguinte, esses algoritmos exigem uma modelagem do modelo à priori, proporcionando custos computacionais adicionais que afetam claramente o desempenho.
- **Baseado em aparência:** os métodos baseados em aparência contemplam aprendizagem a partir de exemplos de imagens. Métodos como PCA, análise de componentes independentes, e *Gabor-wavelets* são usados para extrair o vetor de características. Estas abordagens são comumente usadas para reconhecimento de rosto ao invés de identificação da pessoa.
- **Técnicas de segmentação de cores:** técnicas de segmentação de cores fazem uso da cor da pele para isolar a face. Qualquer região com cor que não seja pele dentro de uma face é visto como um candidato para os olhos ou boca. O desempenho de tais técnicas em bancos de dados de imagens faciais é bastante limitado, devido à diversidade de etnias.

A Tabela 2 mostra uma comparação das técnicas de extração de recursos faciais, exibindo as vantagens e desvantagens de ambas as técnicas (BAGHERIAN e RAHMAT, 2008).

Tabela 2 – Um comparativo das técnicas de detecção de característica. Adaptado de Bagherian e Rahmat (2008).

Técnicas de extração de características faciais	Vantagens	Desvantagens
Baseado em geometria	<ul style="list-style-type: none"> • Com imagens de rostos de baixa resolução. • Características geométricas. • Base de dados pequena. • Formas simples. 	<ul style="list-style-type: none"> • Grande número de características (olhos, nariz, boca, sobrancelhas, queixo e entre outros) são usadas.
Baseado em modelo	<ul style="list-style-type: none"> • Formas simples, flexíveis. • Eficientes 	<ul style="list-style-type: none"> • Descrição entre modelos e imagens tem um longo período de tempo devido a variações de pose, de escala e de forma de faces.
Baseado em aparência	<ul style="list-style-type: none"> • Base de dados pequena. • Formas simples. 	<ul style="list-style-type: none"> • Descontinuidade entre as cores. • Em perfil e com os olhos fechados tem um problema.
Técnicas de segmentação de cores	<ul style="list-style-type: none"> • Pequeno número de características. 	<ul style="list-style-type: none"> • Precisa de imagens de boa qualidade, grande tamanho da base de dados e iluminação.

A seção seguinte apresenta uma breve descrição do modelo de CANDIDE-3, técnica baseado em modelo.

2.3.2.1 MODELO CANDIDE-3

Neste trabalho usamos o modelo CANDIDE-3 (Figura 10) para rastrear a cabeça e as características faciais. O CANDIDE-3 é um modelo de face parametrizado, especificamente desenvolvido para codificação de faces humanas. É uma técnica baseado em modelo e é usado em vários projetos citados na literatura (DORNAIKA e AHLBERG, 2004; JIAN *et al.*, 2012).

Embora tenha um número baixo de polígonos, consistindo de 113 vértices, captura os mais importantes pontos característicos da face humana. Portanto, pode proporcionar um nível relativamente elevado de modelo e ao mesmo tempo permite uma reconstrução rápida com capacidade computacional moderada.

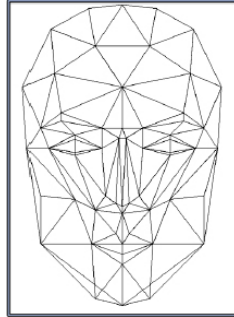


Figura 10 - Modelo CANDIDE-3 (KOTSIA *et al.*, 2007).

Este modelo de CANDIDE-3 contém um número razoável de pontos característicos e polígonos, de modo que esse modelo atinge um nível aceitável de qualidade, mantendo, ao mesmo tempo, o custo de funcionamento relativamente baixo. A Figura 11 exemplifica o mapeamento de alguns dos principais pontos característicos da face (FCPs).

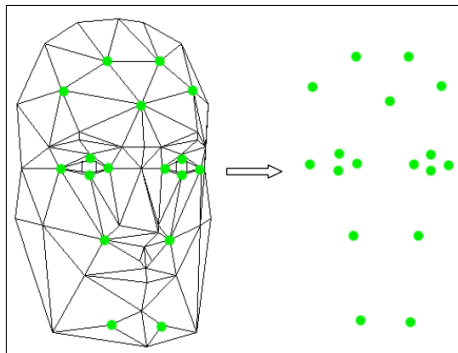


Figura 11 – Exemplo do CANDIDE-3 capturando alguns pontos característicos da face (MATTHEWS e BAKER, 2004).

O CANDIDE-3 é controlado pelas Unidades de Ações global e local (UAs) (AHLBERG, 2001) citado por (JIAN *et al.*, 2012). A UA global corresponde a uma rotação ao redor dos três eixos. Já a UA local controla as mímicas da face de modo que diferentes expressões podem ser obtidas (Figura 12). O modelo usando UAs locais permite fazer animações faciais como, por exemplo, abrir a boca ou levantar a sobrancelha. Com as

unidades de forma (*Shape Units* - SUs) é possível mudar a forma do modelo como, por exemplo, a localização da boca ou dos olhos (Figura 12).

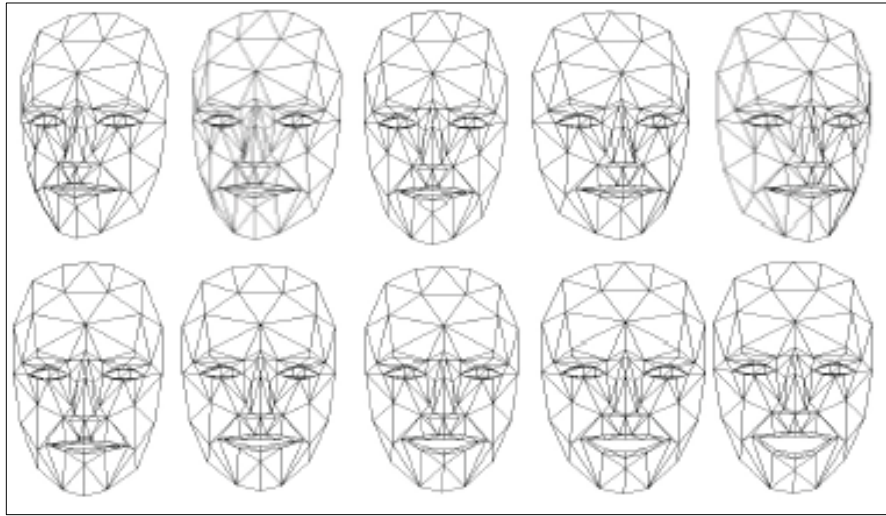


Figura 12 – O modelo CANDIDE-3 refletindo várias formas e expressões faciais. Adaptado de (MATTHEWS e BAKER, 2004).

A forma do modelo é dada por um conjunto de vértices e triângulos (AHLBERG, 2001) citado por (JIAN *et al.*, 2012). Pode representar o modelo como um g' (vetor- $3N$) contendo as coordenadas dos vértices e , onde N é o número de vértices, g' é a forma base do modelo que pode ser reformada usando g (AHLBERG, 2001) citado por (JIAN *et al.*, 2012):

$$g = g' + Sx + Ay \quad (1)$$

Onde S e A são respectivamente as unidades da forma e as unidades de animação, x e y respectivamente contém os parâmetros da forma e animação para remodelar a forma base. Incorporando o movimento global, a equação é reescrita para:

$$g = Rs(g' + Sx + Ay) + t \quad (2)$$

Onde R é a matriz rotação, s é a escala e t é o vetor translação. Ou pode ser escrito como:

$$g = M(g' + Sx + Ay) \quad (3)$$

Onde M é uma matriz transformação afim. O modelo pode então ser parametrizado por uma transformação afim, parâmetros de forma e animação.

Resumidamente, a estrutura do CANDIDE-3 é baseada em matriz. O conteúdo pode ser dividido em seis partes:

1. A primeira parte é uma descrição sobre as modificações de cada versão (versão 1 até a versão 3);

2. A segunda parte é uma lista de vértices. Há 113 vértices, e para cada ponto, coordenadas x, y e z são oferecidas. Então, é uma matriz 113 por 3. Todas as coordenadas numa escala entre -1.0 até 1.0;
3. A terceira parte é uma lista de faces, as quais são triângulos. Há 184 faces, e para cada face, 3 vértices são oferecidos. Então, é uma matriz 184 por 3. Nesta parte, somente índice de vértices são mencionados para desenhar uma face;
4. A quarta parte é uma lista de unidades de animação. Sessenta e cinco diferentes unidades de animações são fornecidas;
5. A quinta parte é uma lista de unidades de forma, as quais informam aos usuários como mudar a forma de cada organismo da face;
6. A última parte é sobre textura. Para este trabalho, somente a segunda e a terceira parte são consideradas como úteis.

Nas duas últimas subseções serão vistas as duas principais bibliotecas para o desenvolvimento de aplicações que envolvem VC. Essas bibliotecas implementam as principais técnicas de detecção de faces e de características faciais.

2.3.3 BIBLIOTECA OPENCV

OpenCV é uma biblioteca, livre para o uso acadêmico e comercial, desenvolvida pela Intel no ano de 2000 e voltada para o desenvolvimento de aplicações de tempo real no campo da VC. Originalmente desenvolvida em C/C++, ela é portátil para todas as plataformas de sistemas operacionais.

O OpenCV possui módulos de processamento de imagens e vídeo, estrutura de dados, álgebra linear, interface gráfica do usuário (GUI), controle de *mouse* e teclado, além de vários algoritmos, muitos dos quais são considerados estado da arte, tais como: segmentação, reconhecimento de faces, aprendizado de máquinas, filtragem de imagens, rastreamento de movimento, entre outros métodos (INTEL, 2007).

Atualmente o OpenCV está disponível para Windows, Unix, *Android*, Linux, Mac OS e GPUs Nvidia (CUDA). Possui um adaptador (*Adapter*) que permite que classes com interfaces incompatíveis possam interagir entre si, facilitando o suporte a programadores que

utilizam as seguintes linguagens de programação: Java (usando JavaCV), Python, Csharp, Ruby e Visual Basic.

Os algoritmos de processamento de imagens e aprendizado de máquina disponíveis na biblioteca OpenCV se encontram divididos em cinco módulos: CV + CVAUX, MLL, HighGUI, CXCORE e CVCam (BRADSKY *et al*, 2006).

As funções para processamento de imagens, como filtros, transformações geométricas, histogramas, detecção de cantos, detecção de bordas, pirâmides, transformações, análise de formas, análise de movimento, detecção de objetos, entre outras, estão implementadas no módulo CV + CVAUX.

O MLL (*Machine Learning Library*) é o módulo de aprendizado de máquina, onde é possível encontrar algoritmos de classificação estatística, regressão e agrupamento de dados, permitindo a utilização dos classificadores mostrados na Tabela 3.

Tabela 3 - Algoritmos de reconhecimento de padrão no OpenCV (BRADSKY *et al*, 2006).

Algoritmo	Descrição
<i>Mahalanobis</i>	Uma medida de distância que verifica a elasticidade do espaço dos dados.
K-médias	Um algoritmo de agrupamento não supervisionado.
Classificador Bayesiano	Um classificador que assume que as características são Gaussianas e estatisticamente independentes.
Árvore de decisão	Um classificador discriminativo.
<i>Boosting</i>	Um classificador de grupo discriminativo. A classificação geral depende da combinação de pesos dada a cada classificador.
Árvores randômicas	Um conjunto de árvores de decisão combinadas.
Classificador <i>Haar</i>	Um classificador baseado em <i>boosting</i> .
Maximização esperada	Um agrupador não supervisionado baseado na técnica generativa.
K-vizinhos	O classificador discriminativo mais simples possível.
Redes neurais	Baseado nos <i>perceptrons</i> de múltiplas camadas.
<i>Support Vector Machine</i>	Um classificador discriminativo capaz de fazer regressões.

O módulo HighGUI contém recursos de Interface Usuário para manipulação de janelas e funcionalidades para abertura, gravação e exibição de imagens estáticas, captura e saída para fluxo de vídeo, eventos de teclado, evento de *mouse* e barras de rolagem. O módulo CXCORE contém as estruturas de dados e funcionalidades para cálculo de matrizes, vetores, álgebra linear, operações lógicas, operações aritméticas, funções de desenho e Entrada/Saída de XML (*Extensible Markup Language*). O módulo CVCam que é um módulo multiplataforma para processamento das *streams* de vídeo de câmeras digitais.

Os principais objetivos vindos dos recursos disponíveis do projeto OpenCV são (BRADSKY *et al*, 2006):

- Avançar na pesquisa em VC, fornecendo não só o código aberto, mas também uma infraestrutura aperfeiçoada para VC;
- Disseminar o conhecimento de VC, fornecendo uma infraestrutura comum para os desenvolvedores (código mais facilmente legível e intransferível);
- Trazer avanços em aplicações comerciais, tornando portátil o código com desempenho otimizado disponível gratuitamente.

2.3.4 BIBLIOTECA DO KINECT

Uma interface natural para o usuário (*Natural User Interface* - NUI) é a mais recente forma de como um usuário interage com o computador (BUXTON, 2010). A NUI é caracterizada como um estilo de interface marcado pela invisibilidade do controle ao usuário. Segundo Buxton (2010), uma interface é natural se explora as habilidades que o usuário adquiriu durante toda a vida ao interagir normalmente com o mundo. Portanto, uma NUI deve ser aprendida e utilizada rapidamente, beneficiando-se e adaptando-se a partir da atuação do corpo humano.

O periférico Kinect, dispositivo desenvolvido pela Companhia PrimeSense em colaboração com a Microsoft, é um dispositivo baseado na NUI, pois elimina a necessidade de utilizar um controle físico (MICROSOFT, 2010). Mesmo que este dispositivo tenha sido criado originalmente para o Microsoft XBOX, ele é facilmente integrável a um computador pessoal com sistemas Windows / Mac OS X ou Linux. Ele é o aparelho de NUI mais vendido e popular no mundo. Segundo Microsoft (2011), até março de 2011 cerca de 10 milhões de

unidades foram vendidas, tornando-se um dispositivo amplamente utilizado na indústria (de jogos, robótica, artistas de teatro, de interfaces naturais, etc) e na pesquisa.

Constituído de um conjunto de câmeras e sistema de projeção, o Kinect retorna uma informação bastante exata de cor e da profundidade associada a cada ponto. O aparelho destaca-se também por devolver a posição (x, y e z) de um conjunto de pontos associados às juntas principais do corpo humano (cabeça, mãos, cotovelos, pernas, etc). O Kinect destaca-se, principalmente, por sua *performance* em reconhecimento de gestos, faces e voz. Logo, é desejável que o Kinect seja aplicado nesse trabalho, utilizando-o no processo de reconhecimento de expressões faciais, pois o Kinect possui os recursos necessários e essenciais para a captura dos pontos característicos da face que mapeiam a emoção de uma pessoa.

Existem duas versões do Kinect. A versão vendida junto com o console XBOX precisa de um cabo especial de alimentação quando usado no computador tradicional. A segunda versão já vem com este cabo integrado. Sua interface de dados é via USB (*Universal Serial Bus*). Esta interface não está criptografada (FRIED, 2011), fato que facilitou desde o início sua utilização fora do XBOX. O Kinect é formado por um emissor de luz infravermelha (invisível ao olho humano), uma câmera infravermelha, uma câmera RGB comum, um conjunto de microfones e um motor, conforme indica a Figura 13. O campo visual do Kinect é de 58 graus na horizontal; 45 graus na vertical; e 70 graus na diagonal, e a faixa de funcionamento situa-se entre 0,8 metros e 3,5 metros.

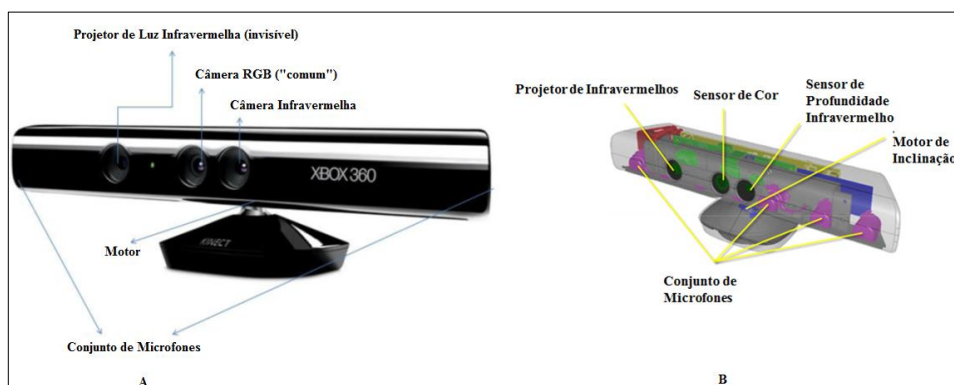


Figura 13 - (A) Visão modelo comercial do Kinect. (B) Visão figurada do modelo Kinect.

Adaptado de Microsoft *Research* (2011).

- **Projeto de luz Infravermelha:** consiste de um laser que emite Raios Infravermelhos (RI). O emissor de laser RI gera um padrão conhecido como luz IR estruturada ruidosa.

- **Câmera de Infravermelho:** é uma câmera que opera em 30 Hz (*Hertz*) que chega a enviar imagens com resoluções de 1200x960 *pixels*. Estas imagens podem ser reduzidas para 640x480 *pixels* com 11-*bits*, o que proporciona 2048 níveis de sensibilidade.
- **Câmera RGB:** é uma câmera RGB que opera em 30 Hz e pode enviar imagens em 640x480 *pixels* com 8 bits por canal. Também pode enviar imagens em 1280x1024 *pixels*, rodando 10 quadros (*frames*) por segundo. A câmera possui um conjunto de recursos, incluindo balanceamento de branco automático, referência preto, para evitar a cintilação, saturação de cor, e correção do defeito. A saída da câmera RGB é filtrada com um padrão de RG (*Red Green*), GB (*Green Blue*).
- **Motor, Acelerômetro e Microfones:** O Kinect tem dentro dois sistemas importantes e inter-relacionados: um motor que inclina a cabeça do Kinect para cima e para baixo, e um acelerômetro que determina qual a posição que a cabeça do Kinect está internamente. O conjunto de microfones apresenta quatro cápsulas de microfones e opera com cada canal de processamento 16-*bits* de áudio a uma taxa de amostragem de 16 kHz (*Quilo Hertz*).

De acordo com arquitetura do Kinect, os seguintes dados principais são repassados para os desenvolvedores de aplicações que utilizam esse dispositivo como área de trabalho (MICROSOFT RESEARCH, 2011) (Figura 14):

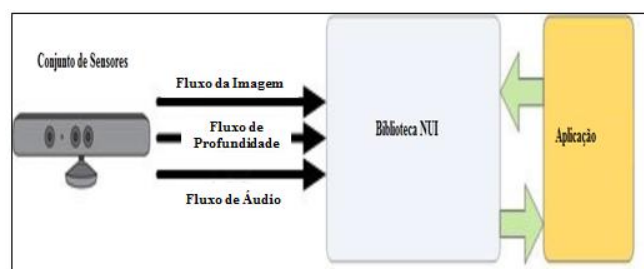


Figura 14 - Resumo da arquitetura básica. Adaptado de Microsoft Research (2011).

- Fluxo de Imagens (imagens): devolve cada *pixel*, representando uma cor, resolução de 640x480 *pixels* em 30 *frames* por segundo (fps), ou 1280x1024 em um máximo de 15 fps. É possível obter tanto a imagem da câmera RGB quanto a da câmera infravermelha;
- Fluxo de Profundidade (informação de profundidade): devolve cada *pixel* indicando a distância do objeto em relação ao aparelho. O aparelho detecta cerca

de 2000 níveis de sensibilidade e percebe objetos presentes de 1.2 a 3.5 metros à frente do aparelho. Além da informação de profundidade, também é possível retornar com exatidão se o *pixel* faz parte do corpo de um ser humano. O Kinect consegue diferenciar até 6 corpos humanos em sua visada.

- Fluxo de Áudio: devolve o fluxo de áudio, utilizando um conjunto de 4 microfones, e a anulação de ruído e eco. O Kinect permite a gravação de áudio e o reconhecimento da fala em inglês.

O principal destaque do Kinect está na no uso de seu projetor de profundidade em jogos e aplicativos, obtendo com bastante exatidão a distância de cada *pixel* de uma imagem em relação ao projetor. O funcionamento deste sensor não se baseia em princípio estereoscópico como nas câmeras de profundidade mais tradicionais. Um padrão infravermelho é projetado e a deformação neste padrão é interpretada como variações na distância do objeto, permitindo a inferência da distância. O padrão utilizado no Kinect é uma patente da companhia PrimeSense (PRIMESENSE, 2011).

Outra característica de sucesso do Kinect é seu aspecto híbrido, não se limitando em apenas devolver os dados de profundidade. Devolve, também, o *pixel* associado a um valor que indica se pertence ao corpo do ser humano. Assim, o aparelho permite a diferenciação de até seis usuários a sua frente. E entre essas seis pessoas, o Kinect também tem a capacidade de conseguir a informação de esqueleto de até dois usuários. Na Figura 15, mostra-se as articulações de um esqueleto reconhecido pelo Kinect.

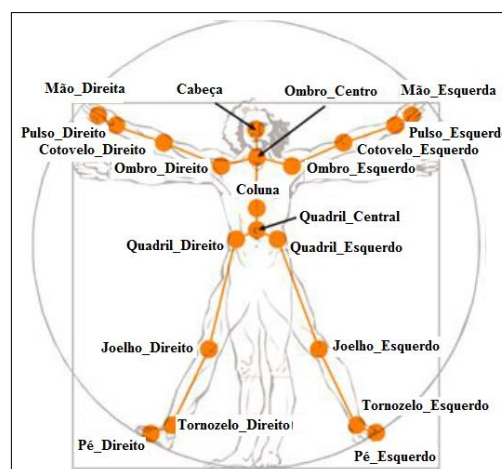


Figura 15 - Articulações do esqueleto do homem. Adaptado de Microsoft *Research* (2011).

As principais funcionalidades utilizadas em aplicações e desenvolvimento de bibliotecas para o Kinect estão disponíveis nos seguintes projetos: OpenNI, OpenKinect e Microsoft Kinect para Windows (CRUZ *et al.*, 2012). Algumas das bibliotecas utilizadas por estes projetos como suas implementações são: OpenCV, Unity3D, PCL, RGBDemo, openframeworks (CRUZ *et al.*, 2012). Dentre esses projetos, as mais robustas e consolidadas atualmente são OpenKinect e Microsoft Kinect para Windows. No entanto, a tendência é que todos esses projetos e outros se tornem mais estáveis.

Tanto o projeto OpenKinect como a OpenNI são abordagens abertas e gratuitas, sem nenhum tipo de restrição de uso ou licença. O projeto OpenKinect usa bibliotecas de código aberto que permitem o Kinect ser usado nas plataformas Linux, Mac e Windows. O foco principal do OpenKinect é o software *libfreenect*. O software *libfreenect* é a biblioteca central para acessar o Kinect por essas plataformas. A biblioteca suporta acesso a RGB e profundidade de imagens, motores, acelerômetro, entre outros. A biblioteca *libfreenect* é escrita em C, mas fornece meios para manipulá-la em várias línguas, como Python, ActionScript, C++, C# e Java.

O projeto Microsoft Kinect para Windows é constituído pelo hardware Kinect e o Kinect SDK (*Software Development Kit*) para Windows, o qual suporta aplicações construídas com C++, C# ou Visual Basic (MICROSOFT RESEARCH, 2011). O Kinect SDK para Windows oferece vários recursos, como reconhecimento de esqueleto sentado, rastreamento de esqueleto, rastreamento facial e reconhecimento de voz. Nesse trabalho, foram utilizados os recursos de rastreamento facial disponibilizados nessa biblioteca. O SDK é limitado a aplicações não comerciais, mas espera-se uma licença comercial nos próximos meses.

Na próxima seção detalha os sistemas de reconhecimento facial. Descreve suas características, fatores que atrapalham o reconhecimento facial e as principais etapas de como se projeta um sistema de reconhecimento facial.

2.4 RECONHECIMENTO DE FACES

O reconhecimento facial é um dos processos de identificação mais utilizados pelos seres humanos, pois permite identificar rapidamente qualquer indivíduo. Além de identificar, podemos perceber o estado emocional de uma pessoa apenas observando sua expressão facial

(EKMAN, 1993) citado por (KSHIRSAGAR *et al.*, 2011). Embora o reconhecimento facial seja uma tarefa simples para o ser humano, não é trivial implementar esse processo em uma máquina, pois não se sabe, ao certo, como o cérebro humano realiza essa tarefa (AGARWAL *et al.*, 2010).

Numerosos trabalhos vêm mostrando um considerável avanço na área de reconhecimento facial (EKMAN, 1993; GHOSAL *et al.*, 2009; KSHIRSAGAR *et al.*, 2011), sendo esta área um campo da VC que trata de imagens faciais e que envolve tarefas que são naturais a um ser humano, como detectar e reconhecer uma face humana em uma imagem (AGARWAL *et al.*, 2010).

Pode-se considerar que a atividade aplicada no processamento de reconhecimento de imagens faciais possui as seguintes fases mais importantes (AGARWAL *et al.*, 2010): Detecção de Faces, Extração e Reconhecimento Facial. Além dessas fases, vale destacar as atividades relacionadas à normalização de faces, as quais estudam métodos que são aplicados às imagens faciais a fim de conseguir uma face que potencialize a eficiência do reconhecimento.

A fase de detecção facial consiste em detectar uma face em uma imagem adquirida. A detecção pode ser feita diferenciando atributos visuais que distinguem uma face humana de outros objetos na cena (YANG *et al.*, 2002) citado por (PATIL *et al.*, 2010). A detecção facial desempenha um papel essencial em um sistema de reconhecimento, pois localiza uma face no cenário complexo e depois recorta a imagem da face, eliminando o fundo remanescente. Desta forma diminui quantidade de informações da imagem a serem reconhecidas, implicando na diminuição do processamento de classificação da imagem. Alguns métodos de detecção utilizam a busca de uma forma elíptica, outros procuram a textura da cor de pele e há os que procuram pelos órgãos da face, como olhos, boca, nariz, etc. A detecção facial é executada em nível de entrada.

A fase de extração envolve capturar características que resultem em informações quantitativas de interesse, que sejam básicas para discriminação entre classes de objetos. Como comentado previamente, a seleção das regiões que representam as características faciais é baseada na intensidade, cor, região com textura semelhante e forma.

A última fase é o reconhecimento, o qual consiste em identificar uma pessoa comparando sua face com a de outras pessoas armazenadas em um banco de imagens. As faces apresentadas para efeito de reconhecimento são comparadas com as faces conhecidas armazenados em uma base de dados, sendo então classificadas como sendo a face de um indivíduo conhecido ou como uma face desconhecida.

Existem pesquisas no desenvolvimento de várias técnicas de representação facial, que podem ser enquadradas em três abordagens distintas: Baseado em Modelo, Baseado em Características e Baseado em Aparência. Estas abordagens foram observadas na Seção 2.3.2.

Apesar de existirem diferentes abordagens, alguns problemas comuns a todas elas prejudicam o desempenho do reconhecimento da face, tais como (YANG *et al.*, 2002):

- **Pose:** Refere-se à variação da face de acordo com a posição da câmera, a qual captura alguns elementos faciais, tais como um olho ou o nariz, que podem aparecer parcial ou totalmente oclusos;
- **Expressão Facial:** Influencia diretamente na aparência da imagem de face;
- **Presença de Elementos Estruturais:** A presença de elementos como barba, bigode e óculos, que podem modificar as características em termos de tamanho e luminosidade da face;
- **Ocultação:** As faces podem aparecer parcial ou totalmente sobrepostas, por objetos ou até mesmo por outras faces;
- **Orientação da Imagem:** Imagens faciais podem variar diretamente sobre o eixo óptico da câmera, possibilitando, por exemplo, a captura de uma imagem de uma face de cabeça para baixo, proporcionando uma disposição não regular dos elementos faciais (olhos, boca, nariz etc);
- **Condições da Imagem:** Fatores como a variação da intensidade e distribuição luminosa, características da câmera (sensor de captura, lentes) afetam a aparência da face.

Nesta dissertação, as pesquisas desenvolvidas se basearam em métodos de representação de faces Baseado em Aparência, usando-se como suporte os autovetores das imagens e seus respectivos autovalores. Estes métodos utilizam o PCA e *Eigenfaces*, os quais decompõem as imagens de faces em um pequeno conjunto de características particulares das imagens, ambos serão detalhados na subseção seguinte.

2.4.1 ANÁLISE DOS COMPONENTES PRINCIPAIS E *EIGENFACES*

A análise dos componentes principais é uma técnica matemática que descreve um conjunto de dados usando “componentes principais”, escrita como combinações lineares dos

dados originais. Os componentes principais são determinados em ordem decrescente de importância. O primeiro componente possui mais informação do que o segundo, e assim por diante. O PCA tenta construir um pequeno conjunto de componentes que resumem os dados originais, reduzindo a dimensionalidade dos mesmos e preservando os componentes mais significativos (KSHIRSAGAR *et al.*, 2011).

Os componentes principais considerados mais importantes possuem as direções em que os dados de entrada têm as maiores variâncias e os outros componentes (componentes com as menores variâncias, chamados de componentes menores) são considerados menos importantes ou associados a ruídos.

O PCA é baseado nos autovetores e autovalores da matriz covariância dos dados (AGARWAL *et al.*, 2010). O objetivo é encontrar um conjunto de vetores ortonormais v_i que melhor descreve a distribuição dos dados de entrada. Seja C a matriz covariância estimada:

$$C = XX^T = VAV^T \in R^{n \times n} \quad (4)$$

Em que $A = \text{diag} \{ \lambda_1, \dots, \lambda_n \}$ é a matriz diagonal contendo n autovalores, v_i é o autovetor associado a λ_i e $V = [v_1, \dots, v_n] \in R^{n \times n}$ é a matriz dos autovetores, v_i , da matriz covariância.

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \quad (5)$$

Onde X é a matriz de dados originais, em que $x_{ij} \in R$, $i = 1, \dots, n$ e $j = 1, \dots, m$. O PCA determina uma transformação linear dos elementos de X , tal que:

$$y = V^T X \quad (6)$$

Onde $y = [y_1, y_2, \dots, y_n]$ é o vetor das componentes principais.

O método *Eigenfaces* busca um conjunto de características que não depende das formas geométricas da face (olhos, nariz, orelhas e boca), utilizando toda a informação da representação facial (KSHIRSAGAR *et al.*, 2011). Seu funcionamento é similar ao funcionamento do PCA, entretanto é utilizada uma leve otimização para reduzir a matriz de covariância, reduzindo o processamento necessário para fazer o cálculo de seus autovetores e autovalores.

Baseadas na Teoria da Informação, as *Eigenfaces* buscam identificar um pequeno número de características que são relevantes para diferenciar uma face de outras faces. Essas características podem ser analisadas apenas com a variação dos valores assumidos pelos *pixels*, em um conjunto de imagens de faces.

Os autovetores da matriz de covariância de um conjunto de imagens de faces descreve a variação dos *pixels* em uma região diferente das imagens, ou seja, significa que cada autovetor descreve a variação dos *pixels* associados a diferentes características faciais (TURK e PENTLAND, 1991) citados por (ROLIM e BEZERRA, 2008).

Depois de encontrar os autovetores, as imagens de faces são projetadas em um espaço de características (“espaço de faces”) que melhor representa a variação entre faces conhecidas. Este espaço é definido pelas *Eigenfaces*, que são os autovetores do conjunto de faces. O reconhecimento de uma face é realizado pela sua projeção no subespaço gerado pelas *Eigenfaces* e então pela comparação da posição obtida com a posição de indivíduos conhecidos (TURK e PENTLAND, 1991) citados por (ROLIM e BEZERRA, 2008).

A sequência dos procedimentos da geração das *Eigenfaces* abordados por esse trabalho foi (CHEON e KIM, 2008):

- I. Adquirir um conjunto de M faces. Onde Γ_i ($i \in N$) é cada face do conjunto da base de treinamento.

$$\Gamma = [\Gamma_1, \Gamma_2, \dots, \Gamma_M] \quad (7)$$

- II. Calcular a face média Ψ (2), para eliminar muita informação redundante na face, isto é, uma imagem de face poderia ser representada com poucas componentes principais devido à sua redundância.

$$\Psi = \frac{\sum_{i=1}^M \Gamma_i}{M} \quad (8)$$

- III. Criar uma matriz com as faces de treino com os *pixels* dispostos em linhas e as M faces do conjunto de treinamento dispostas em colunas.
- IV. Subtrair a imagem média Ψ de cada imagem de Γ , gerando uma nova matriz A (4) que contém somente as variações de cada face em relação à face média. O vetor Φ (3) contém todas as variações de uma determinada face Γ em relação à face média Ψ .

$$\Phi = \Gamma - \Psi \quad (9)$$

$$A = [\Phi_1, \Phi_2, \dots, \Phi_m] \quad (10)$$

- V. Sendo M menor que a dimensionalidade (largura * altura das imagens de treinamento) das imagens em Φ , calcular a matriz de covariância C (5).

$$C = AA^T \quad (11)$$

- VI. Calcular os autovetores v e autovalores λ da matriz C .
- VII. Criar a matriz de transformação. A matriz μ (6) conterá $M - 1$ autovetores significativos. Entretanto ainda é possível realizar a eliminação de alguns desses $M-1$ autovetores pela ordem de importância de seus correspondentes autovalores λ , gerando M' autovetores escolhidos.

$$\mu = v\Phi \quad (12)$$

VIII. Normalizar os M' vetores da matriz μ .

IX. As imagens do conjunto de treinamento são projetadas no espaço de faces, efetuando-se a operação de transformação (7). Em seguida o novo padrão passa para a fase de classificação.

$$\Omega = \mu^T \Phi \quad (13)$$

A Figura 16 exemplifica a face média de um conjunto de imagens em treinamento e algumas *Eigenfaces* calculadas para o conjunto de imagens de treino de exemplo.



Figura 16 - Exemplo de face média (acima) e *Eigenfaces* mais importantes no treinamento.

Fonte própria.

A próxima seção detalha o que são AVAs, descrevendo suas principais características. Cita e descreve o Moodle, um dos mais importantes AVAs citados na literatura. E, por fim, detalha um AVA baseado em agentes de apoio à PBL no qual o trabalho proposto foi integrado.

2.5 AMBIENTES VIRTUAIS DE APRENDIZAGEM

Os AVAs são ambientes de EaD que utilizam a Internet como meio de comunicação com os usuários. Ele disponibiliza vários recursos que melhoram a comunicação dos tutores, ou professores, com os seus alunos. Alguns desses recursos são fóruns, *chats*, murais, listas de discussão, entre outros, que ajudam na elaboração de atividades didáticas. Esses recursos apresentam informações de maneira organizada, realizam troca de mensagens, permitem a inserção e acesso a conteúdos educacionais e o gerenciamento de informações, além de possibilitar o gerenciamento de tarefas pedagógicas e de atividades administrativas.

Existe, atualmente, uma grande quantidade de AVAs, alguns gratuitos, como MOODLE, e outros pagos, como Blackboard/WebCT (KUMAR *et al.*, 2011). Segundo

Kumar *et al* (2011), o MOODLE é considerado um dos melhores AVAs *Open Source*. A subseção seguinte detalha o MOODLE.

2.5.1 MODULAR *OBJECT-ORIENTED DYNAMIC LEARNING ENVIRONMENT* (MOODLE)

O MOODLE é um *Framework* formado por ferramentas de comunicação, interação, inserção de documentos e gerenciamento de informações (KUMAR *et al.*, 2011). Disponibiliza conteúdos facilmente, permitindo a colaboração e a comunicação entre o professor e o aluno. Os cursos regulares, geralmente, são acessados quando os usuários se cadastram ou são previamente cadastrados e têm acesso ao conteúdo do curso apenas com autorização através de *login* e senha.

Durante o acesso, o ambiente rastreia e monitora, automaticamente, as ações do aluno no curso, registrando todas suas interações com o ambiente e apresentando estas informações de forma que o professor possa acompanhar a frequência, a participação e o progresso dos alunos (KUMAR *et al.*, 2011).

No entanto, o MOODLE registra relatórios participativos somente contendo informações como número de acessos, data e hora do último acesso, e frequência de acesso do aluno durante um determinado período (KUMAR *et al.*, 2011). Ou seja, só é registrado de quem informou o *login* e a senha de acesso, mas não garante a identidade real de quem está acessando e participando durante todo o processo de ensino e aprendizagem do aluno (FRANÇA e SOARES, 2011). Além disso, não possui mecanismo de adquirir informações afetivas do aluno durante toda participação do aluno no curso do MOODLE.

O ambiente do MOODLE reúne várias ferramentas para auxiliar no ensino e aprendizado a distância. Ele possui tanto ferramentas síncronas, nas quais usuários estão conectados no ambiente ao mesmo tempo, como uma conversa em um *chat* ou videoconferência; como também ferramentas assíncronas, onde usuários acessam a mesma ferramenta em tempo diferente, como listas de discussão e fórum. Possui também ferramentas avaliativas em que os alunos podem realizar exercícios, simulados com questões de múltipla escolha, discursivas e associativas.

Em síntese, o MOODLE é formado por um conjunto sofisticado de ferramentas síncronas e assíncronas, além de possuir alguns recursos de monitoramento das atividades dos

seus usuários. No entanto, esses recursos ainda são insuficientes, pois os docentes continuam tendo dificuldades em processos colaborativos e no acompanhamento dos alunos durante toda realização das atividades.

A implementação do MOODLE foi feita na linguagem PHP (*Hypertext Preprocessor*) e permite criar extensões, o que é primordial para a construção de módulos que complementam as funcionalidades já existentes assim como novas funcionalidades. A construção da base de dados é feita com o MYSQL, que é um verdadeiro servidor de banco de dados multiusuário e multitarefa. Ele apresenta uma arquitetura cliente-servidor multiplataforma com suporte a uma série de sistemas operacionais (KUMAR *et al.*, 2011).

A arquitetura padrão do MOODLE é a cliente-servidor, onde nos computadores Clientes rodam os programas que requisitam serviço de um ou mais computadores Servidores e estes respondem as solicitações (FRANÇA e SOARES, 2011). A Figura 17 apresenta um servidor Web, chamado Servidor EAD, onde está disponível o AVA (páginas em PHP) com o Sistema de Gerenciamento de Banco de Dados (SGBD), que é o MYSQL, que armazena todas as informações obtidas durante os cursos envolvidos. O Servidor EAD pode ser utilizado pelos administradores, coordenadores, docentes, tutores e alunos, que executam o AVA em computadores Clientes, trocando mensagens via Internet (KUMAR *et al.*, 2011).

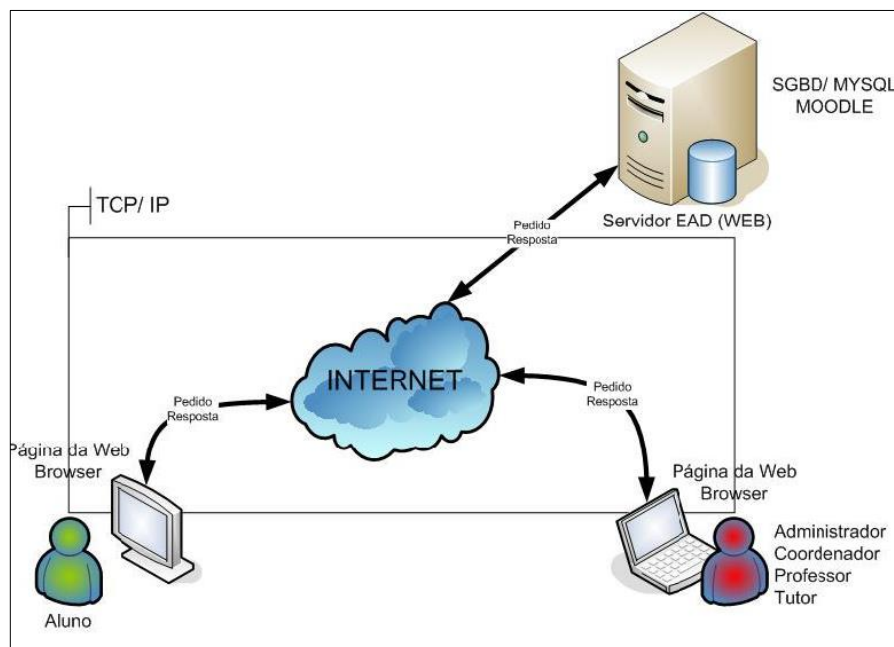


Figura 17 - Arquitetura padrão do MOODLE (SILVA, 2008).

2.5.2 AMBIENTE VIRTUAL DE APRENDIZAGEM BASEADO EM AGENTES DE APOIO À PBL

No trabalho de Fontes *et al.* (2012), foi desenvolvido um AVA composto por um agente pedagógico animado e outros quatro agentes de software para auxiliar na aplicação correta da PBL (*Problem-Based Learning - PBL*), com o objetivo de interagir afetivamente com o estudante, de modo a facilitar a relação ensino-aprendizagem.

Segundo Fontes *et al.* (2012), a PBL é um método no qual os estudantes aprendem através da resolução de um problema que, em geral, não possui uma solução trivial e uma única solução correta. Os estudantes trabalham em pequenos grupos colaborativos para identificar o que eles necessitam aprender para resolução do problema, e o professor atua como facilitador do processo de aprendizagem ao invés de apenas transmitir conhecimentos.

De acordo com Fontes *et al.* (2012), os agentes pedagógicos oferecem instrução personalizada, aumentam a motivação dos estudantes e agem pedagogicamente, por conta própria ou com o auxílio do facilitador. Já os ambientes virtuais de aprendizagem (AVAs) agregam valor ao processo educativo, gerando novas possibilidades de educação. Logo, a utilização da junção de agentes pedagógicos com AVAs se torna uma abordagem promissora para o aprendizado de um usuário que utiliza um computador.

Contudo, a implantação de um método de ensino com base na PBL nos AVAs é desafiador, pois o facilitador (por exemplo, o professor) nem sempre pode detectar possíveis problemas na colaboração, nem possui todas as informações necessárias para aplicar as técnicas de aprendizagem deste método.

Portanto, Fontes *et al.* (2012) elaborou uma arquitetura baseada em agentes para aperfeiçoar a aplicação da PBL nos seguintes aspectos: detecção de estudantes passivos, detecção de conversações fora do contexto, formação de grupos e recomendação de Objetos de Aprendizagem (OAs) sensível ao contexto do estudante. Esses agentes trabalham colaborativamente com o objetivo de detectar e corrigir situações indesejadas, melhorando o processo de aprendizagem. A arquitetura do AVA baseado em agentes de apoio à PBL está esquematizada na Figura 18. O ambiente utiliza a arquitetura cliente-servidor padrão do MOODLE (FRANÇA e SOARES, 2011). No lado cliente, são executadas as páginas em PHP, o Agente Pedagógico Animado (AgPA) e o sistemas de reconhecimento facial e de expressões faciais (RedFace) proposto neste trabalho. Todos solicitam serviços ao servidor da aplicação, através da Internet. No lado servidor localiza-se o SGBD, onde são armazenados

todos os dados referentes aos cursos, aos alunos cadastrados (inclusive suas imagens faciais), além dos quatro tipos de agentes: Agente Detector de Problemas (AgDP), Agente Formador de Grupos (AgFG), Agente Recomendador (AgR) e Agente DF (*Directory Facilitator*) (FONTES *et al.*, 2012).

O meio padrão de acesso ao MOODLE (FRANÇA e SOARES, 2011) se dá através de *login* e senha, o que não garante a veracidade da identidade do aluno. Como pode ser visto na Figura 18, o aluno deve, inicialmente, se autenticar no MOODLE (KUMAR *et al.*, 2011), utilizando o módulo de reconhecimento facial do sistema RedFace, e, depois de autenticado, acessa algum dos cursos nos quais esteja matriculado. O módulo de reconhecimento facial identifica o aluno, via *webcam*, tanto no acesso como durante a utilização do AVA, capturando, em tempos predefinidos, a imagem facial do aluno. Este módulo de reconhecimento facial foi implementado usando a biblioteca OpenCV. Desta forma, o RedFace registra a identidade de quem está realizando as atividades pedagógicas e avaliações no processo da PBL, reforçando assim a garantia da participação do aluno.

O RedfFace possuem também o módulo de reconhecimento de expressões faciais. O módulo de reconhecimento de expressões faciais reconhece as emoções dos alunos, utilizando as expressões faciais capturadas por uma câmera (Kinect), em tempos predefinidos. O RedFace registra na base de dados a emoção do aluno, a qual vai ser utilizada pelo Agente Pedagógico Animado (AgPA) para acompanhar os estudantes durante o processo de aplicação da PBL, além de manter os estudantes sempre motivados (FONTES *et al.*, 2012).

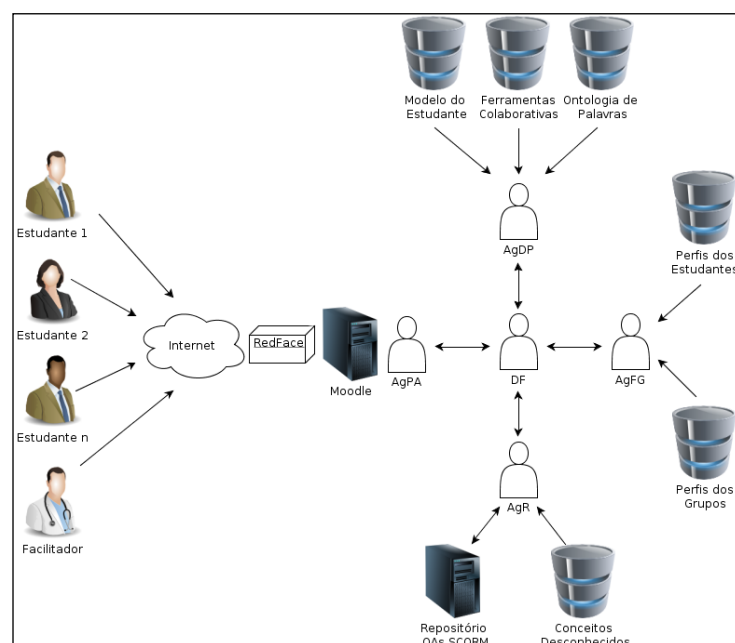


Figura 18 – Arquitetura do AVA baseado em agentes de apoio à PBL (FONTES *et al.*, 2012).

Os agentes desenvolvidos por Fontes *et al* (2012) presentes na arquitetura (Figura 18) foram implementados utilizando o JADE (*Java Agent Development Framework*) (JADE, 2012). As principais características desses agentes são descritas a seguir (FONTES *et al.*, 2012).

a. Agente Pedagógico Animado – AgPA

O AgPA consiste em um avatar 3D que provê uma maior interatividade ao ambiente, oferecendo aos estudantes a ideia de um tutor sempre presente durante todo o processo que define a PBL. Ele se comunica com os outros agentes, de forma colaborativa, e atua de acordo com o que for constatado no ambiente. Por exemplo, uma vez que o AgDP tenha detectado uma mudança na emoção do aluno, o AgPA será acionado para tentar motivar este estudante a interagir mais com o ambiente.

Como forma de motivar o aluno, o AgPA expressará emoções similares às dos seres humanos, conforme ilustrado na Figura 19. Foram modeladas quatro animações para expressar as emoções do AgPA, conforme ilustrado na Figura 19: felicidade, que remete, por exemplo, momentos em que o estudante esteja interagindo com o ambiente; tristeza, quando algum problema de colaboração for detectado como, por exemplo, a detecção de estudantes passivos; expectativa, durante os questionamentos do AgPA para o estudante; e dúvida, quando o estudante permanecer muito tempo em uma página.

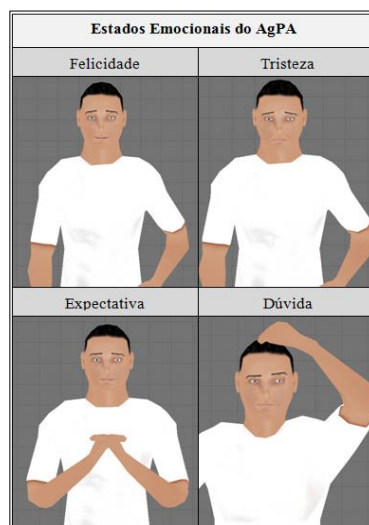


Figura 19 - Animações referentes aos estados emocionais do AgPA (Fontes *et al.*, 2012).

b. Agente Detector de Problemas - AgDP

O AgDP possui dois comportamentos: detecção de estudantes passivos e detecção de conversações fora do contexto do problema. Para que o AgDP consiga detectar estudantes passivos, é necessário que ele atualize os perfis dos estudantes conforme o uso das ferramentas colaborativas disponíveis no ambiente. Em seguida, todas as informações relativas ao estudante passivo são enviadas para o AgPA. Já para que o AgDP detecte as conversações fora de contexto do problema, utiliza-se uma abordagem baseada no uso das ferramentas colaborativas disponíveis no ambiente e uma ontologia do problema.

c. Agente Formador de Grupos - AgFG

O AgFG é o responsável pela formação automática dos grupos. Faz a geração e alocação de candidatos que estão aptos a participar de determinado grupo. Esse processo é feito analisando os perfis dos estudantes e os perfis dos grupos. O perfil dos estudantes é composto por habilidades, conhecimentos e deficiências, onde cada um possui um nível, que pode ser baixo, médio ou alto, podendo um estudante ter uma ou mais habilidades, deficiências e conhecimentos.

d. Agente Recomendador - AgR

O AgR tem o intuito de detectar OAs adequados ao contexto do estudante. O AgR encontra, inicialmente, o OA que seria mais adequado de acordo com a recomendação baseada em conteúdo. O mecanismo de recomendação baseada em conteúdo considera as informações de horário preferido de estudo e área de interesse do estudante, as quais estão contidas na ontologia de contexto estático.

e. *Directory Facilitator* - DF

Tem como objetivo ser o mediador da comunicação entre os agentes, o qual é provido pela própria plataforma JADE, conforme exigência da especificação FIPA (*Foundation for Intelligent Physical Agents*) (JADE, 2012; FIPA, 2011).

A Seção 2.6 aborda o conceito de emoções, resumindo suas definições e características. Mostra as diferentes maneiras de comunicar as emoções. Detalha a face como medida de avaliação da expressão emocional. E, por fim, são descritas as teorias existentes que diferenciam as emoções.

2.6 EMOÇÕES

Emoções são reações universais presentes em todos os seres humanos e que independem de fatores culturais; assim, diferentemente da linguagem, que se desenvolve de acordo com fatores culturais, o reconhecimento e as expressões emocionais são os mesmos nas diferentes culturas (KELTNER e EKMAN, 2002) citado por (KWANG-EUN e KWEE-BO, 2010; ABDAT *et al*, 2011). No entanto, há diferenças na forma e na intensidade com que as pessoas exprimem essas emoções (BESCHE-RICHARD e BUNGENER, 2008).

As emoções são consideradas tipos de estado afetivo que podem ser caracterizadas através de suas manifestações, tais como: motivação, raiva, tristeza, alegria, medo, vergonha e desespero (ver Figura 20). As emoções são manifestadas por: (i) reações expressivas, como sorrisos, cenho franzido e dentes trincados; (ii) por reações fisiológicas, como aumento dos batimentos cardíacos, produção de lágrimas, calores e vermelhidão no rosto; (iii) por comportamentos instrumentais, como correr e juntar as mãos; (iv) por comportamentos instrumentais situacionais, como digitar com força desmesurada uma tecla e gritar um impropério qualquer; (v) por cognições, como pensamento de injustiça para si ou para outros e sensação de impotência frente a problemas; e (vi) por sentimentos que integram os fenômenos fisiológicos e cognitivos, como a tristeza (sensação de um aperto no peito, lágrimas nos olhos e a lembrança do evento que gerou tais sensações) (BERCHT, 2001).

As expressões emocionais também se caracterizam por serem automáticas e involuntárias. Por isso que não é fácil reproduzir artificialmente uma expressão emocional genuína. Por exemplo, para expressar o sorriso autêntico, em oposição aos sorrisos falsos, envolve a contração de um músculo dos olhos e da região lateral do *orbicularis oculi*, também conhecido por músculo de Duchene (CARLSON, 2002). Conforme Duchene, o primeiro (músculo zigomático principal) obedece à vontade, mas o segundo (*orbicularis oculi*) é apenas ativado pelas doces emoções da alma. “A alegria simulada, a gargalhada falsa, não pode provocar a contração deste último músculo” (CARLSON, 2002).

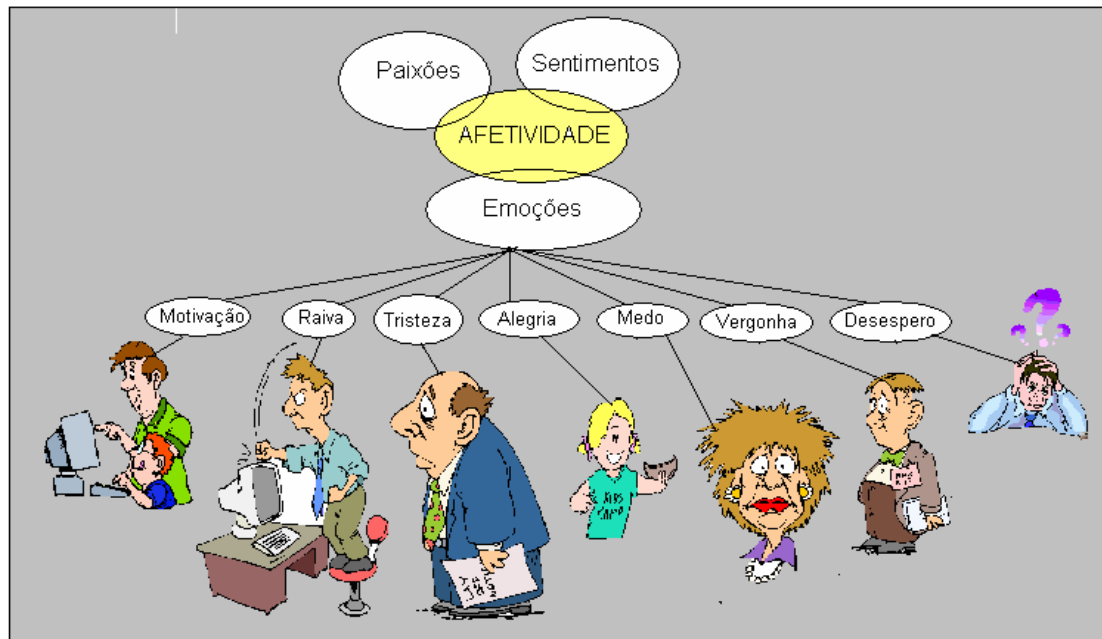


Figura 20 – Afetividade e manifestações do estado afetivo (QUARTO, 2006).

Ainda não existe um consenso quanto à definição de emoções, estas podem ocupar uma lista de até vinte tipos (PICARD, 1997) citados por (SUN e LU, 2010). Segundo a visão de alguns pesquisadores da educação, psicopedagogos e de psicólogos, muitos estudos têm sido desenvolvidos para definir emoção. Para esses pesquisadores, emoções são: (i) fenômenos complexos multifacetados (BERCHT e VICCARI, 2004); (ii) partes necessárias da vida que afeta em como nos sentimos, como nos comportamos e pensamos; (iii) um tipo de estado afetivo (JAQUES e VICCARI, 2004); (iv) reações a eventos, agentes ou objetos, com suas naturezas particulares, sendo determinadas pelo modo na qual a situação extraída é construída (JAQUES e VICCARI, 2004); (v) sistemas de atitudes que correspondem cada uma a uma determinada espécie de situação (MANRIQUE, 2004).

As emoções são processos complexos e são comunicadas de diferentes maneiras, pela expressão facial, pelo olhar, pela postura, pelos gestos, pela voz e pelas palavras, sendo as primeiras do tipo de comunicação não verbal e a última, de comunicação verbal (KELTNER e EKMAN, 2002) citado por (KWANG-EUN e KWEE-BO, 2010; ABDAT *et al*, 2011). A linguagem verbal permite-nos comunicar a respeito das emoções por meio de palavras. Já o comportamento não verbal, como a postura corporal ou a expressão facial, possibilita-nos comunicar e interagir uns com os outros através de outras manifestações do corpo. No entanto, as medidas de avaliação utilizadas para analisar tão diversificadas respostas emocionais são diversas, pois não é trivial compreender o complexo processo emocional (ROSERBERG e EKMAN, 2002) citado por (ABDAT *et al*, 2011).

A medida de avaliação da expressão emocional mais usada em investigações é pela face, e os diferentes instrumentos de avaliação são: o reconhecimento emocional por meio do julgamento do observador, os sistemas de codificação e os sistemas eletrofisiológicos (ROSERBERG e EKMAN, 2002) citado por (ABDAT *et al*, 2011). As investigações sobre a universalidade das emoções e o desenvolvimento de métodos para medir objetivamente as expressões faciais contribuíram para o nosso conhecimento sobre a emoção (EKMAN, 1993) citado por (ABDAT *et al*, 2011).

Assim, as expressões faciais têm uma importância vital na comunicação de emoções, tanto ao nível da comunicação verbal como da comunicação não verbal (LEDERMAN *et al.*, 2007). Pode ser vista como um palco onde os sinais emocionais atuam. Como, por exemplo, o sorriso, que além de transmitir a emoção de felicidade, também é sinal de cooperação; o franzir das sobrancelhas pode transmitir ira, e, portanto, um conflito interpessoal. Ela também permite transmitir o que por vezes não queremos fazer por palavras, revelando emoções e pensamentos genuínos, que por vezes, a linguagem tenta enganar ou esconder (ESCADA, 2004).

O Ekman (1999) citado por (ABDAT *et al*, 2011) propôs uma teoria de emoções bastante difundida, conhecida como Modelo de Emoções Básicas, que recebe este nome por que as emoções têm as mesmas manifestações corporais em diferentes culturas. Segundo Ekman (1999) citado por (ABDAT *et al*, 2011), as respostas emocionais primárias da face não variam interculturalmente porque os seres humanos possuem o mesmo conjunto muscular no rosto. Segundo Ekman e Friesen (2002) citado por (OLIVEIRA e JAQUES, 2008), os sujeitos nascem com a capacidade de reconhecer apenas um conjunto de seis expressões faciais de emoções: raiva, medo, tristeza, desgosto, alegria e surpresa, representadas na Figura 21. As demais emoções e suas expressões faciais são adquiridas através da aprendizagem e da interação com o meio.



Figura 21 – Seis expressões faciais emocionais básicas (OLIVEIRA e JAQUES, 2008).

Esse modelo de emoções básicas, proposto por Ekman (1999) citado por (ABDAT *et al.*, 2011), apresenta uma grande quantidade de trabalhos relacionados, principalmente no que se refere ao reconhecimento computacional de emoções através de expressões faciais. Neste trabalho é utilizado o sistema de codificação de classificação facial (FACS) para inferência de emoções (EKMAN *et al.*, 2002) citado por (ABDAT *et al.*, 2011), que é baseado psicologicamente na Teoria de Emoções Básicas de Ekman (1999) citado por (ABDAT *et al.*, 2011), que será abordada na próxima seção.

A próxima seção descreve as principais características do FACS para inferência de emoções. Detalha também o que são Unidades de Ação (UAs) e mostra o mapeamento dos UAs para a classificação das emoções.

2.7 MODELO PSICOLÓGICO DE CLASSIFICAÇÃO FACIAL – FACS

Segundo Ekman e Friesen (2002) citado por (OLIVEIRA e JAQUES, 2008), as expressões faciais estão classificadas em seis grandes classes de emoções básicas: alegria, tristeza, surpresa, desgosto, raiva e medo. Cada emoção básica é definida levando em conta a expressão facial que caracteriza essa emoção univocamente. No entanto, esta representação tornou-se bastante limitativa. Pois, é comum para o ser humano a manifestação de várias emoções simultâneas, o que resulta numa mescla entre várias expressões emocionais básicas.

Para solucionar este problema, Ekman e Friesen (2002) citado por (OLIVEIRA e JAQUES, 2008) desenvolveram o sistema FACS (EKMAN *et al.*, 2002) citado por (OLIVEIRA e JAQUES, 2008), o qual permite a especificação precisa da morfologia e dinâmica de movimentos faciais. Ele é baseado em conhecimentos de anatomia, vídeos e fotografias ilustrativos que mostram como a contração de cada músculo facial desempenha um papel na alteração de expressão de um dado indivíduo. Este sistema tornou-se uma das maiores referências na área de reconhecimento de expressões faciais.

Foram determinados, no sistema FACS, os músculos que podem ser controlados voluntariamente e de forma independente, e como eles alteram a aparência da face. Tais mudanças foram mapeadas, dando origem às UAs que compõem as expressões faciais. Cada UA apresenta o conjunto de músculos que devem ser relaxados ou contraídos para resultar em um pequeno movimento em uma parte do rosto. Nesse sistema foram definidas 44 UAs

(variando de 1 a 44), que combinadas geram um grande número de expressões faciais, podendo expressar mais de 7000 expressões faciais diferentes, cobrindo todas as expressões faciais possíveis (Ekman *et al.*, 2002) citado por (OLIVEIRA e JAQUES, 2008).

A Figura 22 mostra o mapeamento da UA 43 com seu respectivo conjunto muscular envolvido. A UA 43 descreve o fechamento dos olhos, que resulta do relaxamento do músculo *Levator Palpebrae Superioris* (FACS, 2007). O Apêndice A mostra todas as UAs do sistema FACS, que foi utilizado na implementação do sistema de reconhecimento de expressões faciais proposto.


AU	Descrição	Músculo Facial	Exemplo de imagem
AU 43	Olhos Fechados	<i>Relaxation of Levator palpebrae superioris; Orbicularis oculi pars palpebralis</i>	

Figura 22 – UA 43 do Sistema FACS. Fonte própria.

Estes estudos sobre FACS foram mais explorados na década de 80 para a criação de outro modelo, EMFACS (EMFACS, 2007), que mapeia e seleciona as UAs que ocorrem na manifestação de emoções. Por exemplo, uma pessoa que está sentindo a emoção de raiva pode ser reconhecido pela ocorrência simultânea das UAs 4, 5, 7, 15 e 24 presentes na sua face (EKMAN *et al.*, 2002) citado por (OLIVEIRA e JAQUES, 2008). A Tabela 4 mostra alguns exemplos de como UAs podem ser combinadas para formar expressões de emoção.

Tabela 4 – Exemplo de combinações de UAs formando algumas expressões de emoções.

Emoção	Combinação de UAs
Alegria	6 + 12 + 25
Raiva	4 + 5 + 7 + 15 + 24
Tristeza	1 + 4 + 7 + 15 + 17
Desgosto	9 + 10

É importante estudar os modelos de codificação de expressões faciais, pois são relevantes para a etapa de detecção de características faciais. Consequentemente, estes

modelos podem ser usados como base para realizar as classificações de emoções presentes nas imagens, utilizando algoritmos de aprendizado de máquina da inteligência artificial.

A próxima seção apresenta os conceitos de afetividade e sua importância nos sistemas educativos. Define o que é afetividade e suas características mais importantes para dar aos computadores as habilidades emocionais. Especifica os principais recursos que fazem os sistemas educativos serem afetivos. Mostra os três principais ramos de pesquisas abordados nos trabalhos sobre a afetividade citados na literatura. Por fim, detalha as principais formas de mapeamento de reconhecimento das emoções de uma pessoa e analisa o uso de Computação Afetiva em AVAs.

2.8 A COMPUTAÇÃO AFETIVA

É cada vez mais frequente ver pessoas passando mais tempo interagindo com computadores do que com outros seres humanos. As novidades tecnológicas, por sua vez, estão cada vez mais presentes na vida das pessoas. Os rumos da tecnologia parecem ir ao encontro da melhoria constante da interação humano-máquina, desenvolvendo-se computadores sensitivos e inteligentes, que chegam ao ponto de serem confundidos com seres autônomos. Segundo Picard (1997) citado por (SUN e LU, 2010), o afeto é parte natural e social da comunicação humana. As pessoas naturalmente o usam quando vão interagir entre elas e quando interagem com computadores. Se os computadores fornecerem qualidades afetivas com o ambiente ou com o usuário, a relação homem-máquina-ambiente tornar-se-á um processo mais natural.

Neste contexto, Picard (1997) citado por (SUN e LU, 2010) usa o termo Computação Afetiva para dar aos computadores as habilidades emocionais. Segundo a autora, para que os computadores adaptem-se e interajam naturalmente com seres humanos, é preciso dotá-los da habilidade de reconhecer e expressar emoções (Picard, 1997) citado por (SUN e LU, 2010).

A Computação Afetiva é um campo da Inteligência Artificial com a finalidade de imergir emoções na máquina, podendo ser por inferência de emoções do usuário, expressão de emoções ou simulação de emoções em máquina (JAQUES e VICCARI, 2005).

Para compreender melhor a Computação Afetiva, é importante entender o conceito de afetividade. Segundo Longhi (2007), afetividade é um termo utilizado para identificar um conjunto de fenômenos psíquicos que inclui as primeiras manifestações afetivas basicamente

orgânicas, bem como suas manifestações relacionadas ao social. Logo, a afetividade está envolvida com conceitos como: emoção, humor, motivação, paixão, atenção, personalidade, temperamento, entre outros. Porém, a emoção é o estado afetivo mais aplicado nos sistemas educativos, através da utilização das seis emoções básicas (EKMAN e FRIESEN, 2002) citado por (OLIVEIRA e JAQUES, 2008), para o estudo do reconhecimento de emoções humanas e da expressão de emoções por máquinas.

Por conseguinte, os sistemas educativos que pretendem ser afetivos aos seus usuários devem ser capazes de: (1) *Reconhecer emoções*, (2) *Expressar emoções* e (3) *Possuir emoções* (PICARD, 1997) citado por (SUN e LU, 2010). A Figura 23 ilustra as relações da Computação Afetiva com o homem.

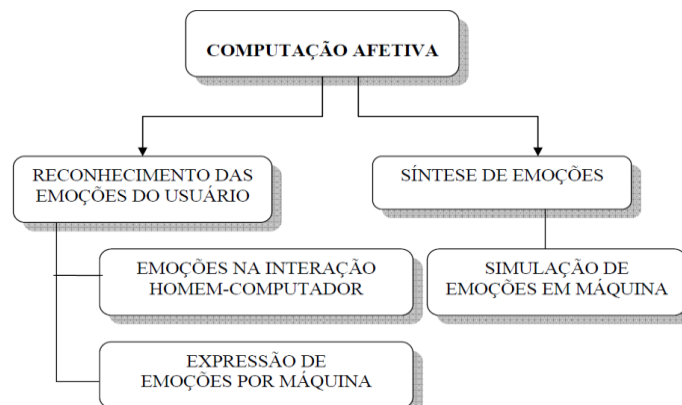


Figura 23 – Ramos da computação afetiva (JAQUES e VICCARI, 2004).

De acordo com as relações da Computação Afetiva com o homem, observadas na Figura 23, os trabalhos sobre afetividade em sistemas educacionais inteligentes podem ser divididos em três importantes pesquisas (JAQUES e VICCARI, 2004):

- **Reconhecimento de emoções de usuários:** o sistema computacional afetivo deve reconhecer as emoções do usuário para respondê-lo adequadamente, a fim de proporcionar um melhor aproveitamento nas interações. Alguns dos mecanismos para reconhecimento de estados afetivos de estudantes são por meio: de gestos/expressões faciais, linguagem (texto diálogo), condutividade da pele, eletromiograma, respiração, ritmo cardíaco, temperatura, acústico, comportamento observável, entre outros (JAQUES e VICCARI, 2004);
- **Expressões de emoções:** objetiva obter motivação em relação ao usuário. Os sistemas computacionais têm procurado enriquecer suas interfaces de forma a expressar emoções através de expressões faciais e gestos de forma mais próxima

do comportamento humano (JAQUES e VICCARI, 2004). Em educação, por exemplo, alguns trabalhos usam agentes representados por personagens animados para apresentar o conteúdo pedagógico para o estudante, fazendo-lhe demonstrações e motivando-o (JAQUES e VICCARI, 2004). Estes agentes têm um papel pedagógico de facilitar e melhorar o aprendizado do aluno durante a interação com o mesmo;

- **Síntese de emoções:** investiga a simulação de emoções em máquinas a fim de descobrir mais sobre emoções humanas. Segundo Picard (2004) citado por (DEWI *et al.*, 2011), agentes computacionais com síntese de emoções terão condições de analisar as situações e eventos do mundo e responder ao mesmo com características próximas ao do comportamento humano.

A maioria das pesquisas sobre emoção no ramo da Informática na Educação está sendo focada principalmente nos campos de reconhecimento e expressões de emoções dos usuários. Sistemas tutores inteligentes, agentes pedagógicos e hipermídia adaptativa têm sido algumas implementações propostas para a educação. Estes sistemas devem ser capazes de reconhecer as emoções dos alunos, verificando se os mesmos estão aptos ou motivados a determinada atividade para, enfim, realizar ações de modo a encorajá-los a estudar (JAQUES e VICCARI, 2004).

2.8.1 MAPEAMENTO DE ASPECTOS AFETIVOS

Quando o interesse é o reconhecimento de estados afetivos, é preciso dispor de interfaces adequadas para captar os sinais emitidos pelo usuário para, em seguida, processá-los e, posteriormente, atribuí-les determinados estados afetivos. Os fenômenos afetivos podem ser deduzidos através de dispositivos de entrada de dados (por exemplo, câmeras que analisam as expressões faciais), analisados através da aplicação de modelos afetivos (PICARD, 2004) citado por (DEWI *et al.*, 2011).

O uso de tecnologia para reconhecer afetividade com mais robustez ainda carece de muita pesquisa e do desenvolvimento de interfaces, que associados às técnicas de reconhecimento de padrões, processamento de sinais (som e imagens) e algoritmos computacionais poderão avaliar e responder os estados afetivos dos aprendizes em tempo real (PICARD, 2004) citado por (DEWI *et al.*, 2011). Logo, a inferência computacional de

fenômenos afetivos, principalmente em ambientes educacionais, é um dos desafios da Computação Afetiva. Há necessidade de apoio de novos meios tecnológicos que ajudem a perceber, identificar, medir, expressar e simular a afetividade em situações de aprendizagem.

Entre as formas de mapeamento de reconhecimento das emoções de uma pessoa, quatro delas se destacam (PICARD, 2004) citado por (DEWI et al., 2011), como mostra o mapeamento dos aspectos afetivos ilustrados na Figura 24.



Figura 24 – Mapeamentos dos aspectos afetivos. Fonte própria.

O mapeamento feito pela voz se faz pela variação sonora da voz dos indivíduos. Pelos comportamentos observáveis podem ser mapeados pelas ações do usuário na interface do software. As expressões faciais são mapeadas observando as mudanças das deformações da face resultantes das emoções emitidas pelo indivíduo. Já o mapeamento feito por sinais fisiológicos pode ser feito pelos batimentos cardíacos, eletromiograma, tensão muscular, condutividade de pele e respiração.

De acordo com a Figura 24, é verificado que a voz, os sinais fisiológicos e as expressões faciais demandam tecnologia específica para o reconhecimento da afetividade, tais como: sensores associados às técnicas de reconhecimento de padrões, processamento de sinais e algoritmos computacionais. Segundo Picard (2004) citado por (DEWI et al., 2011), estas tecnologias permitem avaliar e responder os estados afetivos dos alunos em tempo real. As emoções podem também ser reconhecidas a partir do comportamento observável do aluno, ou seja, observando a interação do mesmo com a interface do ambiente educacional, onde é analisado, por exemplo, o tempo de realização de um exercício, sucesso ou falha na realização de tarefas, pedido de ajuda ou negação de uma ajuda, entre outros.

2.8.2 USO DE COMPUTAÇÃO AFETIVA EM AMBIENTES VIRTUAIS DE APRENDIZAGEM

O desenvolvimento da computação afetiva representa um aprimoramento na interação humano-computador. As contribuições potenciais das pesquisas nessa área são significantes tanto teórica quanto empiricamente para um melhor entendimento da emoção e da cognição, para o aperfeiçoamento dos computadores, para avanços na possibilidade de comunicação homem-máquina, e para a melhor compreensão da influência que a tecnologia exerce no próprio desenvolvimento humano (PICARD, 1998) citado por (ZHOU *et al.*, 2010). Além da intenção de colocar emoções nos computadores, a computação afetiva também inclui muitas outras funções, dentre elas a de dar aos computadores a capacidade de responder inteligentemente à emoção humana.

A seguir, são apresentadas algumas aplicações e suas características que utilizam técnicas da área da computação afetiva em ambientes de EaD.

2.8.2.1 AGENTE PEDAGÓGICO PAT

O agente PAT (JAQUES e VICCARI, 2004) tem como objetivo motivar o estudante e promover emoções positivas nele. De acordo com Jaques e Viccari (2004), para o PAT fornecer suporte emocional ao aluno, o agente deve conhecer as emoções do aluno. O agente reconhece as seguintes emoções do aluno: alegria, tristeza, satisfação, frustração, raiva e vergonha, a partir do comportamento observável na interface do sistema educacional. Já no agente proposto por Fontes *et al* (2012) reconhece as emoções de felicidade, tristeza, dúvida e expectativa, utilizando o comportamento observável na interface do AVA e pelas expressões faciais reconhecidas pelo módulo de reconhecimento de expressões faciais.

O agente PAT reconhece emoções dos estudantes fundamentadas no modelo psicológico OCC (*Ortony, Clore e Collins*) (JAQUES e VICCARI, 2004), o qual consiste de um modelo psicológico cuja teoria é baseada na abordagem cognitiva da emoção (JAQUES e VICCARI, 2004). As emoções são vistas nesse modelo a partir da avaliação cognitiva que um indivíduo faz através de três aspectos do mundo real: eventos, agentes e objetos.

- Os eventos são formas ou maneiras de como as pessoas percebem as coisas;

- Os agentes representam as pessoas, animais, objetos abstratos;
- Os objetos representam as coisas visíveis, palpáveis.

A arquitetura do agente PAT é composta por dois módulos principais: módulo Mente e o módulo Corpo. O módulo Corpo é responsável por capturar as ações do aluno na interface do sistema, realizar a comunicação com os outros agentes e mostrar os comportamentos animados e mensagens escolhidos pelo módulo Mente. Este módulo foi implementado na linguagem Java. O módulo Mente é responsável por inferir os estados afetivos do aluno a partir de seu comportamento observável e escolher as táticas pedagógicas afetivas de acordo com o modelo afetivo do aluno.

O modelo afetivo do estudante e o módulo Mente do agente PAT são desenvolvidos com base em crença, intenção e desejo (*Belief Desire Intention* - BDI) (JAQUES e VICCARI, 2004). A abordagem BDI descreve o processamento interno do agente através de estados mentais (crenças, desejos e intenções) e define a arquitetura de controle que racionalmente seleciona o curso de ações do agente (GIRAFFA 1999) citado por (JAQUES e VICCARI, 2004). Para a modelagem e implementação em BDI da mente, foi utilizada a ferramenta X-BDI (MÓRA *et al.*, 1998) citado por (JAQUES e VICCARI, 2004). O personagem animado foi desenvolvido em *Microsoft Agent* (BOCCA *et al*, 2003) citado por (JAQUES e VICCARI, 2004). Esse personagem foi desenhado em estudo feito com psicólogos e pedagogos. Para voz, fala e a linguagem foi usada a *Microsoft Speech API* (BOCCA *et al*, 2003) citado por (JAQUES e VICCARI, 2004).

Toda essa arquitetura projetada para o PAT permite uma maior aceitação do ambiente por parte do aluno, deixando-o mais realista e menos robótico (JAQUES e VICCARI, 2004). Na Figura 25, podem ser observadas as características físicas do agente PAT.

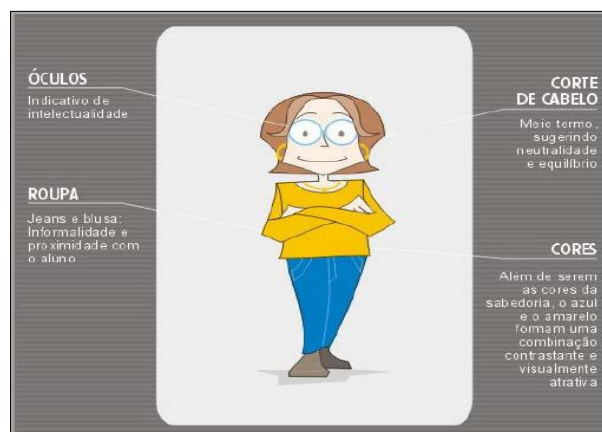


Figura 25 – Agente PAT (JAQUES e VICCARI, 2004).

2.8.2.2 AGENTE PEDAGÓGICO DÓRIS

O agente pedagógico animado *Dóris 3D* é resultado do projeto de um grupo de pesquisadores do departamento de Informática da Universidade de Santa Cruz do Sul (FROZZA *et al.*, 2009). Segundo Frozza *et al.* (2009), o agente Dóris foi implantado como um agente pedagógico emocional, com expressões em 3D, visando qualidade nos métodos de aprendizagem por meio de sistemas virtuais, a fim de oferecer maior motivação e entusiasmo nos processos de aprendizagem, gerando autonomia e proporcionando atividades desafiadoras aos estudantes. Foram utilizadas as emoções alegria, tristeza, expectativa, indignação, surpresa, atenção e dúvida.

A arquitetura do agente Dóris, ilustrada na Figura 26, é composta dos módulos: perceptivo, cognitivo e reativo. Esta estrutura também possui uma base de conhecimento interna sobre as táticas de ensino disponíveis no ambiente em relação a um material instrucional.

O módulo perceptivo é o responsável pela extração e armazenamento das informações referentes à interação do estudante com o sistema. É por meio dele que é realizada a tarefa de monitoramento das ações do estudante. O módulo cognitivo é o responsável por realizar as inferências sobre a base de conhecimento, determinando quais as ações que devem ser executadas pelo agente, a partir de suas percepções. Utilizando este módulo é feita a escolha das mensagens que serão emitidas pelo agente na sua interação com o estudante. O módulo reativo é o responsável por executar as ações indicadas pelo módulo cognitivo. Por meio do módulo reativo são apresentadas as mensagens dos agentes aos estudantes.

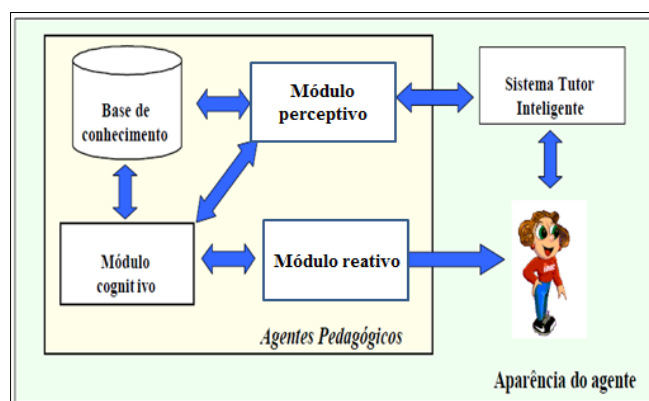


Figura 26 – Arquitetura do agente pedagógico Dóris (FROZZA *et al.*, 2009).

O agente Dóris ainda possui um módulo de tomada de decisão, que é responsável por analisar as características do estudante e relacioná-las com as táticas de ensino a serem utilizadas para a apresentação do material instrucional. Ele tem sua execução através de regras de produção, na qual as condições são as características dos estudantes e as ações são as táticas de ensino. Nessas regras foram incluídas as emoções a serem expressas pelo agente Dóris durante a execução das suas ações de interação com os estudantes.

Para a modelagem e animação do agente Dóris 3D foi utilizada a ferramenta *Blender* 3D, considerando o fato de ser um *software* de código aberto, multiplataforma, entre outros motivos. Na modelagem da forma do agente foi utilizada a técnica de malha poligonal e na animação do agente foi usada a técnica conhecida como *key frame* (FROZZA *et al.*, 2009).

Dos resultados desse trabalho, os autores concluíram a importância das expressões durante a interação entre os indivíduos, que fazem um diferencial na compreensão do assunto abordado, porém, saber qual expressão utilizar, e em qual momento, é um grande desafio.

A próxima seção descreve as características dos sistemas automatizados de reconhecimento de emoções de um indivíduo. Detalha a importância do reconhecimento de expressões faciais na descoberta das emoções de uma pessoa. Verifica os principais elementos que compõem a expressão facial. Discute alguns dos principais fatores que influenciam no reconhecimento das emoções faciais. E, por fim, resume as principais etapas envolvidas no reconhecimento da expressão facial, detalhando algumas técnicas para o reconhecimento da expressão.

2.9 RECONHECIMENTO DE EXPRESSÕES FACIAIS

Como foi visto na Subseção 2.8.1, o homem manifesta a emoção durante todos os momentos da sua vida. São várias as formas de uma pessoa manifestar suas emoções, e mais variados ainda são os métodos capazes de reconhecerem a emoção transmitida. Os principais meios de reconhecimento de emoções são por meio da voz, das ações do usuário na interface com o sistema, das expressões faciais e dos sinais fisiológicos (JAQUES e VICCARI, 2005).

Atualmente, os sistemas automatizados de reconhecimento de emoções de uma pessoa tendem a aproximar-se do reconhecimento de emoções feitos pelo próprio homem e, em alguns casos, chega a exceder. Segundo Sebe *et al* (2005), a média do reconhecimento de

emoções faciais por humanos é de aproximadamente 85% de precisão, enquanto alguns algoritmos computacionais obtêm sucesso entre 74% e 98%. Sebe *et al* (2005) afirma também que existe um equilíbrio entre humanos e máquinas no reconhecimento de emoções pela voz, por volta de 65%. Entretanto, alguns algoritmos alcançaram o nível de quase 80% de precisão. O principal problema existente na identificação de emoção pela voz é a interferência provocada por ruídos que pode existir na captação sonora.

Por conseguinte, dentre os métodos de reconhecimento de emoções, a expressão facial é um dos elementos mais importante e espontâneo para reconhecer o estado emocional humano. O reconhecimento de expressões faciais é uma modalidade que visa aumentar a familiaridade da comunicação com os usuários, fazendo com que usuários interajam com um sistema computacional como se fosse uma interação entre pessoas.

Segundo Fasel e Luetin (2003), uma análise apenas da expressão labial, por exemplo, não tem como concluir se um sorriso refere-se realmente a uma emoção de alegria ou é apenas uma pose, mas fornece artefatos que podem aumentar essa possibilidade. Uma pessoa pode tentar expressar e convencer uma emoção que não sente, mas alguns músculos faciais acionados de determinado modo, somente quando algum tipo verdadeiro de emoção é manifestado, podem desmentir essa tentativa (EKMAN, 1993). Conforme ilustra a Figura 27, uma expressão facial é composta por vários elementos, sendo um destes o sentimento da emoção.

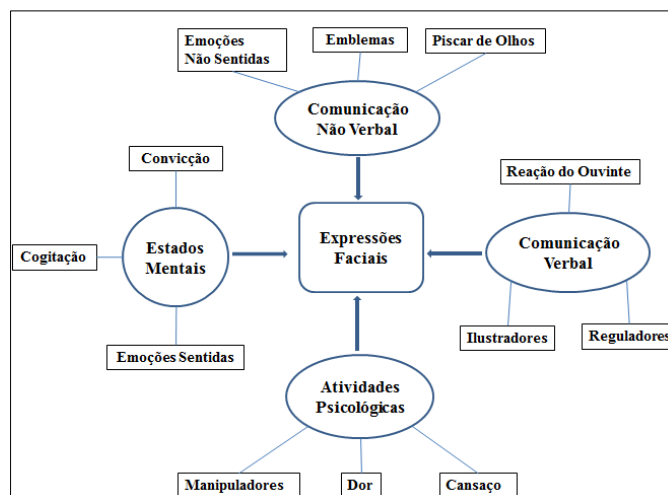


Figura 27 - Elementos envolvidos nas expressões faciais. Adaptado de (FASEL e LUETTIN, 2003).

As pesquisas sobre o potencial expressivo de um rosto humano estão cada vez mais sendo amplamente exploradas. Pois, apesar de ser simples e rápido para um ser humano

reconhecer as expressões faciais, para um sistema computacional envolve uma série de restrições, implicando a adoção de um conjunto de técnicas e algoritmos relativamente complexos para reconhecer as expressões faciais (ROSARIO, 2008).

O reconhecimento automático de expressões faciais deve levar em consideração fatores como a luminosidade do meio ambiente, a posição do utilizador em relação à câmara, as diferenças individuais entre as pessoas (decorrentes da idade, sexo, doença, ou raça, por exemplo) e características do próprio usuário, como a presença de pêlos faciais e óculos. Além disto, os objetivos da sua aplicabilidade determinam outros requisitos, como o desempenho a nível temporal e quantização da imagem, que determina se os dados de entrada consistem em vídeo ou quadros individuais (ROSARIO, 2008).

Para que uma expressão facial seja reconhecida, é preciso seguir uma metodologia coerente. As aplicações dos conhecimentos de VC são as mais adotadas como metodologia para a detecção de emoção de expressões faciais pelo computador. Existem processos computacionais que tornam possível realizar a extração e interpretação de dados de imagens (ROSARIO, 2008). Sistemas automáticos de reconhecimento de expressão facial aderem a um modelo clássico de reconhecimento de padrões comuns na maioria dos sistemas de VC. As principais etapas desse modelo são mostradas na Figura 28, detalhando suas etapas com exemplos de algumas técnicas para o reconhecimento da expressão.

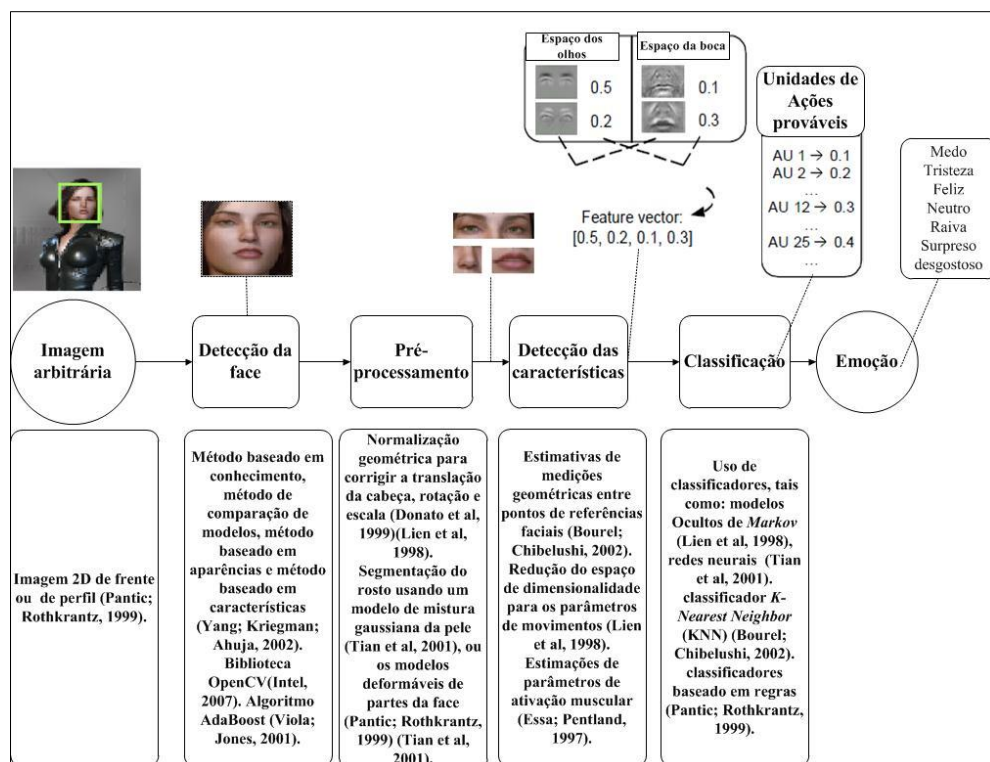


Figura 28 – Metodologia básica para o reconhecimento de padrões. Fonte própria.

A seguir, serão descritas as principais características das etapas desse modelo clássico de reconhecimento de padrões (Figura 28), comuns na maioria dos sistemas de VC.

- **Aquisição da imagem arbitrária:** É o primeiro passo para um sistema de VC. Trata-se do processo de aquisição de uma imagem ou de um conjunto de imagens a partir de sensores de câmeras. A imagem pode ser bidimensional, tridimensional ou uma sequência de imagens.
- **Detecção da face:** A detecção de face consiste na utilização de métodos computacionais que verificam a existência de uma face, em uma determinada imagem digital, de vídeo ou fotografia (KASHEM, 2011). No entanto, pode ser que estejam presentes diversos artefatos exteriores ao rosto. Logo, para que o processamento seja eficaz e eficiente, é necessário eliminar informação desnecessária que bloqueia a identificação da face.
- **Pré-processamento:** A etapa de pré-processamento aplica métodos específicos que facilitam a identificação de um objeto, tais como: correção da iluminação, contraste, distorções, nitidez, destaque de contornos, bordas e destaque de figuras geométricas (PATIL *et al.*, 2010). Esta etapa realiza o processo de normalização que executa as possíveis correções necessárias nas imagens ou quadros para serem segmentadas, para em seguida passarem ao processo de detecção de características. Dentre as principais correções realizadas pela normalização, destacam-se a eliminação de ruído e das possíveis rotações, translações naturais aplicadas pela face (YANG *et al.*, 2002) citado por (PATIL *et al.*, 2010).
- **Detecção das características faciais:** A etapa de extração de características compreende as técnicas necessárias para extrair os pontos de controle de cada uma das características faciais segmentadas. Ou seja, são extraídas as características mais relevantes da face para a classificação da expressão. Normalmente, é dada mais ênfase às sobrancelhas, olhos, nariz e boca como as características faciais mais importantes, pois as UAs utilizadas para o reconhecimento da emoção estão presentes, em sua maioria, nestas partes do rosto (EKMAN *et al.*, 2002) citado por (ABDAT *et al.*, 2011).
- **Classificação da expressão facial:** É a etapa da categorização da expressão facial. É realizada por um classificador, que utiliza uma técnica de reconhecimento de padrão da área de Aprendizagem de Máquina (PATIL *et al.*, 2010). Os dois

principais tipos de classes usadas em reconhecimento da expressão facial são as UAs (OLIVEIRA e JAQUES, 2008) e o protótipo das expressões faciais, definidas por Ekman (1982), que se relacionam com os estados emocionais de felicidade, tristeza, surpresa, raiva, medo e desgosto. Uma expressão resulta, tipicamente, da combinação de várias UAs (EKMAN, 1982) citado por (OLIVEIRA e JAQUES, 2008).

- **Emoção:** Esta etapa é o estado final do modelo arquitetural básico da maioria dos sistemas de reconhecimento de expressões faciais e corresponde a saída das emoções captadas pela câmera. Os estados emocionais frequentemente usados como saídas na maioria dos sistemas de reconhecimento de emoções da face são os de felicidade, tristeza, surpresa, raiva, medo e desgosto, seguindo o modelo proposto por Ekman (1982) citado por (OLIVEIRA e JAQUES, 2008).

No próximo capítulo, são apresentados alguns trabalhos relacionados, abordando tanto o desenvolvimento de sistemas de reconhecimento facial como sistemas de reconhecimento de emoções faciais.

3 TRABALHOS RELACIONADOS

Este capítulo descreve alguns trabalhos relacionados ao tema proposto nesta dissertação. A Seção 3.1 detalha tanto trabalhos desenvolvidos para o reconhecimento de faces de aplicação em geral como trabalhos de reconhecimento facial aplicados em AVAs. Já a Seção 3.2 descreve trabalhos desenvolvidos para o reconhecimento de expressões faciais.

3.1 RECONHECIMENTO DE FACES

O reconhecimento facial é um dos métodos biométricos de reconhecimento de padrão que utiliza características particulares da face. Este método apresenta várias vantagens sobre outras tecnologias biométricas (STAN e ANIL, 2011). A investigação nesta área tem sido conduzida há mais de 30 anos e, como resultado, o estado atual da tecnologia de reconhecimento facial está bem avançado (STAN e ANIL, 2011).

Existem duas abordagens básicas para o reconhecimento da face (STAN e ANIL, 2011). A primeira abordagem é baseada na extração de vetores característicos de partes básicas de uma face, tais como: olhos, nariz, boca e queixo. Esta abordagem utiliza modelos deformáveis e análise matemática para extrair informação das componentes básicas da face e, em seguida, converter essa informação em um vetor característico (STAN e ANIL, 2011). A segunda abordagem é baseada nos conceitos da Teoria da Informação (STAN e ANIL, 2011). Nesta abordagem a informação que melhor descreve uma face é derivada a partir da imagem da face toda.

Geralmente, modelos deformáveis de face visando extrair características usam retângulos para localizar os principais elementos da face (tais como boca, olhos e sobrancelhas) e, em seguida, são projetados modelos de curvas e funções de energia que caracterizam olhos, boca, queixo e contorno da face (ZHANG e RUAN, 2006).

Em (GHOSAL *et al.*, 2009), a extração das características de imagens da face pode ser obtida usando análise matemática com a transformada de *Wavelet Gabor*. Tais características podem ser obtidas através do algoritmo *Random Forest*, classificador de imagens da face com base nas características extraídas. O uso da transformada *Wavelet Gabor* conduz a um número extremamente elevado de características, a maioria das quais são redundantes para efeitos de classificação. Tal problema pode ser contornado eficientemente usando a técnica de *Random Forest* para calcular as características mais importantes de *Wavelet Gabor*, uma vez que o espaço das características é reduzido, acelerando o processo da classificação facial.

Em (KSHIRSAGAR *et al.*, 2011), métodos PCAs e estatísticos foram utilizados. Em (STAN e ANIL, 2011), foi implementado um sistema de reconhecimento de face utilizando as técnicas de Eigenfaces, PCA e Redes Neurais. Neste trabalho foi implementado o algoritmo PCA, o qual extrai as características mais relevantes que representam uma identidade única da imagem, características usadas como entradas da rede neural classificadora da face. Foi comprovado que os Eigenfaces têm a capacidade de fornecer as características significativas e reduz o tamanho de entradas da rede neural, aumentando a velocidade no processo de reconhecimento. No entanto, esta abordagem é sensível a imagens com condições de iluminação não controladas.

Na educação, poucos trabalhos abordando reconhecimentos faciais foram encontrados. Um dos trabalhos relacionados foi proposto pelos autores Rolim e Bezerra (2008). Neste trabalho, os autores apresentaram um sistema de identificação de rosto e integraram no AVA MOODLE. Esse sistema monitora os alunos remotos através de uma *webcam* para a captura periódica de imagens da face do aluno, tirada em intervalos de tempo predefinidos,

habilitando a segurança digital na identificação de alunos *online* no AVA. Segundo Rolim e Bezerra (2008), o sistema de reconhecimento facial utiliza uma técnica baseada em coeficientes DCT (*Discrete Cosine Transform*) para extrair as características da face presente. A Figura 29 ilustra a arquitetura do sistema.

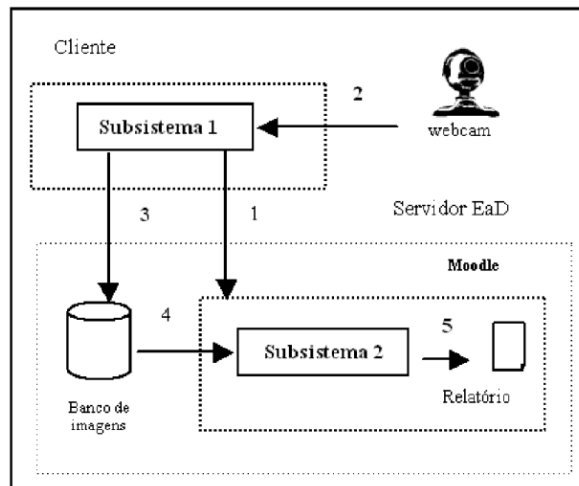


Figura 29 - Arquitetura do Sistema proposto por Rolim e Bezerra (2008).

A arquitetura proposta por Rolim e Bezerra (2008) é dividida em dois subsistemas. O Subsistema 1 faz a execução das páginas PHP no ambiente MOODLE (1). Em seguida, inicia o processo de captura de imagens em intervalos regulares a partir da webcam (2). As imagens capturadas são enviadas ao servidor que realiza o reconhecimento das faces e as armazenam em um banco de imagens. O subsistema 2 é uma extensão do MOODLE que recupera as imagens (4) e as apresenta nos relatórios do próprio ambiente (5), contendo informação do resultado do reconhecimento (afirmativo ou negativo) e mostra as imagens dos alunos durante sua presença no curso.

O resultado, verificado no sistema apresentado em (ROLIM e BEZERRA, 2008), atingiu acima dos 90% o qual também foi observado no trabalho proposto. Os experimentos feitos por Rolim e Bezerra (2008) foram feitos com imagens de 14 pessoas, cada uma com três poses. No trabalho proposto, os testes foram feitos com imagens de 64 pessoas, cada uma com cinco poses. Além disso, o trabalho de Rolim e Bezerra (2008) não utilizou o método de detecção de face Viola-Jones, o que acarreta aumento do custo computacional nas etapas de extração de características faciais e reconhecimento facial.

Agulla *et al.* (2008) desenvolveram um sistema, BioTracker, que processa imagens das faces de usuários. Utilizou-se uma arquitetura cliente/servidor. No lado do cliente, o sistema detecta a face, baseada na estimativa de obtenção de uma boa imagem (face frontal).

No lado do servidor, o processamento encarrega-se de extrair as características da imagem da face e compará-las com as presentes na base de dados. O sistema suporta tanto verificação colaborativa, na qual o aluno permite voluntariamente a captura de sua face, quanto não colaborativa, na qual o usuário não sabe que está sendo monitorado.

O sistema BioTracker auxiliou os processos de controle de acesso durante a participação do aluno nas atividades e avaliações. Segundo Agulla *et al* (2008), o sistema garante que o estudante está presente no computador, interagindo com o AVA, e valida quanto tempo o usuário gastou navegando pelo curso no sistema. No entanto, não assegura que os usuários não consigam violar o sistema. O BioTracker pode ser integrado aos sistemas MOODLE e Claroline (AGULLA *et al.*, 2008).

Neste trabalho dissertativo, é proposta a detecção da face com a técnica de Viola-Jones (VIOLA e JONES, 2004), que diminui os efeitos do fundo das imagens, além da técnica de Eigenfaces (TURK, 1991), que restringe as informações desnecessárias, aplicando, inicialmente, o classificador *K-Nearest Neighbors* (K-NN), depois o *Random Forest*, e por último, o *K-Star* na classificação das faces, como forma de obter o melhor resultado.

3.2 RECONHECIMENTO DE EXPRESSÕES FACIAIS

Ultimamente, o reconhecimento de expressões faciais tem estimulado o interesse dos pesquisadores, devido as suas potencialidades para a criação de interfaces baseadas no comportamento do ser humano (CHANGJUN *et al.*, 2011). O reconhecimento de expressões faciais tem uma função primordial na interação homem-computador. Esse recurso pode ser usado em várias aplicações, tais como: na realidade virtual, na videoconferência, para traçar o perfil do utilizador, para satisfazer o consumidor, e variados serviços Web, além de tornar mais real e amigável a interligação entre o homem e o computador (CHANGJUN *et al.*, 2011).

O problema do reconhecimento da expressão facial é abordado fazendo a distinção entre imagens estáticas ou sequências de imagens. No primeiro caso, encontram-se abordagens baseadas em modelos, redes neurais ou regras. No último, existem métodos baseados em modelos, em regras ou híbridos.

Foram encontrados, durante o levantamento bibliográfico, vários trabalhos que tratam sobre a automatização do reconhecimento de expressões faciais emocionais. Em sua maioria,

estes trabalhos tentam inferir se a expressão facial realizada por uma pessoa se ajusta entre uma das seis expressões básicas (alegria, desgosto, raiva, tristeza, medo e surpresa). São utilizadas, para isto, duas abordagens principais: classificadores sobre emoções (utilizando, por exemplo, redes neurais, máquinas de vetores de suporte (*Support Vector Machines* - SVM), análise discriminante linear (*Linear Discriminant Analysis* - LDA), entre outros) e o sistema FACS.

Na Tabela 5, resumem-se algumas das tentativas na área do reconhecimento de expressões faciais que tiveram como objetivo a automatização do reconhecimento de expressões faciais, com os seus respectivos métodos aplicados.

Tabela 5 - Resumo de trabalhos de reconhecimento de emoções faciais.

Referência	Método	Ações Faciais	Caso de Teste	Imagens	Precisão
(Pantic; Rothkrantz, 2000)	Baseado em regras	31 UAs	496 imagens de 8 sujeitos	Estáticas	89%
(Black; Yacoob, 1997)	Baseado em regras	Não Abordou	70 sequências de 40 sujeitos	Sequenciais	88%
(Cohn <i>et al</i> , 1998)	Baseado em padrão	8 UAs + 7 UAs e combina-ções	504 sequências de 100 sujeitos	Sequenciais	88%
(Edwards <i>et al</i> , 1998)	Baseado em padrão	Não Abordou	200 imagens de 25 sujeitos	Estáticas	74%

Se um método de classificação baseado em modelo é aplicado, a expressão facial encontrada é comparada com os modelos definidos para cada categoria da expressão. A melhor correspondência decide a categoria da expressão mostrada. Na classificação baseada em redes neurais, uma expressão facial é classificada de acordo com o processo de categorização que a rede aprendeu durante uma fase de treinamento. Uma rede neural é representada por ligações entre elementos de processamento a que se atribuem pesos, que consistem nos parâmetros que definem a função não linear efetuada pela rede neural. A determinação destes parâmetros é denominada de treino (*training*) ou aprendizagem (*learning*), sendo que as redes neurais são do tipo adaptativo.

Os métodos baseados em regras classificam as expressões faciais nas categorias básicas de emoção com base nas codificações das ações faciais. As expressões prototípicas, as quais caracterizam as categorias de emoções, são primeiramente descritas em termos das ações faciais. Em seguida, a expressão mostrada, descrita em termos de ações faciais, é comparada com as expressões prototípicas definidas para cada uma das categorias de emoções para, enfim, se classificar na categoria que melhor se encaixe na emoção definida.

Contudo, antes de abordar tais métodos, é importante considerar algumas questões gerais que estão relacionadas com a classificação de expressões faciais para ter um melhor proveito da capacidade do sistema. Primeiro, o sistema deve ser capaz de analisar qualquer indivíduo (homem ou mulher de qualquer idade e etnia). Em outras palavras, o mecanismo de classificação não pode depender da variabilidade fisionômica do indivíduo observado. Segundo, é importante compreender que a interpretação da linguagem corporal depende da situação. A informação sobre o contexto em que uma expressão facial aparece é muito difícil de se obter de uma forma automática. Esse problema não tem sido tratado pelos sistemas atualmente existentes.

Na educação, poucos trabalhos abordando reconhecimentos 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 SVM (CHANGJUN *et al.*, 2011), com o objetivo de estimar o estado afetivo dos alunos que utilizam um sistema de EaD baseado em computação afetiva. Um rastreador de faces reúne um conjunto de características de deslocamentos de movimento no sistema de EaD. Estes são posteriormente utilizados 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-se 265 imagens faciais de 42 pessoas como testes de conjunto. Atingindo uma taxa de reconhecimento total é de 84,55%. O modelo de sistema de EaD, proposto por Changjun *et al* (2011), com base em computação afetiva é apresentado na Figura 30. O modelo de sistema é composto por dois módulos principais:

- **Módulo de interface:** a entrada de computação afetiva (entrada do reconhecimento da emoção) é adicionada à interface homem-máquina do sistema de ensino baseado na web, que coleta informações das emoções dos alunos.

- **Módulo do reconhecimento da emoção:** é composto de entrada, pré-processamento da extração das características faciais, treinamento SVM, classificação SVM e da saída do reconhecimento de emoções.

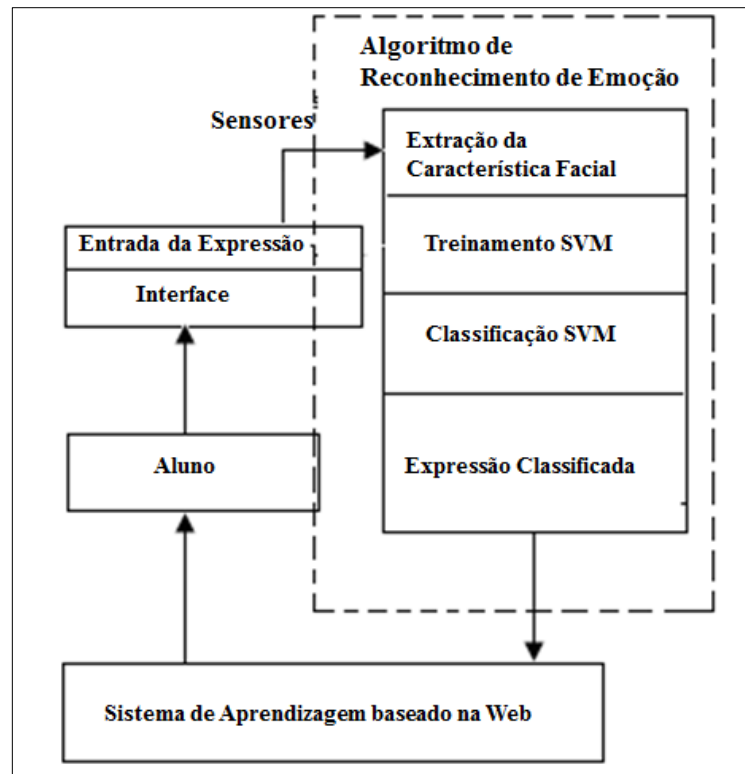


Figura 30 - Passos de reconhecimento de expressões faciais (CHANGJUN *et al.*, 2011).

No trabalho proposto, apresenta-se um sistema de reconhecimento de expressões faciais que infere 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 UAs. 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 e JONES, 2004) e o modelo CANDIDE-3 (AHLBERG, 2001).

No próximo capítulo será apresentado o sistema proposto. Detalhando o desenvolvimento do sistema e de seus componentes implementados. E, por fim, serão descritos e analisados os resultados do sistema implementado.

4 REDFACE: SISTEMA DE RECONHECIMENTO DE EXPRESSÕES FACIAIS DE APOIO A UM AVA

Este capítulo descreve a construção do módulo de reconhecimento facial e do módulo de reconhecimento de expressões faciais que compõem o sistema proposto neste trabalho. Mostra as arquiteturas desses módulos, descrevendo o funcionamento dos principais componentes envolvidos nessas arquiteturas. São destacadas as técnicas empregadas em cada etapa do desenvolvimento dos dois módulos propostos. E, por fim, são apresentados os resultados obtidos da avaliação do sistema e dos métodos utilizados sobre a base de dados coletada.

Para desenvolver as arquiteturas dos dois módulos para aplicação em um grande número de usuários, algumas decisões de projetos foram levantadas por motivos da complexidade computacional das técnicas utilizadas. Os algoritmos de PI e VC apresentam, em geral, grande consumo de recursos computacionais, dada a natureza matricial das imagens e sua complexidade algorítmica. Logo, o processamento do sistema foi projetado para ser distribuído entre o cliente e servidor, a fim de diminuir a quantidade de informação trocada entre os clientes e o servidor e, conseqüentemente, diminuir o tempo de resposta. Portanto, ao projetar a arquitetura dos módulos de reconhecimento facial e de expressões faciais, os seguintes itens foram levados em consideração:

- **Tráfego na rede:** transmite-se somente o vetor de características ao invés do *streaming* de vídeo, buscando diminuir o tráfego para somente o necessário à identificação do indivíduo;
- **Balanceamento do processamento:** a aplicação cliente é responsável por rastrear e extrair as características das faces, deixando para o servidor a tarefa da classificação, visto que ele será usado por todos os clientes;
- **Portabilidade:** é possível executar os módulos propostos nos principais sistemas operacionais (Linux ou Windows).

Na próxima seção será detalhada a implementação do sistema de reconhecimento facial. Descrevendo sua arquitetura e os comportamentos dos componentes existentes nas etapas do processo de reconhecimento facial. E, por fim, serão descritos os experimentos e seus resultados.

4.1 IMPLEMENTAÇÃO DO MÓDULO DE RECONHECIMENTO FACIAL

Nesta seção é descrito o projeto do módulo de reconhecimento facial utilizando, inicialmente, as técnicas de Análise de Componentes Principais (PCA) e Eigenfaces (autofaces) para a extração de características da face e, em seguida, aplicando os classificadores *K-Nearest Neighbors* (K-NN), *Random Forest* e *K-Star* na classificação das faces. Para a construção do protótipo foram observadas as características discutidas na Seção 2.3, que trata sobre as características e componentes de sistemas biométricos.

As subseções seguintes detalham a arquitetura geral proposta, os componentes, sua distribuição entre cliente e servidor, os experimentos e resultados. E, por fim, as discussões dos experimentos.

4.1.1 ARQUITETURA DO MÓDULO

Considerando o AVA proposto por Fontes *et al* (2012), assim como o levantamento do estado-da-arte no desenvolvimento de abordagens e sistemas computacionais na área de processamento facial, é proposto um sistema de arquitetura distribuída que identifica automaticamente um usuário que se submete ao uso do AVA através do reconhecimento de sua face em imagens digitais capturadas por uma webcam. A identidade do usuário será reconhecida dentre um conjunto de imagens de treinamento cadastradas no AVA (FONTES *et al*, 2012), e será anexada aos relatórios do AVA.

O módulo de reconhecimento facial proposto foi projetado e desenvolvido baseado na arquitetura Cliente/Servidor de modo a encapsular o acesso ao AVA proposto por Fontes (2012). Esse módulo consiste basicamente das seguintes etapas: (a) aquisição das imagens; (b) pré-processamento; (c) extração de características; e (d) classificação e verificação (STAN e ANIL, 2011).

A arquitetura do módulo de reconhecimento facial, apresentada na Figura 31, mostra a disposição dos subsistemas, distribuída entre cliente e servidor. Portanto, alguns subsistemas são implantados do lado Servidor e outros subsistemas no lado cliente, sendo que o conjunto de imagens deve ser armazenado no servidor com o propósito de manter o conjunto atualizado sempre que for preciso, assim como aumentar a privacidade e segurança dos dados do usuário, além disso, no treinamento, é gerado o vetor característico importante na

identificação do usuário. A carga de processamento do sistema deve ser distribuída entre cliente e servidor para se chegar a um tempo de resposta razoável, bem como limitar a quantidade de informação trocada entre os clientes e o servidor. A Figura 31 mostra o fluxo de controle e de informação dos subsistemas do módulo de reconhecimento facial, detalhando também como é feita a integração entre esses subsistemas com o Moodle.

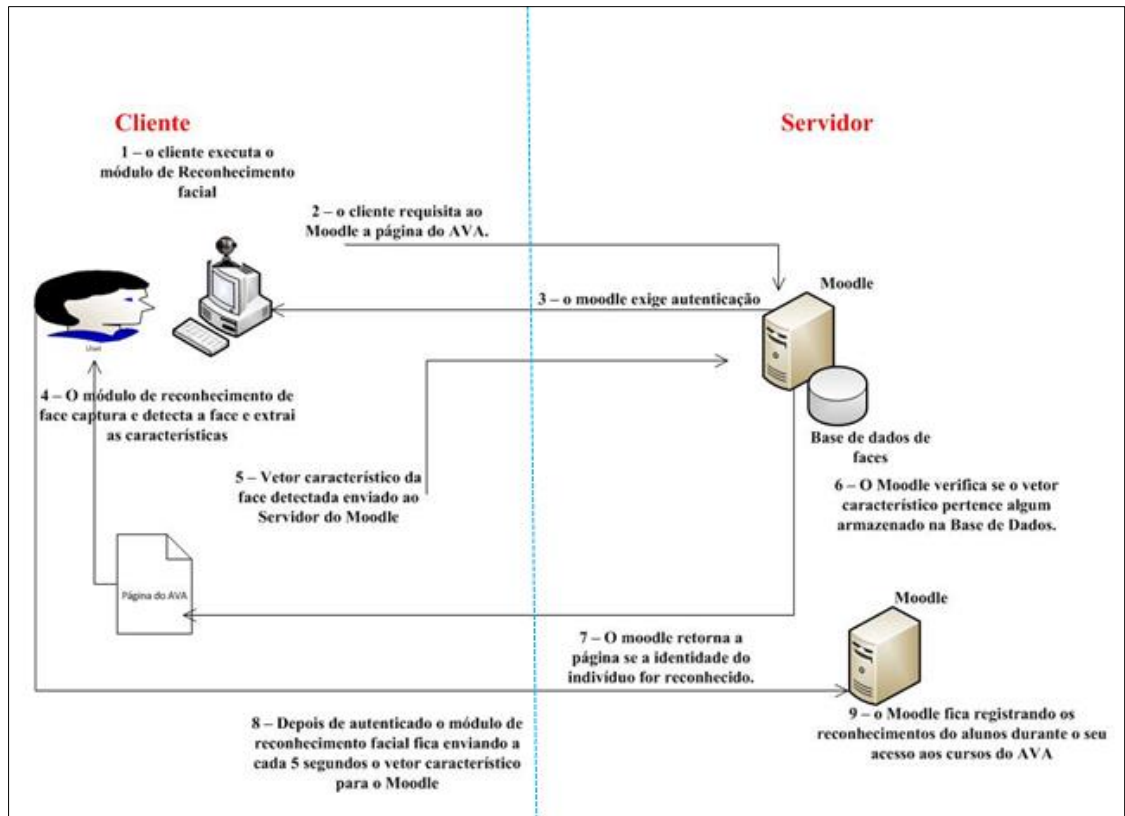


Figura 31 - Interação entre os subsistemas do módulo de reconhecimento facial localizados no cliente e no servidor do Moodle. Fonte própria.

O aluno, para poder acessar o curso, executa o módulo de reconhecimento facial e requisita, pelo navegador *web*, a página do curso para o servidor *Web*. O Moodle do AVA verifica se a página *Web* sendo requisitada exige autenticação biométrica. Este poderia ser o caso para uma avaliação ou um conteúdo protegido, por exemplo. Se não for, a página é enviada de volta ao cliente. Caso contrário, a aplicação cliente começa a capturar o vídeo por meio da *webcam* do usuário. A aplicação processa o vídeo, detecta, pré-processa as faces presentes nos quadros amostrados, extrai os vetores de características e envia-os para o servidor. O servidor consulta a base de dados pelo traço biométrico e retorna a resposta para o cliente *desktop*. A aplicação *desktop* então bloqueia ou retorna a página *Web*.

Assim, utilizando a face no reconhecimento das pessoas, a arquitetura distribuída proposta neste trabalho se mostra viável para o problema da identificação remota de indivíduos pela Internet com aplicação em sistemas de *e-Learning*, fazendo a distribuição da carga computacional entre cliente e servidor com pequena quantidade de informação trafegada. Em comparação com o trabalho de Rolim e Bezerra (2008) está arquitetura se mostra mais eficiente em termos de tempo de processamento quando aplicados no contexto da análise de imagens pela Internet, pois no trabalho de Rolim e Bezerra (2008) todo o processo de reconhecimento de face se realiza no lado do cliente, enviando para o servidor uma imagem rotulada como identificada ou não. Esse processo proposto por Rolim e Bezerra (2008) provoca uma sobrecarga maior no lado do cliente em comparação com a proposta neste trabalho, pois, pode prejudicar outros requisitos de projetos como, por exemplo, o tempo de resposta o qual é essencial em um ambiente *Web*. Já no trabalho de Agulla *et al.* (2008) o sistema de reconhecimento de face se realiza totalmente no lado do servidor, sobrecarregando demais o lado do servidor, provocando um aumento no custo computacional e no tempo de resposta.

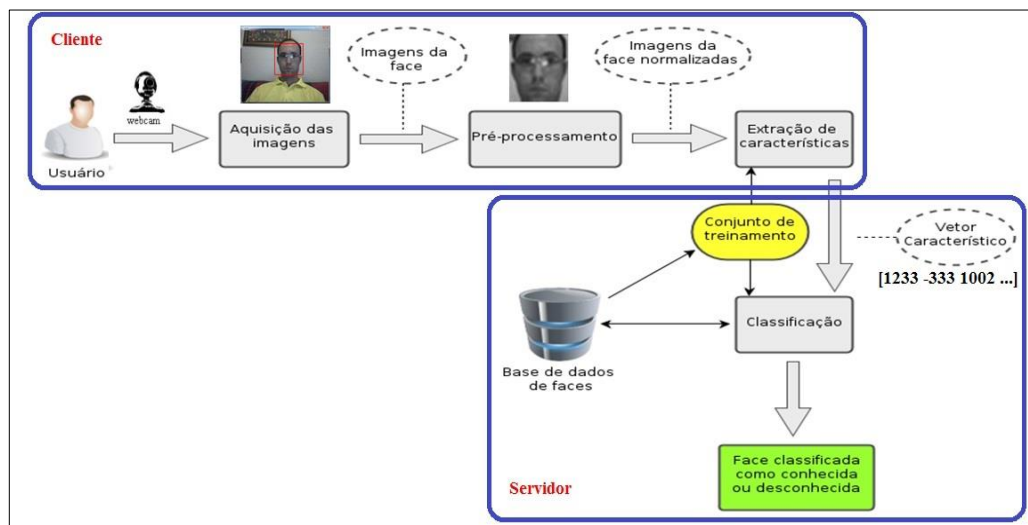


Figura 32 - Arquitetura do Módulo de Reconhecimento Facial. Fonte própria.

De acordo com a Figura 32, o início do processo de reconhecimento de face se dá a partir do subsistema de aquisição da imagem. Utilizando a *webcam*, o subsistema captura a imagem da face do usuário, em tempos predefinidos (a cada 5 segundo), a qual será utilizada como entrada para o módulo de reconhecimento facial. Nesse subsistema foi utilizado o algoritmo para detecção de face *Viola-Jones* (VIOLA e JONES, 2001) (visto na Subseção 2.3.1.1). O algoritmo *Viola-Jones* foi adotado neste trabalho por possuir uma detecção facial

bastante rápida em relação aos outros métodos existentes na literatura. A execução desse algoritmo faz com que propicie uma taxa de detecção tão boa quanto os outros métodos de detecção existentes na literatura, porém, com um tempo de processamento consideravelmente menor. 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.

O algoritmo *Viola-Jones* tenta encontrar em uma imagem características que codificam alguma informação da classe a ser detectada. Para tal tarefa, são usadas as chamadas características de *Haar*, responsáveis por codificar informações sobre a existência de contrastes orientados entre regiões da imagem. Neste trabalho, as imagens foram exibidas quadro a quadro e as faces foram detectadas e extraídas usando o algoritmo Viola-Jones que se encontra implementado na biblioteca OpenCV (INTEL, 2007) (detalhada na Subseção 2.3.3). A Figura 33 mostra a execução do algoritmo Viola-Jones atuando sobre um quadro do vídeo e detectando a face em tempo real.



Figura 33 – Execução do algoritmo de detecção de face Viola-Jones em um vídeo. Fonte própria.

O próximo subsistema é o de pré-processamento da imagem. Neste subsistema, as imagens da face detectadas pelo algoritmo Viola-Jones são automaticamente normalizadas e corrigidas para melhorar o reconhecimento da face. As imagens tiveram correções nas suas dimensões (todas foram redimensionadas somente para as dimensões 120x90, 32x24 e 12x9), na resolução (300 *dpi*) e na iluminação (foi utilizada a mesma em todas as fotos) e foram transformadas para escala de cinza. Esse subsistema é útil para poder fazer as comparações

com as imagens de faces cadastradas na base de dados, a qual é formada somente por imagens das dimensões 120x90, 32x24 e 12x9.

Foram utilizadas essas dimensões (120x90, 32x24 e 12x9) a fim de encontrar a melhor dimensão da imagem para reconhecimento da face, pois algoritmos de visão computacional apresentam, em geral, grande consumo de recursos computacionais, dada a natureza matricial das imagens. Para aplicações Web, em que vários usuários podem interagir simultaneamente com o sistema, esta deve ser uma preocupação, pois a experiência do usuário exige baixos tempos de resposta. Este subsistema também se tornou um diferencial comparado com os trabalhos de Rolim e Bezerra (2008) e Agulla *et al.* (2008), pois foi possível verificar resultados de reconhecimento de indivíduos com imagens de três dimensões diferentes. Tornando possível a escolha da melhor imagem para aplicações *web* (imagem com menor dimensão e que possui a melhor taxa de reconhecimento).

Após o pré-processamento, a imagem da face normalizada serviu de entrada para o subsistema de extração de características, objetivando encontrar as principais características a serem utilizadas para a classificação. Cada imagem representante do indivíduo foi transformada em uma matriz de tamanho $w \times h$, onde w e h são, respectivamente, os números de *pixels* referentes à largura e à altura da imagem. Cada valor do *pixel* da imagem corresponde a uma componente do vetor.

Devido à alta dimensionalidade dos vetores, foi utilizada, nesta etapa, a técnica de PCA, a qual reduz a quantidade de características de uma imagem. Logo, conseqüentemente, diminui o custo computacional e melhora a precisão do classificador. Se o espaço característico contiver somente as características mais visíveis, o classificador será mais rápido e ocupará menos memória (AGARWAL *et al.*, 2010). Juntamente com a técnica PCA, foi utilizada a técnica do Eigenfaces. O algoritmo Eigenfaces objetiva fornecer um conjunto de vetores de distribuições probabilísticas para resolver o problema da detecção de padrões em imagens e seu fundamento básico é a utilização desses vetores para gerar uma informação matemática do rosto de um ser humano para sua futura identificação. Ambas as técnicas foram descritas na Subseção 2.4.1.

Com a ajuda de um classificador de padrões, as características extraídas da imagem da face são comparadas com as amostras do conjunto de treinamento de face. A imagem da face é então classificada como conhecida ou desconhecida. Neste trabalho foram analisados três algoritmos de reconhecimento de padrões: K-NN (JIANGSHENG, 2002), *Random Forest* (BREIMAN, 2001) e *K-Star* (CLEARY e TRIGG, 1995). Estes dois últimos foram utilizados

a partir do software de mineração de dados Weka (*Waikato Environment for Knowledge Analysis*) (WEKA, 20011).

O último subsistema da Figura 32 é o conjunto de treinamento. Ele visa encontrar características apropriadas para a representação de padrões de entrada, sendo o classificador treinado para particionar o espaço de características, formando padrões de todas as classes existentes.

As subseções seguintes detalham os algoritmos de classificação utilizados neste trabalho. Todos esses algoritmos foram selecionados e utilizados neste trabalho por serem simples na codificação e de conhecimento técnico pelo autor deste trabalho.

4.1.1.1 CLASSIFICADOR *K-NEAREST NEIGHBORS*

O K-NN é um dos algoritmos de classificação mais utilizados na área de aprendizagem de máquina (JIANGSHENG, 2002). É baseado na procura dos k vizinhos mais próximos do padrão de teste. A busca pela vizinhança é feita utilizando uma medida de distância nessa procura. Nos experimentos desenvolvidos neste trabalho, o K-NN foi implementado utilizando as distâncias *Euclidiana*, *Manhattan* e a *Euclidiana* normalizada (JIANGSHENG, 2002). Além disso, todos os experimentos com K-NN foram realizados com peso pela distância e não pela frequência.

Dessa maneira, o padrão é classificado de acordo com a soma dos pesos dos k vizinhos – o peso é o inverso das distâncias. O padrão é dito pertencer à classe que apresentar maior frequência dentre os k vizinhos utilizados (JIANGSHENG, 2002). O k escolhido será o que proporcionar menor quantidade de erros na classificação das imagens.

Na Figura 34 temos um exemplo do funcionamento do *K-NN*. Se X tem três vizinhos “halterofilistas” e dois vizinhos “jôqueis”. A medida mais utilizada é a distância *Euclidiana* (JIANGSHENG, 2002). Classifica-se um dado padrão associando a ele a classe que apresentar a maior soma dos pesos entre os k vizinhos mais próximos. Neste exemplo X será classificado como halterofilistas.

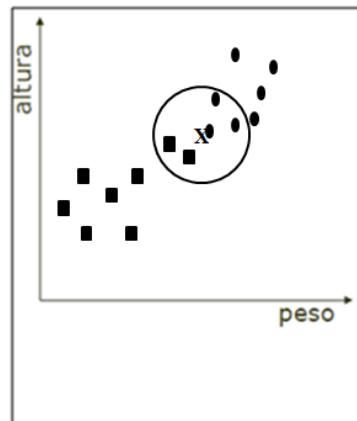


Figura 34 - Exemplo de funcionamento do K-NN. Fonte própria.

4.1.1.2 CLASSIFICADOR *RANDOM FOREST*

O método de classificação conhecido como *Random Forest* foi proposto em (BREIMAN, 2001). Ele consiste em uma técnica de agregação de classificadores do tipo árvore, construídos de forma que a sua estrutura seja composta de maneira aleatória (GHOSAL *et al.*, 2009). Para determinar a classe de uma instância, o método combina o resultado de várias árvores de decisão, por meio de um mecanismo de votação. Cada árvore dá uma classificação, ou um voto para uma classe. A classificação final é dada pela classe que recebeu o maior número de votos entre todas as árvores da floresta (BREIMAN, 2001).

O classificador é baseado no método *Bagging* (BREIMAN, 2001). Para cada árvore gerada é utilizado um conjunto de treinamento diferente, formado por n instâncias de treinamento escolhidas aleatoriamente (i.e., uma amostra ou *bootstrap*). Para cada nó da árvore gerada, são escolhidos aleatoriamente m atributos que orientam o direcionamento do nó, baseado na melhor discriminação de classes do conjunto de treinamento, de acordo com uma métrica. Em geral, o valor de m deve ser bem menor que o total de atributos da base, de maneira que possam ser geradas árvores distintas, que são combinadas para classificar uma nova instância (BREIMAN, 2001).

O modelo gerado elege a classe mais frequente entre as opções individuais de cada árvore. Dessa forma, a seleção de atributos é feita no instante de construção do modelo de classificação, caracterizando a seleção do tipo embutida (*embedded*). A vantagem desse classificador é que ele permite bases de dados com um número grande de atributos, contudo é suscetível a super ajuste (*overfitting*) em determinadas bases (SEGAL, 2004).

Como mencionado anteriormente, esse algoritmo foi utilizado a partir da ferramenta Weka (WEKA, 2011), que proporciona a variação do número de árvores e o número de características a serem analisadas. Escolhemos essa ferramenta por ser amplamente utilizada em atividades de aprendizado de máquina e por ser adequada a conjuntos de dados ruidosos, apresentando um bom desempenho na análise de imagens (ZAINAL *et al.*, 2008).

4.1.1.3 CLASSIFICADOR *K-STAR*

K-Star é um classificador baseado em exemplos, isto é, baseia-se na classe das instâncias de formação semelhante, conforme determinado por uma função de similaridade (CLEARY e TRIGG, 1995). Ele difere de outros por utilizar funções de distância baseadas na entropia e assume que os exemplos similares terão classes similares. O algoritmo *K-Star* utilizado consiste em uma implementação em Java, sendo parte integrante do pacote Weka (FRANK *et al.*, 1999). O *K-Star* possui os parâmetros de configuração para a execução, os quais são *globalBlend* e *missingMode* (WEKA, 2011).

Para a execução do *K-Star*, é necessário configurar dois parâmetros (FRANK *et al.*, 1999; WEKA, 2011):

- ***Global Blend***: Parâmetro correspondente à mistura, relacionado ao uso de cálculos de entropia, com valores definidos de 0 a 100 e expressos em porcentagem (B=20);
- ***Missing Mode***: Determina como os valores de atributos faltantes são tratados.

Na próxima subseção são apresentados os experimentos, descrevendo os materiais e a base de dados utilizados. São descritos os resultados obtidos do módulo e dos métodos abordados na Subseção 4.1.1

4.1.2 EXPERIMENTOS E RESULTADOS

Para a captura dos vídeos, bem como para os métodos de detecção de faces (Viola-Jones) foram utilizadas as implementações padrão da biblioteca de código livre e multiplataforma OpenCV (INTEL, 2007). Esta biblioteca contém vários algoritmos e métodos

de VC escrito nativamente em C e com rotinas otimizadas para processadores da arquitetura *Intel*. Já na representação das faces (PCA) foi utilizada a biblioteca padrão do Java.

Como ambiente integrado de desenvolvimento do sistema implementado foram utilizadas as IDEs (*Integrated Development Environment*) *Microsoft Visual Studio 2010* com a linguagem C++ padrão para extensões da biblioteca OpenCV com suas bibliotecas de acesso a dados, interface com o usuário e comunicação entre processos. O banco de dados escolhido para armazenar as faces foi MySQL para o armazenamento das faces cadastradas pelos alunos do AVA proposto por Fontes (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 das técnicas e dos seus parâmetros. Todos os algoritmos e técnicas (assim como suas variações) foram executados de forma automática e exaustiva, de modo que foi analisado somente o melhor resultado da taxa de acerto para o conjunto de testes (número de acertos dividido pelo número total de testes), utilizando o 1-NN (primeiro vizinho mais próximo) e distância *Euclidiana* normalizada. Primeiro foi executado o redimensionamento (*resize*) das imagens, resultando em três dimensões diferentes: 120x90 – tamanho normal das imagens; 32x24 – tamanho aproximado de um ícone e 12x9 – melhor resultado dos testes automáticos. Além disso, todas as imagens faciais tiveram correções nas suas iluminações, ou seja, todas as imagens tiveram as mesmas iluminações, pois com imagens pouco iluminadas o método *Eigenface* perde sua eficiência na extração das características.

Nas simulações deste trabalho, foi utilizada uma base de dados própria, possuindo 1280 imagens de um total de 64 pessoas. Essas imagens são de pessoas conhecidas dos autores deste trabalho. Essas pessoas estão cadastradas no AVA (Fontes, 2012) no qual esse sistema de reconhecimento facial foi integrado, sendo este o um principal motivo para a utilização dessa base de dados própria para os testes dos métodos abordados neste trabalho. No entanto, em trabalhos futuros serão utilizadas bases de imagens padronizadas (GROSS *et al.*, 2008; MISLAV *et al.*, 2011; SHARMA e REILLY, 2003) para comparações de resultados, verificando a taxa de acerto em relação à literatura existente. A Figura 35 ilustra algumas faces armazenadas na base de dados utilizada.



Figura 35 - Imagens de faces da base de dados. Fonte própria.

Cada indivíduo integrante da base de dados é representado por 20 imagens nas seguintes posições: 4 imagens olhando diretamente para a câmera, 4 imagens olhando 45 graus à esquerda da câmera, 4 imagens olhando 45 graus à direita da câmera, 4 imagens olhando 90 graus à esquerda da câmera e 4 imagens olhando 90 graus à direita da câmera, como pode ser visto na Figura 36.

Dessas 20 imagens, 5 (cada uma de uma posição) compõem o conjunto de treinamento do sistema e as 15 restantes compõem o conjunto de teste.



Figura 36 - Esquema de posicionamento da face. Fonte própria.

Em seguida, os testes foram desenvolvidos tentando encontrar o melhor número de características restantes após o PCA (número de autovetores do *Eigenfaces*). Foram utilizados testes com valores de características no intervalo de 15 a 120 para encontrar os melhores resultados tanto no reconhecimento da face como no custo computacional. Em caso de empate (mesma taxa de acerto para número de características diferentes), o menor número de características foi escolhido como melhor por garantir menor custo computacional. O resultado obtido para cada dimensão foi: 120x90 – 17 características; 32x24 – 19 características e 12x9 – 20 características.

A Figura 37 mostra os resultados da taxa de acerto segundo as configurações: as três dimensões das imagens; número de características variando de 15 a 20 (não são mostrados os testes completos por serem extensos e por não terem apresentado bons resultados); e k no intervalo de 1 a 10; fazendo uso das três medidas de distância já citadas.

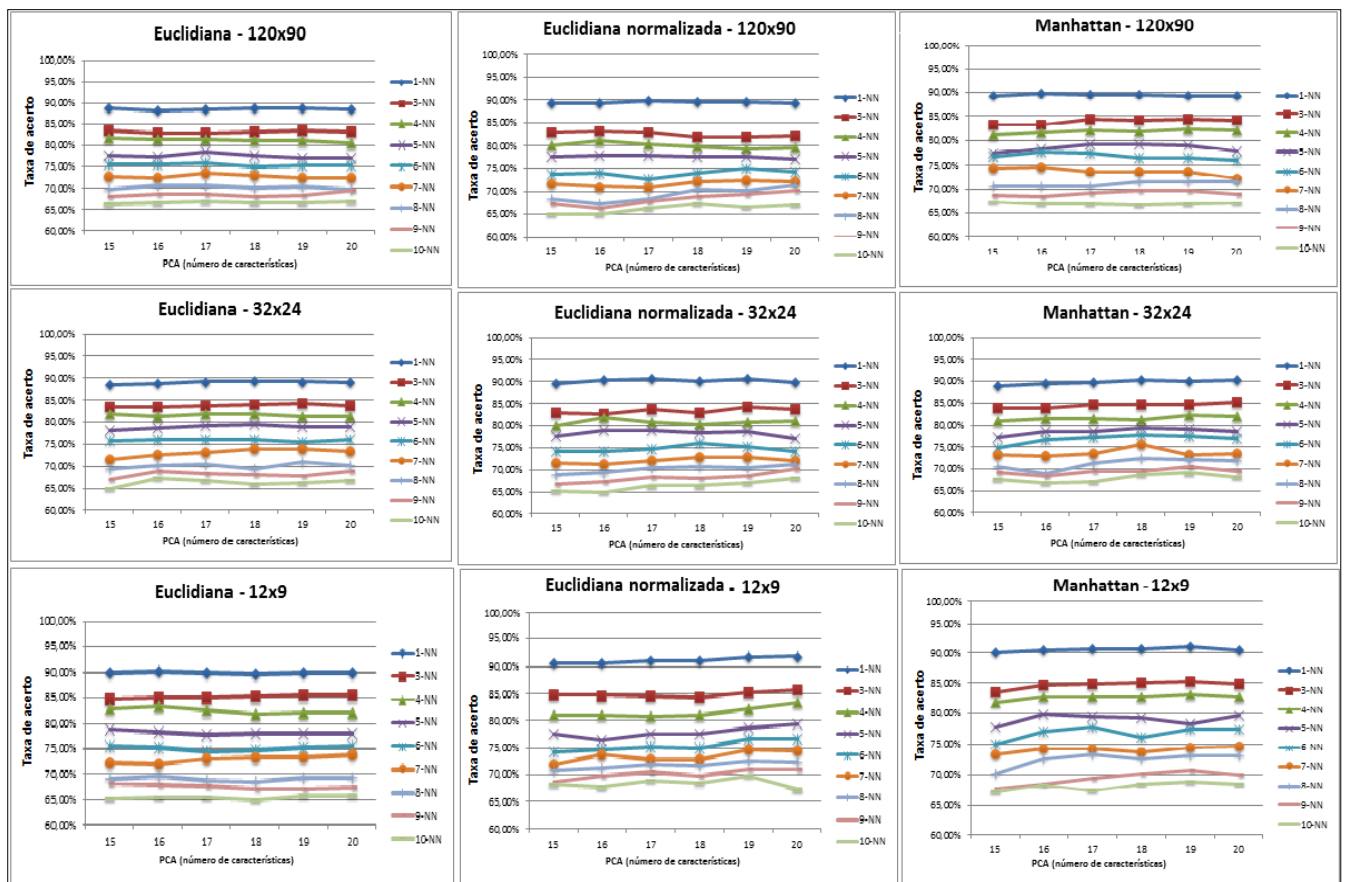


Figura 37 - Resultados da classificação para as imagens de dimensões 120x90, 32x24 e 12x9.

Fonte própria.

Os algoritmos *Random Forest* e *K-star* foram simulados utilizando o Software Weka (WEKA, 2011). Os experimentos seguiram a metodologia *10-fold-cross-validation* (GHOSAL *et al.*, 2009). As sequências foram aleatoriamente separadas em 10 grupos de tamanhos aproximadamente iguais. Um grupo foi utilizado como conjunto de novas sequências apresentadas ao sistema para o aprendizado e os 9 grupos restantes foram utilizados como sequências de treinamento dos algoritmos. No *Random Forest* varia-se o número de árvores e o número de características a serem analisadas. As variações testadas e os resultados se encontram na Tabela 6.

Verifica-se, na Tabela 6, que ao aumentar o número de características e mantendo o número de árvores, implica na diminuição da taxa de acertos. Isso pode ser visto nas linhas três e quatro da Tabela 6, com ambas mantendo número de árvores de cem, e também nas linhas nove e dez da Tabela 6, com ambas mantendo o número de árvores de quatrocentos e cinquenta. Nesta tabela, pode-se observar que a melhor taxa de acerto resultante foi 91,4%, utilizando 450 árvores e 2 características.

Tabela 6 - Testes *Random Forest*.

Número de árvores	Número de Características	Taxa de acerto
10	5	78,2292%
50	5	87,5%
100	3	89,6875%
100	5	88,5417%
200	3	90,9375%
250	3	90,4167%
300	3	90,625%
400	3	90,8333%
450	2	91,4583%
450	3	91,0417%

No *K-Star*, os valores testados para o *global blend* (descrito na Seção 3.3), assim como os resultados obtidos estão descritos na Tabela 7. A melhor taxa de acerto resultante foi aproximadamente 91%, utilizando um *Global Blend* de 100%.

Tabela 7 - Testes *K-Star*.

<i>Global Blend</i>	Taxa de acerto
20%	88,4375%
50%	89,375%
100%	91,0417%

Analisando os três métodos, ambos tiveram como melhores resultados valores parecidos, acima de 90%. Destacando o melhor resultado obtido de testes com a técnica *K-NN* em imagens de dimensões 12x9 e utilizando números de 20 características resultantes do PCA, atingindo uma taxa de reconhecimento de aproximadamente 92%.

4.1.3 DISCUSSÕES DOS EXPERIMENTOS

Na Figura 37, nos gráficos para imagens de face de dimensão 120 x 90 verifica-se a maior taxa de acerto (89,9%), usando o $k=1$, número de características (autovetores) igual a 17 e distância *Euclidiana* normalizada. A menor taxa de acerto encontra-se quando se usa o $k=10$ e distância *Euclidiana* normalizada. Vale destacar aqui que para os outros valores de k os melhores resultados se encontram utilizando a distância *Manhattan*. Nos gráficos das faces de dimensão 32x24 (Figura 37) verifica-se que a maior taxa de acerto (90,63%) é usando o $k=1$, número de autovetores igual a 19 e distância *Euclidiana* normalizada. A menor taxa de acerto encontra-se quando se usa o $k=10$ e distância *Euclidiana*. Vale destacar aqui também que para os outros valores de k os melhores resultados se encontram utilizando a distância *Manhattan*. Além disso, há uma uniformidade dos melhores resultados, se encontrando entre os valores característicos do PCA variando entre 18 e 20.

Nos gráficos das faces de dimensão 12x9 (Figura 37) verifica-se que a maior taxa de acerto (91,98%) foi usando o $k=1$, número de autovetores igual a 20 e distância *Euclidiana* normalizada. A menor taxa de acerto encontra-se quando se usa o $k=10$ e distância *Euclidiana*. Vale destacar aqui que para os outros valores de k os melhores resultados se encontram utilizando a distância *Euclidiana* normalizada, com os melhores resultados se encontrando entre os valores característicos do PCA variando entre 19 e 20. Concluímos com os resultados obtidos nas três dimensões (Figura 37) que, se os valores de k crescem, implica na diminuição da precisão da classificação. Além disso, quanto menor a dimensão do *pixel*, maior é a resolução espacial da imagem. Quanto maior a resolução espacial da imagem, maior serão a definição e detalhe da imagem. Uma imagem com resolução alta ficará menos borrada e com mais detalhes, melhorando o reconhecimento do padrão. Isto favoreceu os melhores resultados encontrados nas imagens de dimensão 12x9.

Analisando os três algoritmos, ambos tiveram como melhores resultados valores parecidos, em torno de 90% acima. Destacando o melhor resultado na taxa de acerto obtido de

testes com a técnica K-NN em imagens de dimensões 12x9, utilizando números de 20 características resultantes do PCA e $k=1$, atingindo uma taxa de reconhecimento de aproximadamente 92%. Resultado útil no trabalho proposto, pois imagens de dimensões menores implica diminuição do vetor da imagem comparado com as outras imagens analisadas de dimensão maiores, conseqüentemente reduz a etapa de pré-processamento. Além disso, com a utilização de somente 20 características diminui o tráfego na rede durante a transmissão do vetor característico para o servidor. E com o $k = 1$ torna a classificação mais rápida no lado do servidor, pois reduz o número de vizinhos próximos a serem comparados, diminuindo o tempo de resposta ao indivíduo no lado do cliente a ser identificado.

Na Seção 4.2 será detalhada a implementação do sistema de reconhecimento de expressões faciais. Descrevendo sua arquitetura, os comportamentos dos componentes existentes nas etapas do processo de reconhecimento da emoção do indivíduo por meio da expressão facial. E, por fim, descreve os experimentos e seus resultados.

4.2 IMPLEMENTAÇÃO DO MÓDULO DE RECONHECIMENTO DE EXPRESSÕES FACIAIS

Esta seção apresenta o projeto do módulo de reconhecimento de expressões faciais que infere as emoções alegria, tristeza, raiva, e desgosto, utilizando imagens 3D (três dimensão) sequenciais capturadas pela câmera do Kinect (MICROSOFT RESEARCH, 2011). A aplicação desenvolvida baseia-se no sistema psicológico de codificação facial FACS, que classifica todos os movimentos musculares faciais, denominados de Unidades de Ação (UAs) (EKMAN *et al.*, 2002). Foi projetado e implementado um novo algoritmo baseado em regras que enferem as emoções. Esta nova abordagem verifica as deformações geométricas da boca, olhos e sobrancelhas, depois mapeia essas deformações àquelas UAs presentes nas expressões faciais das emoções inferidas. Por fim, verifica as UAs encontradas, inferindo a emoção que possui a maior quantidade de UAs presentes durante a expressão facial realizada atual.

Para isso, este trabalho utilizou a técnica Viola-Jones (VIOLA e JONES, 2004) para a detecção da face e o rastreamento dos pontos característicos da face que se baseou no modelo de CANDIDE o qual é uma máscara facial parametrizada, baseada em modelo, especificamente desenvolvida para a codificação de rostos humanos (AHLBERG, 2001). Seu baixo número de polígonos (cerca de 100) permite a reconstrução rápida com poder de

computação moderado (AHLBERG, 2001). Ambas as técnicas já se encontram implementadas na API (*Application Programming Interface*) do Kinect (MICROSOFT RESEARCH, 2011).

As subseções seguintes detalham a arquitetura geral proposta, os componentes, sua distribuição entre cliente e servidor, os experimentos e os resultados.

4.2.1 ARQUITETURA DO MÓDULO

O módulo de reconhecimento de expressões faciais proposto também foi projetado e desenvolvido no padrão Cliente/Servidor de modo a encapsular o acesso ao AVA proposto por Fontes (2012). A arquitetura do módulo de reconhecimento de expressões faciais, apresentada na Figura 39, mostra a disposição dos seus subsistemas. Todos os subsistemas são implantados do lado Cliente os quais, em tempo real, capturam, a cada 5 segundos, imagens 3D sequenciais de um indivíduo, detecta e extrai as características da face, e, por fim, classifica as expressões faciais e infere as emoções. Ao encontrar as emoções, envia para o servidor do AVA proposto por Fontes *et al* (2012) que registra as emoções encontradas na base de dados. A Figura 38 mostra o fluxo de controle e de informação do módulo de reconhecimento de expressões faciais com o AVA.

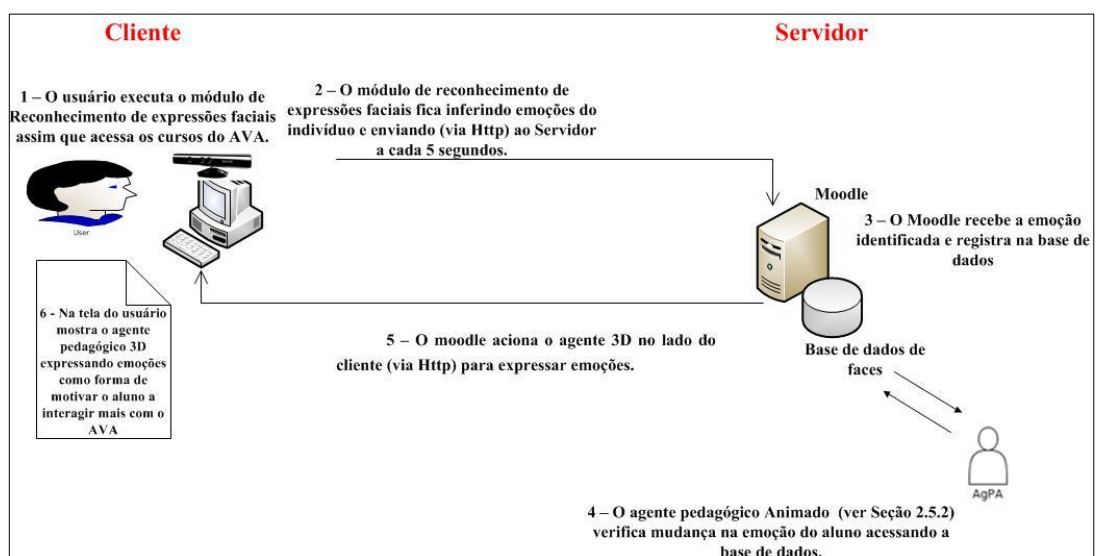


Figura 38 – Interação do módulo de reconhecimento de expressões faciais com o AVA. Fonte própria.

Conforme visto na Figura 38, quando o aluno acessa algum curso do AVA composto por agentes (FONTES *et al.*, 2012), executa o módulo de reconhecimento de expressões faciais. Ao ser executado, este módulo processa, no lado do Cliente, todas as etapas da arquitetura do reconhecimento de expressão facial (ver Figura 39) a fim de identificar as emoções do aluno, e fica enviando essas emoções para o Servidor a cada 5 segundos. Ao receber as emoções, o Moodle as registra na base de dados. O agente do AVA, em tempos pré-definidos, verifica a mudança da emoção do aluno acessando a base de dados, e, de acordo com a emoção verificada, aciona o agente 3D animado para expressar emoções com o objetivo de motivar o aluno a interagir mais no AVA (FONTES *et al.*, 2012).

O módulo de reconhecimento de expressões faciais consiste basicamente das seguintes etapas (ver Figura 39): (a) Captura das imagens; (b) Rastreamento das características faciais (subetapa de detecção da face e detecção das características faciais); (c) Classificação da expressão facial; (d) Inferência da Emoção.

Este módulo propõe inferir algumas das emoções básicas (EKMAN, 1993) (raiva, desgosto, alegria e tristeza) captadas pelas expressões faciais através de imagens provenientes da câmera do Kinect. Essas emoções serviram como entrada para um AVA, composto por um agente inteligente pedagógico que expressa emoções. Este agente tem como propósito proporcionar atividades desafiadoras que oferecem uma maior motivação e entusiasmo nos processos de ensino-aprendizagem.

De acordo com a arquitetura do módulo de reconhecimento de expressão facial, apresentada na Figura 39, o processo de reconhecimento da emoção proposto é composto pelos seguintes passos:

1. Primeiro é realizada a captura da imagem por uma câmera;
2. Depois é rastreada a face e, conseqüentemente, os pontos característicos da face (*Facial Characteristic Points* - FCPs) dos locais de interesse (olhos, boca e sobrancelhas);
3. As coordenadas dos FCPs são submetidas às regras que identificam a ação das características faciais (ex. boca aberta, olho fechado);
4. As ações faciais são traduzidas para UAs, as quais correspondem às expressões faciais específicas (ex. UA1 = cantos das sobrancelhas internas levantadas);
5. Por fim, as combinações das ações faciais (as UAs) encontradas são traduzidas em emoções.

Os passos 3, 4 e 5 são os que implementam o algoritmo proposto baseado em regras que enferem as emoções.

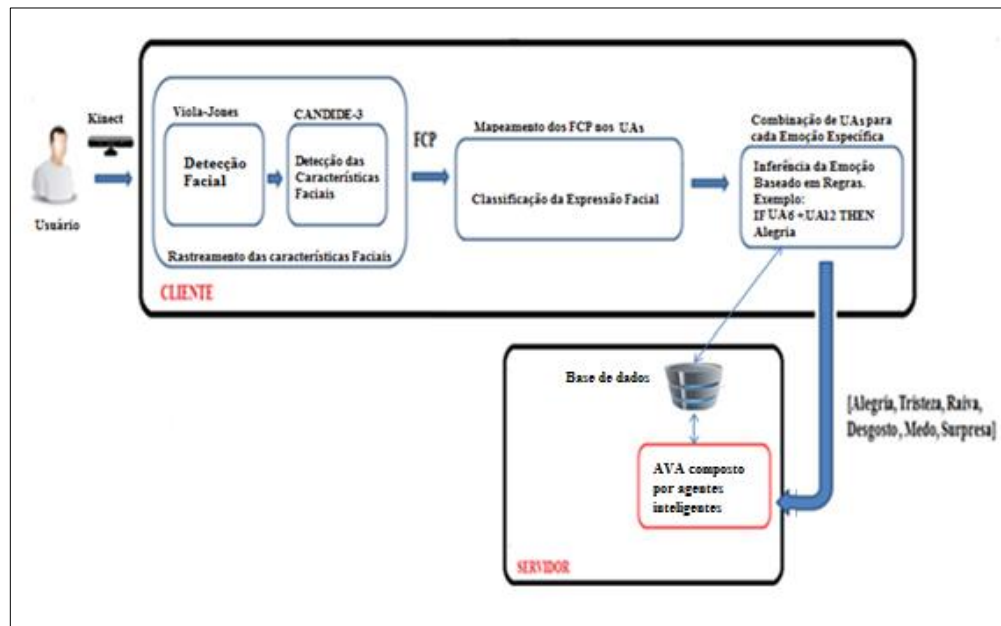


Figura 39 – Arquitetura do módulo de reconhecimento de expressão facial. Fonte própria.

Nas subseções seguintes serão descritas as etapas desta arquitetura, as quais são agrupadas em três subsistemas: rastreamento das características faciais, classificação da expressão facial e inferência da emoção.

4.2.1.1 RASTREAMENTO DAS CARACTERÍSTICAS FACIAIS

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 subsistema de rastreamento das características faciais. O motor de rastreamento das características faciais, utilizando o Kinect, rastreia em torno de 4 a 8 milissegundos, dependendo do computador.

No entanto, nesse módulo de reconhecimento de expressões faciais foi atribuído um tempo de 5 segundos para o rastreamento, a fim de dar tempo suficiente para que a emoção encontrada chegue ao AVA e seja avaliada pelos agentes pedagógicos. Além disso, a duração

de uma expressão espontânea normalmente varia entre 0,5 segundos e 5 segundos (EKMAN, 2003).

Nesse subsistema foi utilizada a técnica de detecção de face *Viola-Jones* (VIOLA e JONES, 2001) e a técnica de detecção das características faciais CANDIDE-3 (AHLBERG, 2001), os quais em conjunto rastreiam os FCPs dos locais de interesse da face (boca, olhos, sobrancelhas e nariz). Ambas as técnicas foram automatizadas, utilizando a biblioteca do Kinect (Kinect SDK) (MICROSOFT RESEARCH, 2011).

A escolha do algoritmo Viola-Jones na detecção da face foi pelos mesmos motivos ditos na Subseção 4.1.1, que trata o reconhecimento das faces. Ou seja, principalmente, porque possui 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. Mais detalhes na Subseção 2.3.1.1.

O CANDIDE-3 é uma técnica de face parametrizada especificamente desenvolvida para codificação de faces humanas baseada em modelo e é usada em vários projetos citados na literatura (DORNAIKA e AHLBERG, 2004). Ele é ajustado à face do indivíduo no primeiro quadro (*frame*) de vídeo e acompanha até ao final do mesmo, enquanto a expressão facial vai sofrendo evoluções, pois é controlado pelas várias UAs.

Embora tenha um número baixo de polígonos (aproximadamente 100 polígonos), consistindo de 113 vértices, captura os mais importantes pontos característicos da face humana. Portanto, pode proporcionar um nível relativamente elevado de modelo e ao mesmo tempo permite uma reconstrução rápida com capacidade computacional moderada. Mais detalhes na Subseção 2.3.2.1. A combinação das duas técnicas (CANDIDE-3 e Viola-Jones) atinge um nível aceitável de qualidade, mantendo ao mesmo tempo um custo computacional relativamente baixo. Logo, essa combinação é essencial para aplicações que têm resposta imediata ao usuário, como é o caso da identificação dos alunos nos AVAs.

A Figura 40 mostra a execução do sistema detectando e rastreando as características faciais de um indivíduo. Com os algoritmos Viola-Jones e CANDIDE-3 atuando sobre um quadro do vídeo, detectando a face, as características da face e os FCPs em tempo real. O retângulo, visualizado na cor rosa, mostra a execução do algoritmo Viola-Jones, já a máscara, visualizada na cor amarela, mostra o funcionamento do algoritmo CANDIDE-3. A tela do lado esquerdo mostra a extração das características da face com os FCPs interligados, entre si, por meio de arestas nos locais de interesse (boca, nariz, sobrancelha e olhos).

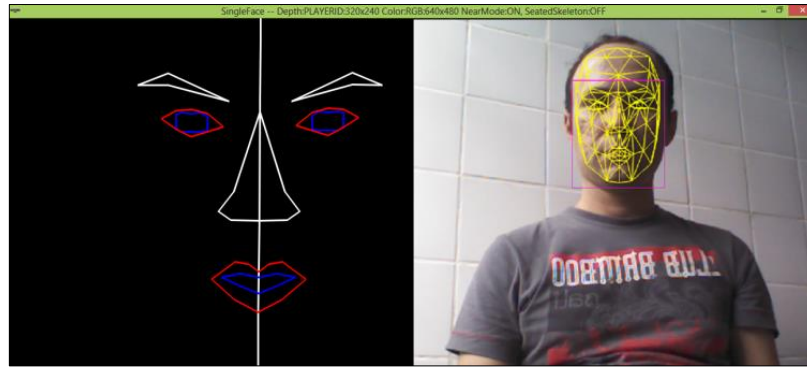


Figura 40 – Execução do rastreamento das características faciais. Fonte própria.

Os FCPs encontrados pelo rastreamento das características faciais, utilizando o CANDIDE-3, totalizam 113 vértices, os quais correspondem às bordas de toda a face (Apêndice C). No entanto, só foram necessários 38 destes vértices no desenvolvimento do sistema de reconhecimento de expressões faciais neste trabalho. Esses vértices foram utilizados como pontos característicos (ilustrados na Figura 41) para fazer os mapeamentos às suas respectivas UAs.

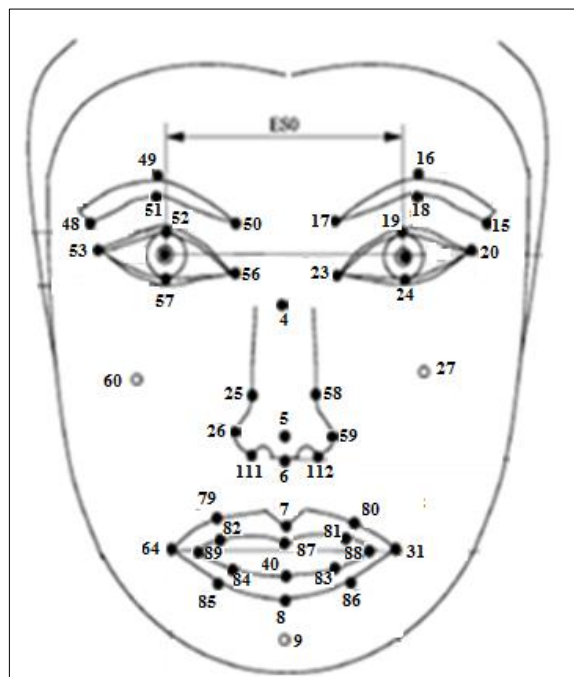


Figura 41 – FCPs utilizados para o mapeamento das UAs. Fonte própria.

Na próxima subseção será detalhada a classificação das expressões faciais, descrevendo o processo do mapeamento dos FCPS nas respectivas UAs.

4.2.1.2 CLASSIFICAÇÃO DAS EXPRESSÕES FACIAIS

Neste subsistema inicia o algoritmo proposto baseado em regras. É realizado o mapeamento dos FCPs, vindas do subsistema anterior, com as respectivas UAs definidas em FACS (EKMAN *et al.*, 2002). Um exemplo que pode ser demonstrado é o levantar da parte interna das sobrancelhas, movimentação que pelo FACS representa a UA 1 (*Inner brow raiser*). Portanto, a saída desse subsistema são as UAs específicas das expressões faciais realizadas pelos indivíduos.

Para encontrar as UAs, é necessário primeiramente fazer algumas transformações geométricas nas coordenadas dos FCPs, a fim de corrigir as distorções provocadas pelas mudanças em orientação, tamanho e formato das imagens 3D (terceira dimensão) capturadas da câmera do kinect. Portanto, o início do algoritmo proposto começa com a execução das principais transformações geométricas (escala, rotação e translação) aplicadas nas coordenadas FCPs, e depois submetidas a critérios que interpretam os seus deslocamentos geométricos. Esses deslocamentos são as ações faciais que submetidas a regras definidas das FACS (EKMAN *et al.*, 2002), resultam nas suas UAs específicas.

Na translação, os 38 pontos dos FCPs utilizados têm suas coordenadas alteradas para que se orientem ao valor de um **ponto de origem** abordado. O ponto de origem utilizado nesse trabalho foi a média entre os centros dos olhos, como pode ser visto Figura 42 (b). Antes da translação, esses pontos têm sua origem no ponto mais extremo do canto superior esquerdo, como pode ser visto na Figura 42 (a).

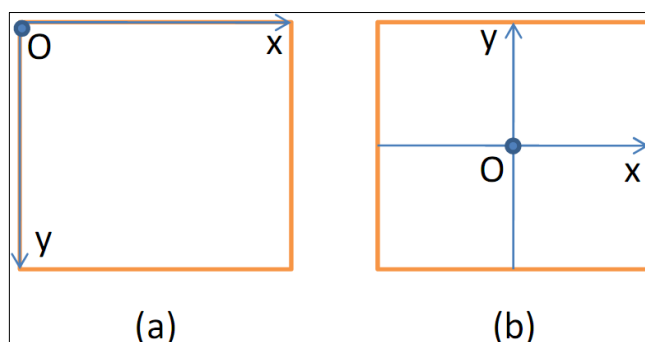


Figura 42 – (a) Sem translação. (b) Com translação. Fonte própria.

Na rotação dos pontos para um ângulo alinhado ao horizonte se toma como base a inclinação dos dois pontos do centro dos olhos. No Apêndice D, é mostrado o código que

realiza a translação e a rotação. A escala é transformada pela divisão dos FCPs pelo **valor base**. O valor base escolhido neste trabalho foi a distância de Manhattan (JIANGSHENG, 2002) entre os dois pontos do centro dos olhos. Realizando essa divisão, são compensadas as variações que podem ocorrer na distância entre a câmera e a face de um indivíduo, mantendo a mesma proporção entre as possíveis variações de distância entre pontos.

Com os pontos FCPs transformados, são realizados os cálculos dos deslocamentos de deformações geométricas nas expressões faciais, usando a Tabela 8 proposta neste trabalho. Por exemplo, o deslocamento D2 da Tabela 8 calcula a média do deslocamento vertical da parte interna das sobrancelhas esquerda e direita até o canto interior dos olhos. No algoritmo proposto, de acordo com análises das expressões faciais, foram projetados e utilizados vinte e um deslocamentos (Tabela 8) para ser mapeado a sua respectiva UAs com intuito de reconhecer as emoções com a melhor taxa de precisão.

Tabela 8 - Deslocamentos das deformações geométricas nas expressões faciais.

Distância	Deformação	Equação
D2	Média vertical da parte interna das sobrancelhas esquerda e direita até o canto interior dos olhos.	$(\text{abs}(\text{pontos}[17].y - \text{pontos}[23].y) + \text{abs}(\text{pontos}[50].y - \text{pontos}[56].y))/2$
D3	Deslocamento horizontal dos cantos esquerdo e direito do lábio até o centro do nariz.	$(\text{abs}(\text{pontos}[31].x - \text{pontos}[5].x) + \text{abs}(\text{pontos}[64].x - \text{pontos}[5].x))/2$
D4	Abertura vertical da boca.	$\text{abs}(\text{pontos}[7].y - \text{pontos}[8].y)$
D5	Deslocamento vertical entre os cantos dos lábios e os olhos.	$\text{abs}(\text{pontos}[31].y - (\text{pontos}[20].y + \text{pontos}[23].y)/2) + \text{abs}(\text{pontos}[64].y - (\text{pontos}[53].y + \text{pontos}[56].y)/2))/2$
D6	Deslocamento vertical entre os cantos externos das sobrancelhas e os cantos externos dos olhos	$\text{abs}(\text{pontos}[15].y - \text{pontos}[20].y) + \text{abs}(\text{pontos}[48].y - \text{pontos}[53].y))/2$
D8	Deslocamento vertical das pálpebras inferiores em relação aos centros dos olhos.	$\text{abs}(\text{pontos}[24].y - (\text{pontos}[20].y + \text{pontos}[23].y)/2) + \text{abs}(\text{pontos}[57].y - (\text{pontos}[53].y + \text{pontos}[56].y)/2))/2$
D9	Deslocamento vertical entre os cantos da boca e o centro do nariz.	$\text{abs}(\text{pontos}[31].y - \text{pontos}[5].y) + \text{abs}(\text{pontos}[64].y - \text{pontos}[5].y))/2$
D10	Deslocamento horizontal entre os cantos dos lábios.	$\text{abs}(\text{pontos}[31].x - \text{pontos}[64].x)$
D11	Deslocamento vertical entre o centro externo do lábio superior e o centro do nariz.	$\text{abs}(\text{pontos}[7].y - \text{pontos}[6].y)$
D12	Deslocamento vertical entre o centro externo do lábio inferior e o centro do nariz	$\text{abs}(\text{pontos}[8].y - \text{pontos}[6].y)$
D13	Deslocamento vertical entre os pontos imediatamente acima da narina para o canto interno do olho.	$(\text{abs}(\text{pontos}[17].y - \text{pontos}[25].y) + \text{abs}(\text{pontos}[50].y - \text{pontos}[58].y))/2$
D14	Abertura vertical dos olhos	$(\text{abs}(\text{pontos}[19].y - \text{pontos}[24].y) + \text{abs}(\text{pontos}[52].y - \text{pontos}[57].y))/2$

D15	Deslocamento vertical entre o canto externo da narina e canto interno da sobrancelha.	$(\text{abs}(\text{pontos}[26].y - \text{pontos}[17].y) + \text{abs}(\text{pontos}[50].y - \text{pontos}[59].y))/2$
D16	Deslocamento horizontal entre os cantos internos das sobrancelhas.	$\text{abs}(\text{pontos}[17].x - \text{pontos}[50].x)$
D17	Deslocamento vertical entre o centro da pálpebra superior e o centro superior da sobrancelha.	$(\text{abs}(\text{pontos}[19].y - \text{pontos}[49].y) + \text{abs}(\text{pontos}[52].y - \text{pontos}[16].y))/2$
D18	Deslocamento vertical entre o canto externo da sobrancelha e o canto interno do olho.	$(\text{abs}(\text{pontos}[15].y - \text{pontos}[23].y) + \text{abs}(\text{pontos}[48].y - \text{pontos}[56].y))/2$
D19	Deslocamento vertical entre o centro do olho e o centro inferior da sobrancelha.	$(\text{abs}(\text{pontos}[18].y - (\text{pontos}[20].y + \text{pontos}[23].y)/2) + \text{abs}(\text{pontos}[51].y - (\text{pontos}[56].y + \text{pontos}[53].y)/2))/2$
D20	Deslocamento vertical entre o centro do olho e o centro do lábio externo superior.	$\text{abs}(\text{pontos}[7].y - (\text{pontos}[20].y + \text{pontos}[23].y)/2) + \text{abs}(\text{pontos}[7].y - (\text{pontos}[56].y + \text{pontos}[53].y)/2))/2$
D21	Deslocamento vertical entre o canto interno do olho e o canto externo da boca.	$(\text{abs}(\text{pontos}[23].y - \text{pontos}[31].y) + \text{abs}(\text{pontos}[56].y - \text{pontos}[64].y))/2$
D23	Deslocamento vertical entre o centro do lábio externo inferior e o centro dos cantos internos das narinas.	$\text{abs}(\text{pontos}[8].y - \text{pontos}[6].y)$
D24	Deslocamento vertical entre o centro do olho e a bochecha.	$(\text{abs}(\text{pontos}[90].y - \text{pontos}[20].y) + \text{abs}(\text{pontos}[91].y - \text{pontos}[53].y))/2$

Depois de calculados os deslocamentos de deformações geométricas da face, são executados os mapeamentos desses deslocamentos às suas respectivas UAs, utilizando a Tabela 9, na qual estão os critérios que definem as UAs. Esta Tabela 9 é composta pela associação das combinações das variações dos deslocamentos de deformações faciais e novas regras definidas pelo autor do trabalho as quais foram projetadas analisando o FACS proposto pelos psicólogos Ekman *et al.* (2002).

O processo compara os deslocamentos de deformação (Tabela 8) de uma face realizando expressão facial com deslocamentos de deformação da mesma face sem expressão (neutra) para a identificação de UAs. Indicando se os deslocamentos de uma face, realizando expressão facial, aumentam ou diminuem em relação à face neutra. Por exemplo, para que uma expressão facial contenha a UA 1, é preciso que o deslocamento D2 de um indivíduo fazendo uma expressão facial seja maior (aumento) que o D2_n (deslocamento D2 com face neutra) do mesmo indivíduo que esteja com face neutra.

Tabela 9 – Critérios de Reconhecimentos das UAs.

UA	Critério de Reconhecimento
1	Aumento do deslocamento D2 em relação ao D2 _n da imagem neutra.
2	Aumento dos deslocamentos D6 e D18 em relação ao

	D6_n e D18_n da imagem neutra.
4	Diminuição dos deslocamentos D2, D16 e D17 em relação ao D2_n, D16_n e D17_n da imagem neutra.
5	Aumento do D14 e D1 em relação ao D14_n e D1_n da imagem neutra.
6	Diminuição dos deslocamentos D24 e D14 em relação ao D24_n e D14_n da imagem neutra.
7	Diminuição do deslocamento D8 em relação ao D8_n da imagem neutra. E aumento dos deslocamentos D14 e D19 em relação a D14_n e D9_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.
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.
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
16	Diminuição dos deslocamentos D4, D12 e D23 em relação ao D4_n, D12_n e D23_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.
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.
23	Diminuição do deslocamento D10 em relação ao D10_n da imagem neutra.
24	Aumento do deslocamento D11 em relação ao D11_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.
26	Aumento dos deslocamentos 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.
27	Aumento dos deslocamentos 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.

A subseção seguinte aborda o subsistema que mapeia a combinação das UAs encontradas na sua respectiva emoção.

4.2.1.3 INFERÊNCIA DAS EMOÇÕES

Depois de encontrar as UAs, a etapa de inferência das emoções é realizado pela conversão dos valores dessas UAs para as emoções encontradas a partir do subsistema de classificação. Para realizar esse processo, foram utilizados os estudos de Ekman *et al* (2002), resumidos na Tabela 10. Nesta Tabela 10 estão agrupadas as UAs que mais comumente são encontradas nas expressões faciais das emoções básicas (alegria, tristeza, desgosto, raiva, surpresa e medo). Tais UAs foram desenvolvidas por Ekman *et al* (2002) nas suas FACS. Cada AU codificada tem como base os envoltimentos dos músculos faciais. O Apêndice A descreve cada UA com seu respectivo músculo envolvido e uma imagem exemplificando a ação do músculo.

Tabela 10 – Associação das seis expressões faciais com suas UAs e combinações. Resumido de Ekman *et al* (2002).

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

A expressão facial é de fato a combinação de UA. De acordo com a Tabela 10, 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. 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 ela é ambígua para associar 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 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) combinando UA principal pertencente a uma expressão facial 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 abordagem proposta permite uma expressão facial ser uma combinação de UAs faciais relevantes, utilizando para isto regras de produção (cadeias de “ifs”), para realizar as classificações de UAs para as suas respectivas emoções. Ou seja, o algoritmo classifica a emoção se tiver a maior quantidade de UAs associadas à respectiva expressão facial analisada. A Figura 43 mostra o pseudocódigo da abordagem proposta baseada em regras que infere as emoções a partir da maior quantidade de UAs presentes.

```

Algoritmo "Encontrar as emoções"
1 - Var quantidadeEmocoes, quantidadeUAs, maximo, quantidadeEmpates, retorno: inteiro
2 - Var Emocoes, EmocaoAuxiliar, EmocoesUA, desempate, UA: Array
3 - CalculaAUs(distanciasAtuais3D, distanciasNeutras3D);

4-   for(i=0; i< quantidadeUAs; i++)
5-   // AUMENTA A PONTUAÇÃO DAS EMOÇÕES QUE TEM A MAIOR QUANTIDADE DE
   UAs
6-       // PRINCIPAIS
7-       for(k=0; EmocoesUA [i][0][k] != FIM; k++)
8-           // SE O UA FOR VERDADEIRO
9-           if(UA[i])
10-               Emocoes[EmocoesUA[i][0][k]]++;
11-
12-       // AUXILIARES
13-       for(k=0; EmocoesUA[i][1][k] != FIM; k++)
14-           // SE O AU FOR VERDADEIRO
15-           if(UA[i])
16-               EmocaoAuxiliar[emocoesAUs[i][1][k]]++;
17-
18-   // DESCOBRE A EMOCÃO QUE TIVER MAIOR QUANTIDADE DE UAs
19-   for(i=0; i< quantidadeEmocoes; i++)
20-       if(Emocoes[i] > Emocoes[maximo])
21-           maximo = i;
22-
23-   // CONTA A QUANTIDADE DE EMOCÕES EMPATADAS NA QUANTIDADE DE UAs
24-   quantidadeEmpates = 0;
25-   for(i=0; i< quantidadeEmocoes; i++)
26-       if(Emocoes[maximo] == Emocoes[i])
27-           desempate[quantidadeEmpates] = i;
28-           quantidadeEmpates++;
29-
30-   // SE HOUVE ALGUM EMPATE, DESEMPATA PELOS UAs AUXILIARES
31-   if(quantidadeEmpates > 0)
32-       // DESCOBRE O MÁXIMO
33-       maximo = desempate[0];
34-       for(i=0; i< quantidadeEmpates; i++)
35-           if(EmocaoAuxiliar [desempate[i]] > EmocaoAuxiliar [maximo])
36-               maximo = desempate[i];
37-
38-   // AQUI MANDA IMPRIMIR A EMOCÃO ENCONTRADA A PARTIR DO MÁXIMO DE
   UAs
39-   if (maximo == NEUTRO)
40-       imprime "NEUTRO"
41-   else if(maximo == TRISTEZA)
42-       imprime "TRISTEZA"
43-   else if(maximo == DESGOSTO)
44-       imprime "DESGOSTO"
45-   else if(maximo == RAIVA)
46-       imprime "RAIVA"
47-   else if(maximo == ALEGRIA)
48-       imprime "ALEGRIA"
49-   FIM

```

Figura 43 – Pseudocódigo do algoritmo de inferência de emoções. Fonte própria.

Atualmente, o módulo de reconhecimento de expressões faciais, proposto neste trabalho, só está inferindo as emoções alegria, tristeza, desgosto e raiva, pois o agente pedagógico do AVA de apoio à PBL (Fontes *et al.*, 2012), atualmente está projetado para entender somente essas emoções. Além disso, a versão atual do rastreamento das características da face da biblioteca do Kinect às vezes rastreia, em tempo real, os FCPs de forma não muito precisa, tornando os cálculos dos deslocamentos de deformações geométricas em alguns momentos imprecisos, principalmente os deslocamentos envolvidos para os reconhecimentos das emoções não utilizadas neste trabalho (surpresa e medo). Ficando essas emoções para trabalhos futuros.

Na próxima subseção são apresentados os materiais utilizados, experimentos e resultados da aplicação proposta.

4.2.2 EXPERIMENTOS E RESULTADOS

Esta seção apresenta os resultados experimentais que demonstram a precisão da abordagem proposta de reconhecimento de expressões faciais baseado em regras. Para a captura das imagens foi utilizado o hardware Kinect (visto na Seção 2.3.4). Na utilização dos métodos de detecção de faces (Viola-Jones) e extração das características (detecção dos FCPs pelo CANDIDE-3) da etapa do rastreamento das características foram utilizadas as implementações padrão da biblioteca do Kinect (MICROSOFT RESEARCH, 2011).

O ambiente integrado de desenvolvimento do sistema utilizado foi a IDE *Microsoft Visual Studio Express 2010*, utilizando 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 banco de dados escolhido para armazenar as emoções reconhecidas foi o MySQL 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. O código de reconhecimento da emoção foi feito em C++ e parte se encontra no Apêndice E.

Foram realizados experimentos durante o desenvolvimento e na fase de testes. Durante o desenvolvimento foram utilizadas pequenas amostras de dados, que 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 com imagens de 10 pessoas em tempo real. 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 dessas 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 com idades na faixa de 16 a 45 anos. A Figura 44 exemplifica basicamente cada emoção feita por essas pessoas. Nestes testes, o rastreamento das características faciais foi realizado de modo automatizado, utilizando a técnicas de Viola-Jones para detectar a face e o método CANDIDE-3 para a detecção dos pontos característicos faciais (FCPs), ambos disponíveis na biblioteca do Kinect. Depois de rastrear esses FCPs, foram calculados os deslocamentos geométricos das emoções (Tabela 8). Em seguida foi utilizado o algoritmo proposto baseado em regras (Apêndice E) para encontrar as UAs de acordo com as variações dos deslocamentos geométricos (Tabela 9) para, em fim, inferir as emoções relacionadas com suas específicas combinações de UAs (Tabela 10).

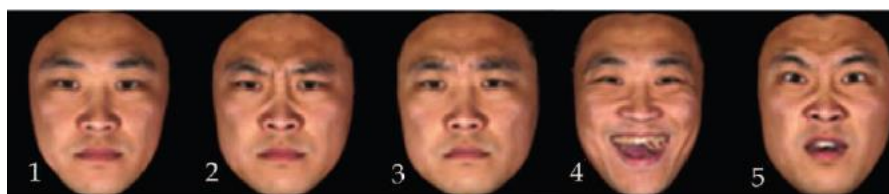


Figura 44 - Expressão Facial para cada emoção específica: 1-Neutra, 2- Raiva, 3-Tristeza, 4- Alegria, 5-Desgosto. Adaptado de YIN *et al* (2006).

A Tabela 11 mostra que o resultado total da taxa de acurácia do algoritmo de reconhecimento de expressões faciais proposto neste trabalho o qual resulta em 85,5%. Esse desempenho total da classificação emocional, vindas das expressões faciais, foi testado no conjunto de 200 imagens da face de 10 indivíduos. Essas imagens são usadas para validar as regras da Tabela 9 da Subseção 4.2.1.

O desempenho da taxa de precisão de cada emoção resultante do algoritmo de reconhecimento de emoção proposto é mostrado na Tabela 11, 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 intuitivo, 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, logo não funciona tão bem. 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. Esse problema também aconteceu com a emoção tristeza, a qual foi confundida com as emoções raiva e neutra. Esses problemas podem ser vistos na Tabela 12, que mostra a matriz de confusão resultante do algoritmo de reconhecimento de expressões faciais baseado em regra proposto neste trabalho.

Tabela 11 – Resultado da precisão das emoções

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

A matriz de confusão mostrada na Tabela 12 apresenta os resultados obtidos durante a utilização da máscara de CANDIDE-3 e o novo algoritmo de reconhecimento baseado em regras proposto. Como pode ser visto, a expressão facial mais ambígua foi desgosto, 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 raiva e alegria, com uma taxa de acerto de classificação similar. No entanto, os resultados das emoções, no geral, apresentam-se satisfatórios 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.*, 2005). Além disso, segundo Sebe *et al* (2005), 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%.

Tabela 12 – 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

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 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 de 10 indivíduos, atingindo um reconhecimento total de 85,5%.

No próximo capítulo são apresentadas as conclusões e discutidos os trabalhos futuros.

5 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Neste trabalho, são tratados problemas para o reconhecimento facial na identificação dos alunos e problemas para o reconhecimento de expressões faciais na inferência de emoções aplicadas ao contexto de um AVA.

No módulo de reconhecimento facial, a abordagem, utilizando Eigenfaces e PCA, provou ter capacidade de fornecer as características significativas e reduzir o tamanho de entrada para os algoritmos de classificação, melhorando o reconhecimento da face para identificação do aluno. O algoritmo de detecção de faces, portanto, mostrou-se eficiente tanto na taxa de acertos quanto no tempo de processamento. No entanto, esta abordagem mostrou-se ser sensível em imagens com condições de iluminação não controladas. Como contribuição, este trabalho apresentou uma alternativa biométrica aos métodos atuais de autenticação de indivíduos em um AVA, normalmente baseada em senhas. Concluimos, a

partir dos resultados desse trabalho, que a autenticação biométrica pode ser usada para garantir a identidade do aluno que está acessando um AVA remotamente.

Experimentalmente, verificamos que esta abordagem de reconhecimento facial, utilizando PCA, *Eigenfaces* e a técnica de classificação K-NN, apresentou bons resultados. O melhor resultado obtido de testes foi atingido com a técnica K-NN, em imagens de dimensões 12x9 e utilizando números de 20 características resultantes do PCA, resultando em uma taxa de reconhecimento de aproximadamente 92%. Como trabalhos futuros, pretende-se utilizar a abordagem de PCA e *Eigenfaces* em uma análise de algoritmos de aprendizado, tais como Redes Neurais e Redes *Bayesianas*. Além disso, paralelamente, a abordagem proposta neste trabalho para o reconhecimento facial será aplicada em bases de dados existentes na literatura para comparar os resultados vindos dessas bases de dados com os resultados alcançados por este trabalho.

No módulo de reconhecimento de expressões faciais, a abordagem utilizando o rastreamento de características faciais por meio das técnicas Viola-Jones e CANDIDE-3 provou ter capacidade de fornecer as características significativas e reduzir o tempo de processamento na classificação das emoções. Além disso, a abordagem 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 baseada em regras, a qual se mostrou eficaz comprovado pelos resultados dos experimentos. Além disso, a contribuição deste trabalho poderá dar algumas referências a pesquisadores acadêmicos.

Como trabalho futuro, com base nas características faciais encontradas, outros métodos de classificação de emoção poderão ser usados. Além disso, Serão feitos testes no sistema de reconhecimento de expressão facial proposto, utilizando bases de dados existentes na literatura a fim de comparar os resultados vindos dessas bases de dados com os resultados alcançados por este trabalho. Também serão estudados outros módulos de reconhecimento de emoções, tais como recursos de áudio e gestos. 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.

REFERÊNCIAS

- ABDAT, F.; MAAOUI, C.; PRUSKI, A. Human-Computer Interaction Using Emotion Recognition from Facial Expression. **Computer Modeling and Simulation (EMS)**. Fifth UKSim European Symposium on, vol., no., pp.196-201, 2011.
- AGARWAL, M.; AGRAWAL, H.; JAIN, N.; KUMAR, M. Face Recognition Using Principle Component Analysis, Eigenface and Neural Network. **International Conference on Signal Acquisition and Processing, ICSAP 10**, vol., no., pp. 310-314, 9-10, 2010.
- AGULLA, et al. Is My Student at the Other Side? Applying Biometric Web Authentication to E-Learning Environments. **8th IEEE International Conference on Advanced Learning Technologies (ICALT)**, p. 551-553, 2008.
- AHLBERG, J. **CANDIDE-3 -- an updated parameterized face**. Report No. LiTH-ISY-R-2326, Dept. of Electrical Engineering, Linköping University, Sweden, 2001.
- BERCHT, M.; VICARI, R. M. Afetividade em Máquina: Uma possibilidade? **Workshop Informática na Educação, INF/UFRGS**, 2004.
- BAGHERIAN, E.; RAHMAT, R. W. O. K. **Facial feature extraction for face recognition: a review**. Information Technology, 2008.
- BERCHT, M. **Em Direção a Agentes Pedagógicos com Dimensões Afetivas**. Tese (Doutorado em Ciência da Computação). Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2001.
- BESCHE-RICHARD, C.; BUNGENER, C. **Psicopatologias, emoções e neurociências**. Lisboa: Climepsi Editores, 2008.
- BLACK, M. J.; YACOOB, Y. Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion, *Int'l J. Computer Vision*, vol. 25, no. 1, pp. 23-48, 1997.
- BOCCA, E.; JAQUES, P. A.; VICCARI, R. W. Modelagem e Implementação da Interface para Apresentação de Comportamentos Animados e Emotivos de um Agente Pedagógico Animado. **RENOTE - Revista Novas Tecnologias na Educação** (ISSN 1679-1916), 2003.
- BOUREL, F.; CHIBELUSHI, C. C.; LOW A. A. Robust Facial Expression Recognition Using a State-Based Model of Spatially-Localised Facial Dynamics. **Proc. Fifth IEEE Int. Conf. Automatic Face and Gesture Recognition**, pp. 106-111, 2002.
- BRADSKY, G. R.; PISAREVSKY, V.; BOUGUET, J. **Learning OpenCV: Computer Vision with the OpenCV Library**. Springer, Estados Unidos, 2006.
- BREIMAN, L. Random Forest. **In Journal of Machine Learning**, Vol.45, pages 5-32. Kluwer Academic, Netherland, 2001.
- BUXTON B. **Entrevista CES 2010: NUI with Bill Buxton**, 2010. Disponível em:

<<http://channel9.msdn.com/posts/LarryLarsen/CES-2010-NUI-with-Bill-Buxton>> Acesso em: 27 Agosto 2011.

CANTONI, V.; CELLARIO, M.; PORTA, M. Perspectives And Challenges In E-Learning: Towards Natural Interaction Paradigms. **Journal of Visual Languages and Computing**, n. 15, p. 333-345. 2003.

CARLSON, N. R. **Fisiologia do comportamento**. 7.ed. Barueri-SP: Manole, 2002.

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.

CHEON, Y.; KIM, D. A Natural Facial Expression Recognition Using Differential-AAM and k-NNS. **Multimedia, Tenth IEEE International Symposium on**, vol., no., pp.220-227, 15-17, 2008.

CHIU, K.; RASKAR, R.; Computer vision on tap, Computer Vision and Pattern Recognition Workshops. **CVPR Workshops 2009. IEEE Computer Society Conference on**, vol., no., pp.31-38, 20-25, 2009.

CLEARY, J.G.; TRIGG, L.E. K*: an instance-based learner using an entropic distance measure. In: **INTERNATIONAL CONFERENCE ON MACHINE LEARNING**, 12., 1995, Tahoe City, California, US . Proceeding. Tahoe City, 1995. p.108-114.

COHN, J.F.; ZLOCHOWER, A.J.; LIEN, J.J.; KANADE, T. Feature-Point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression. **Proc. Int'l Conf. Automatic Face and Gesture Recognition**, pp. 396-401, 1998.

CRUZ, L.; LUCIO, D.; VELHOZ, L. 2012. **Kinect and RGBD Images: Challenges and Applications**. Disponível em: <http://www.visgraf.impa.br/Data/RefBib/PS_PDF/kinect12/kinect-survey.pdf>. Acesso em: 20 Novembro 2012.

DEWI, D.A.; SANI, Z.H.B.A.; JEREMIAH, P. A computational system approach to develop student's based emotion system: An alternative feedback tool for lecturers to enhance teaching and learning. **Software Engineering (MySEC), 2011 5th Malaysian Conference in**, vol., no., pp.342-348, 2011.

DONATO, G.; BARTLETT, M. S.; HAGER, J. C.; EKMAN, P.; SEJNOWSKI, T.J. Classifying Facial Actions, **IEEE Trans. Pattern Analysis and Machine Intelligence**, Vol. 21, No. 10, pp. 974-989, 1999.

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.

EDWARDS, G.J.; COOTES, T. F.; TAYLOR, C. J. Face Recognition Using Active Appearance Models. **Proc. European Conf. Computer Vision**, vol. 2, pp. 581-695, 1998.

EKMAN, P. **Facial Expression and Emotion**. American Psychologist, v.48, n.4, p. 384-392, 1993.

EKMAN, P. Darwin, deception, and facial expression. **In: Annals New York Academy of sciences**, 1000. [S.l.: s.n.], p. 205–221, 2003.

EKMAN, P. **Emotion in the Human Face**. Cambridge University Press, 1982

EKMAN, P. Facial Expressions. **In: DALGLEISH, T.; POWER, T. (Eds.). The Handbook of Cognition and Emotion**. Sussex. Reino Unido: John Wiley & Sons Ltd. p.301-320, 1999.

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.

EMFACS. **EMFACS: Scoring for Emotion with FACS**. Disponível em: <<http://www.face-andemotion.com/dataface/facs/emfacs.jsp>>. Acesso em: 12 Novembro 2011.

EPHRAIM, T.; HIMMELMAN, T.; SIDDIQI, K. Real-Time Viola-Jones Face Detection in a Web Browser," **Computer and Robot Vision, 2009. CRV '09. Canadian Conference on** , vol., no., pp.321,328, 25-27, 2009.

ESCADA, A. **Emoções**. Porto: Raridade Editora, 2004

ESSA, I. A.; PENTLAND, A. P. Coding, Analysis, Interpretation, and Recognition of Facial Expressions. **IEEE Trans. Pattern Analysis and Machine Intelligence**, Vol. 19, No. 7, pp. 757-763, 1997.

FACS. **Description of Facial Action Coding System (FACS)**, 2007. Disponível em: <<http://www.face-andemotion.com/dataface/facs/description.jsp>>. Acesso em: 12 Novembro 2011.

FALCÃO, A. X.; LEITE, N. J. **Fundamentos de processamento de imagem digital**. Disponível em: <<http://www.dcc.unicamp.br/~cpg/material-didatico/mo815/9802/curso.pdf.gz>>. Acesso em: 18 Agosto 2009.

FASEL, B.; LUETTIN, J. **Automatic facial expression analysis: a survey**. Pattern Recognition, v. 36, n. 1, p. 259–275, 2003.

FIGLIARESE, M.; TAROUCO, L. M. R. Uma Proposta de Autenticação de Usuários para Ensino a Distância. **18º Simpósio Brasileiro de Redes de Computadores**. Instituto de Informática da UFRGS. Porto Alegre - RS- Brasil, 2006.

FIPA. Welcome to the Foundation for Intelligent Physical Agents. **Site Oficial do Padrão FIPA**, 2011. Disponível em: <<http://www.fipa.org/>>. Acesso em: 08 Dezembro 2011.

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.

FRANÇA, A. B.; SOARES, J. M. Sistema de apoio a atividades de laboratório de programação via Moodle com suporte ao balanceamento de carga, **in Anais do Simpósio Brasileiro de Informática na Educação (SBIE)**, Aracaju, 2011.

FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. **The WEKA data mining software: an update.** SIGKDD Explorations, v. 11, n. 1, University Mark Hall, 1999.

FRIED, L. **DIY Kinect Hacking**, 2011. Disponível em: <<http://www.ladyada.net/learn/diykinect/>>. Acesso em: 29 Agosto 2011.

FROZZA, R.; SILVA, A.; LUX, B.; CRUZ, M.; BORIN, M. Dóris 3D: Agente Pedagógico baseado em Emoções. **In: XX Simpósio Brasileiro de Informática na Educação, 2009, Florianópolis. XX Simpósio Brasileiro de Informática na Educação.** Florianópolis: Universidade Federal de Santa Catarina, 2009.

GHOSAL, V.; TIKMANI, P.; GUPTA, P. Face Classification Using Gabor Wavelets and Random Forest. **In: Proceedings of the Canadian Conference on Computer and Robot Vision (CRV '09), IEEE Computer Society.** Washington, DC, USA, 68-73, 2009.

GIRAFFA, L. M. M. **Uma Arquitetura de Tutor Utilizando Estados Mentais.** Porto Alegre: CPGCC/UFRGS, Tese de Doutorado, 1999.

GONZALEZ, R.; WOODS, R. "**Digital Image Processing**", 2nd Ed. Prentice Hall, 2002.

GROSS, R.; MATTEWS, I.; COHN, J.; KANADE, T.; BAKER, S. Multi-PIE Automatic Face & Gesture Recognition, **FG '08. 8th IEEE International Conference on**, vol., no., pp.1-8, 17-19, 2008.

HERNÁNDEZ *et al.* **Biometrics In Online Assessments: A Study Case in High School Students.** 18th International Conference on Electronics, Communications and Computers (CONIELECOMP), p. 111-116, 2008.

INTEL. **OpenCV: Open source Computer Vision Library**, 2007. <<http://www.intel.com/technology/computing/opencv/index.htm>>. Acesso em: 15 Novembro 2011.

JADE. **Java development framework: an open-source platform for peer-to-peer agent based applications**, JADE (2012). Disponível em: <<http://jade.tilab.com/>>. Acesso em: 24 Julho 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.

- JAQUES, P. A.; VICCARI, R. M. Considering Students's emotions in computer-mediated learning environments. **In: Zongmin Ma (Org). Web-based intelligent e-learning systems: Technologies and applications.** Hershey, 2004.
- JAQUES, P. A.; VICARRI, R. M. Estado da Arte em Ambientes Inteligentes de Aprendizagem que Consideram a Afetividade do Aluno. **Revista Informática na Educação: Teoria & Prática**, v.8, n.1. 2005.
- JIANGSHENG, Y. **Method of k-Nearest Neighbors.** Institute of Computational Linguistics, Peking University, China, 2002.
- JIAN H.; ZILING S.; RUOMEI W. 3D Face Reconstruction Based on Improved CANDIDE-3 Model. **Digital Home (ICDH), 2012 Fourth International Conference on** , vol., no., pp.438,442, 23-25, 2012.
- KASHEM, M. A. et al. Face Recognition System Based on Principal Component Analysis (PCA) with Back Propagation Neural Networks (BPNN). **Canadian Journal On Image Processing And Computer Vision**, Canadá, n., p.36-45, 04 abr. 2011. 2.
- KELTNER, D.; EKMAN, P. **Emotion: An Overview.** Encyclopedia of Psychology, 162-166, 2002.
- KOTSIA, I.; NIKOLAIDIS, N.; PITAS, I. Facial expression recognition in videos using a novel multi-class support vector machines variant. **In: IEEE Trans. Image Process**, 2007.
- KSHIRSAGAR, V. P.; BAVISKAR, M. R.; GAIKWAD, M. E. Face recognition using Eigenfaces. **Computer Research and Development (ICCRD), 3rd International Conference on**, vol. 2, no., pp. 302-306, 11-13, 2011.
- KUMAR, S.; GANKOTIYA, A. K.; DUTTA, K. A Comparative Study of Moodle with other e-Learning Systems. **In: International Conference on Electronics Computer Technology - ICECT**, 3. Kanyakumari: IEEE, 2011.
- KWANG-EUN, K.; KWEE-BO, S. Development of a Facial Emotion Recognition Method Based on Combining AAM with DBN. **International Conference on Cyberworlds (CW)**, vol., no., pp.87-91, 20-22 Oct. 2010
- LEDERMAN, S.; KLATZKY, R.; ABRAMOWICZ, A.; SALSMAN, K.; KITADA, R.; HAMILTON, C. **Haptic Recognition of Static and Dynamic Expressions of Emotion in the Live Faces.** Psychological Science, 18 (2), 158-164, 2007.
- LIEN, J. J.; KANADE, T.; COHN, J. F.; Li, C-C. Automated Facial Expression Recognition Based on FACS Action Units. **Proc. Third IEEE Int. Conf. Automatic Face and Gesture Recognition**, pp. 390-395, 1998.
- LIENHART, R.; MAYDT, J. An extended set of haar-like features for rapid object detection. **In: ICIP (1)**. [S.l.: s.n.], 2002. p. 900-903.

LONGHI, M. et al. Um estudo sobre os Fenômenos Afetivos e Cognitivos em Interfaces para Softwares Educativos. **In: Revista Novas Tecnologias na Educação - Universidade Federal do Rio Grande do Sul**, Porto Alegre, v. 5, n. 1, 2007.

MA, E. L. H. **Avaliação de Características Haar em Um Modelo de Detecção de Face**. 70 p. Monografia (Graduação) — Departamento de Ciência da Computação, Universidade de Brasília, Brasília, 2007.

MANRIQUE, ANA LÚCIA. **A afetividade manifestada por professores participantes de um processo de formação em geometria**, PUC/SP, 2004.

MARAIS, E.; ARGLES D.; VON SOLMS B. Security issues specific to e-assessments, **8th Annual Conference on WWW Applications**, 2006.

MATTHEWS, I.; BAKER, S. Active Appearance Models Revisited. **Int. J. Comput. Vision**, vol. 60, no. 2, pp. 135-164, 2004.

MICROSOFT. **Introducing Kinect for XBOX 360**, 2010. Disponível em: <<http://www.xbox.com/en-US/Kinect>>. Acesso em: 10 Julho 2011.

MICROSOFT. **Kinect Sales Surpass Ten Million**, 2011. Disponível em: <<http://www.xbox.com/en-us/press/archive/2011/0308-ten-million-kinects>>. Acesso em: 10 Julho 2011.

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.

MISLAV, G.; KRESIMIR, D.; SONJA, G. **SCface --- surveillance cameras face database**. *Multimedia Tools Appl.* 51, 3 (February 2011), 863-879, 2011.

MÓRA, M. C. et al. BDI models and systems: reducing the gap. **In: Proceedings do Agents Theory, Architecture and Languages Workshop**. Canarias. London: Springer-Verlag, 1998.

OLIVEIRA, E.; JAQUES, P. A. Inferindo as emoções do usuário pela face através de um sistema psicológico de codificação facial. **In: Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais**. Porto Alegre. : SBC/ACM, p. 156-165, 2008.

PANTIC, M.; ROTHKRANTZ, L. J. M. An Expert System for Multiple Emotional Classification of Facial Expressions. **Proc. 11th IEEE Int. Conf. on Tools with Artificial Intelligence**, pp. 113-120, 1999.

PATIL, R. A.; SAHULA, V.; MANDAL, A. S.; Automatic recognition of facial expressions in image sequences: A review. **Industrial and Information Systems (ICIIS), 2010 International Conference on**, vol., no., pp.408-413, July 29 2010-Aug. 2010.

PENTEADO, B. E.; MARANA, A. N. A Video-Based Biometric Authentication for ELearning Web Applications. **In: Eleventh International Conference on Enterprise**

Information Systems, ICEIS 2008. Milão. LNBIP - Lecture Notes on Business Information Processing. Berlin : Springer-Verlag, v. 24. p. 770-779, 2009.

PICARD, R. W. **Affective Computing.** M.I.T Media Laboratory Perceptual Computing Section Technical Report, n.321, 1995.

PICARD, R. W. **Affective Computing.** Cambridge, EUA: The M.I.T. Press, 1997.

PICARD, R. W. **Affective Computing.** MIT Press, Cambridge, 1998

PICARD, R.W. Affective Computing. MIT Press, Cambridge, EUA, 1997. BERCHT, Magda; VICARI, Rosa Maria. Afetividade em Máquina: Uma possibilidade? **Workshop Informática na Educação**, INF/UFRGS, 2004.

PRIMESENSE. **Site da empresa Primesense**, 2011. Disponível em: <<http://www.primesense.com/>>. Acesso em: 13 Setembro 2011.

QUARTO, C. C. **Inferindo Fatores Sócio-Afetivos em Ambientes de Ensino-Aprendizagem Colaborativos Assistidos por Computador.** Dissertação de Mestrado. Departamento de Pós-Graduação em Engenharia de Eletricidade. Universidade Federal do Maranhão, MA, 2006.

RABUZIN, K.; BACA, M.; SAJKO, M. **E-learning: Biometrics as a Security Factor.** International Multi-Conference on Computing in the Global Information Technology , 6., 2006. p. 64.

ROLIM, A. L.; BEZERRA, E. P. **Um Sistema De Identificação Automática De Faces Para Um Ambiente Virtual De Ensino E Aprendizagem.** In: XIV Simpósio Brasileiro de Sistemas Multimídia e Web, Vila Velha. 2008.

ROSARIO, S. V. P. G. **Facial Emoticons: Reprodução de Informação Associada a Expressões Faciais por Via do seu Reconhecimento.** Dissertação — Instituto Superior Técnico, Universidade Técnica de Lisboa, Portugal, 100p, 2008.

ROSERBERG, E.L.; EKMAN, P. **Emotion: Methods of Study.** Encyclopedia of Psychology, 171-175, 2002.

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.

SEGAL, M. R. **Machine learning benchmarks and random forest regression.** Disponível em: <http://www.ics.uci.edu/_mlearn/MLRepository.html>. Center for Bioinformatics & Molecular Biostatistics, 2004.

SHARMA, P.; REILLY, R.B. A colour face image database for benchmarking of automatic face detection algorithms, **Video/Image Processing and Multimedia Communications, 2003. 4th EURASIP Conference focused on** , vol.1, no., pp. 423- 428 vol.1, 2-5, 2003.

SILVA, M. G. C. N. **Um mecanismo de identificação fisionômica para ambientes virtuais de aprendizagem**. João Pessoa: UFPB, 2008. Dissertação (Mestrado) – Universidade Federal da Paraíba, João Pessoa.

STAN, Z. Li; ANIL, K. J. **Handbook of Face Recognition**. 2nd Edition, Springer, 2011.

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.

TIAN, Y-L.; KANADE, T.; COHN, J. F. Recognizing Action Units for Facial Expression Analysis. **IEEE Trans. Pattern Analysis and Machine Intelligence**, Vol. 23, No. 2, pp. 97-115, 2001.

TURK, M.; PENTLAND, A. Eigenfaces for recognition. **Journal of Cognitive Neuroscience**, 3(1):71{86, 1991.

VIOLA, P.; JONES, M. **Robust real-time object detection**. Technical report, University of Cambridge, 2001.

WEKA (software). **Data Mining Software in Java The University of Waikato**. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/>>, Acesso em: 05 de nov. de 2011.

WHITEHILL, J.; BARTLETT, M.; MOVELLAN, J. **Measuring the Perceived Difficulty of a Lecture Using Automatic Facial Expression Recognition - MPLab TR**, University of California, San Diego 2008 – Disponível em: <<http://www.physorg.com/news133590374.html>>. Acesso em: 09 Outubro 2011.

YANG, M.; KRIEGMAN, D. J.; AHUJA, N. Detecting Faces in Images: A Survey. **IEEE Transactions On Pattern Analysis And Machine Intelligence**, v.24, n.1, 2002.

YIN, L.; WEI, X.; SUN, Y.; WANG, J.; ROSATO, M. A 3d facial expression database for facial behavior research. **In Proceedings of International Conference on FGR**, pp. 211-216, UK, 2006.

ZAINAL, A.; MAAROF, M.A.; SHAMSUDDIN, S.M.; ABRAHAM, A. Ensemble of One-class Classifiers for Network Intrusion Detection System. In: **Proceedings of Fourth International Conference on Information Assurance and Security**, pages 180-185, 2008.








ZHANG, B.; RUAN, Q. Facial feature extraction using improved deformable templates. **Signal Processing, 8th International Conference on**, vol.4, no., 16-20, 2006.










ZHOU SHENG; LIN ZHU-YING; DONG WAN-XIN. The model of E-learning based on affective computing. **Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on**, vol.3, no., pp.V3-269-V3-272, 20-22 Aug. 2010.




APÊNDICE A – UAs do sistema FACS

A Tabela 13 apresenta o mapeamento de algumas UAs com seus respectivos conjuntos musculares envolvidos, descrevendo o que cada músculo faz, e apresenta uma imagem exemplificando a ação do músculo da respectiva UA.

Tabela 13 – Exemplos de UAs do sistema FACS (Ekman *et al.*, 2002).

AU	Descrição	Músculo Facial	Exemplo de imagem
1	Levantador de Sobrancelha Interna	<i>Frontalis, pars medialis</i>	
2	Levantador de Sobrancelha externa	<i>Frontalis, pars lateralis</i>	
4	Abaixador de Sobrancelha	<i>Corrugator supercilii, Depressor supercilii</i>	
5	Levantador de Pálpebra Superior	<i>Levator palpebrae superiores</i>	
6	Levantador de Bochechas	<i>Orbicularis oculi, pars orbitalis</i>	
7	Apertador de Pálpebra	<i>Orbicularis oculi, pars palpebralis</i>	
9	Enrugador de Nariz	<i>Levator labii superioris alaquae nasi</i>	

10	Levantador de Lábio Superior	<i>Levator labii superiores</i>	
12	Puxador de Canto do Lábio	<i>Zygomaticus major</i>	
15	Depressor de Canto do Lábio	<i>Depressor anguli oris (a.k.a. Triangularis)</i>	
16	Depressor do Lábio Inferior	<i>Depressor labii inferioris</i>	
17	Levantador de Queixo	<i>Mentalis</i>	
20	Esticador de Lábio	<i>Risorius w/ platysma</i>	
23	Apertador do Lábio	<i>Depressor labii inferioris or relaxation of Mentalis, or Orbicularis oris</i>	
24	Pressionador de Lábio	<i>Orbicularis oris</i>	
25	Separador de Lábios	<i>Depressor labii inferioris or relaxation of Mentalis, or Orbicularis oris</i>	

26	Boca Caída	<i>Masseter, relaxed Temporalis and internal Pterygoid</i>	
27	Boca Alongada	<i>Pterygoids, Digastric</i>	
28	Sucção de Lábios	<i>Orbicularis oris</i>	

APÊNDICE B – CÓDIGO *ADABOOST* E O ALGORITMO DE TREINAMENTO

Algoritmo *AdaBoost* modificado para os classificadores Viola e Jones (Figura 45) e em seguida o algoritmo de Treinamento (Figura 46), abordados no trabalho de Rolim (2008).

- Dados exemplos de Imagens $(x_1, y_1), \dots, (x_n, y_n)$ onde $x \in \mathbb{R}$ e $y_i = 0, 1$ para exemplos negativos e positivos respectivamente.
- Inicialize os pesos $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ para $y_i = 0, 1$ respectivamente, onde m e l são números negativos e positivos respectivamente.
- Para $t = 1, \dots, T$:
 1. Normalize os pesos,

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$$
 onde w_t é uma distribuição de probabilidade.
 2. Para cada característica j , treine um classificador h_j restrito ao uso de apenas uma única característica. O erro é avaliado levando em conta $w_i, e_j = \sum w_{t,i} |h_j(x_i) - y_i|$.
 3. Escolha um classificador, h_t , com o menor erro e_t .
 4. Atualize os pesos

$$w_{t+1} = w_{t,i} \beta_t^{1-e_t}$$

onde $e_i = 0$ se exemplo x_i for classificado corretamente,
 e_i se não for, e $\beta_t = \frac{e_t}{1-e_t}$
 5. O classificador final forte é definido por:

$$h(x) = \begin{cases} 1 & \text{Se } \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{Caso Contrário} \end{cases}, \text{ onde } \alpha_t = \log \frac{1}{\beta_t}$$

Figura 45 – Algoritmo *AdaBoost* (ROLIM, 2008).

- Selecione valores para:
 1. f : o máximo valor de aceitação para a taxa de falso positivo por camada
 2. d : o mínimo valor de aceitação da taxa de detecção por camada.
- Selecione o valor total da taxa de falso positivo, F_{alvo} .
- P = conjunto das amostras positivas.
- N = conjunto das amostras negativas.
- $F_o = 1.0$; $D_o = 1.0$; $i = 0$
- Enquanto $F_i > F_{alvo}$
 - $i=i+1$;
 - $n_i = 0$; $F_i = F_{i-1}$
 - Enquanto $F_i > f \times F_{i-1}$
 - * $n_i = n_i + 1$
 - * Use P e N para treinar com n_i características pelo AdaBoost.
 - * Avaliar a atual cascata de classificador no conjunto de validação para determinar F_i e D_i .
 - * Diminuir o limiar do-ésimo classificador até que a cascata atual tenha uma taxa de detecção pelo menos $d \times D_{i-1}$ (Isso também afeta F_i).
 - $N=0$
 - Se $F_i > F_{alvo}$ então avalie o atual detector em um conjunto de imagens negativas e coloque qualquer detecção falsa no conjunto N .

Figura 46 – Algoritmo de treinamento (ROLIM, 2008).

APÊNDICE C – Vértices do CANDIDE-3 com suas descrições

A Tabela 14 lista todos os vértices do CANDIDE-3 com as suas respectivas descrições no momento do rastreamento das características faciais.

Tabela 14 – Vértices do CANDIDE-3 com suas descrições. Adaptado de Ahlberg (2001).

Vértices	Descrição
0	Topo do crânio.
1	Fronteira média entre o cabelo e a testa.
2	Meio da testa.
3	Ponto médio entre as sobrancelhas.
4	Não usado.
5	Ponta do nariz.
6	Extremidade inferior do meio do nariz.
7	Ponto médio do lábio superior externo.
8	Ponto médio do lábio inferior externo.
9	Ponto médio do queixo.
10	Parte inferior do queixo.
11	À esquerda do topo do crânio.
12	À esquerda do topo do crânio.
13	Borda esquerda entre o cabelo e a testa.
14	Lado esquerdo do crânio.
15	Canto externo da sobrancelha esquerda.
16	Ponto mais alto da sobrancelha esquerda.
17	Canto interno da sobrancelha esquerda.
18	Contorno inferior da sobrancelha esquerda, em linha reta abaixo do vértice 16.
19	Centro de pálpebra superior externa esquerda.
20	Canto externo do olho esquerdo.
21	Centro de pálpebra interior superior esquerda.
22	Centro da pálpebra inferior interior esquerda.
23	Canto interno do olho esquerdo.
24	Centro de pálpebra inferior e externa esquerda.
25	Borda do nariz esquerdo.
26	Borda narina esquerda exterior.
27	Osso da bochecha esquerda.
28	Ponto de contato interior entre orelha esquerda e o rosto.
29	Ponto de contato superior entre orelha esquerda e o rosto.
30	Canto esquerdo do osso da mandíbula.
31	Canto esquerdo do contorno do lábio exterior.
32	Canto esquerdo do queixo.
33	Ponto mais alto do contorno labial esquerda exterior.
34	Fronteira média entre o cabelo e a testa.
35	Não usado (idêntico ao 2).
36	Não usado (idêntico ao 3).
37	Não usado (idêntico ao 4).
38	Não usado (idêntico ao 5).
39	Não usado (idêntico ao 6).
40	Ponto médio do contorno do lábio interno inferior.

41	Não usado (idêntico ao 8).
42	Não usado (idêntico ao 9).
43	Não usado (idêntico ao 10).
44	Direita do topo do crânio.
45	Direita do topo do crânio.
46	Margem direita entre o cabelo e a testa.
47	Lado direito do crânio.
48	Canto externo da sobrancelha direita.
49	Ponto mais alto da sobrancelha direita.
50	Canto interno da sobrancelha direita.
51	Contorno inferior da sobrancelha direita, em linha reta abaixo do 49.
52	Centro da pálpebra exterior superior direita.
53	Canto externo do olho direito.
54	Centro de pálpebra interna superior direito.
55	Centro da pálpebra inferior direito interno.
56	Canto interno do olho direito.
57	Centro da pálpebra inferior exterior direita.
58	Borda direita do nariz.
59	Borda da narina direita.
60	Osso da bochecha direita.
61	Ponto inferior de contato entre a orelha direita e o rosto.
62	Ponto superior de contato entre orelha direita e o rosto.
63	Canto direito da mandíbula.
64	Canto direito do contorno do lábio exterior.
65	Canto direito do queixo.
66	Ponto mais alto do contorno dos lábios externa direita.
67	Íris esquerda, canto superior externo delimitador (quadrado) da caixa.
68	Íris esquerda, canto inferior externo delimitador (quadrado) da caixa.
69	Íris direita, canto superior externo delimitador (quadrado) da caixa.
70	Íris direita, canto inferior externo delimitador (quadrado) da caixa.
71	Íris esquerda, canto superior interno delimitador (quadrado) da caixa.
72	Íris esquerda, canto inferior interno delimitador (quadrado) da caixa.
73	Íris direita, canto superior interno delimitador (quadrado) da caixa.
74	Íris direita, canto inferior interno delimitador (quadrado) da caixa.
75	Lado esquerdo da ponta do nariz.
76	Lado direito da ponta do nariz.
77	Extremidade superior esquerda do osso do nariz.
78	Extremidade superior direita do osso do nariz.
79	Ponto médio entre 64 e 7 no contorno dos lábios superior externo.
80	Ponto médio entre 31 e 7 no contorno dos lábios inferior externo.
81	Ponto médio entre 89 e 87 no contorno dos lábios superior interno.
82	Ponto médio entre 88 e 87 no contorno dos lábios superior interno.
83	Ponto médio entre 89 e 40 no contorno dos lábios superior interno.
84	Ponto médio entre 88 e 40 no contorno dos lábios superior interno.
85	Ponto médio entre 64 e 8 no contorno dos lábios superior externo.
86	Ponto médio entre 31 e 8 no contorno dos lábios superior externo.
87	Ponto médio do contorno dos lábios interno superior.
88	Canto esquerdo do contorno dos lábios interior.
89	Canto direito do contorno dos lábios interior.
90	Centro da face esquerda.
91	Centro da face direita.
92	Extremidade inferior esquerda do osso do nariz.
93	Extremidade inferior direita do osso do nariz.

94	Extremidade inferior do meio do osso do nariz (ou galo do nariz).
95	Extremidade superior externa da pálpebra superior esquerda.
96	Extremidade superior externa da pálpebra superior direita.
97	Extremidade inferior externa da pálpebra superior esquerda.
98	Extremidade inferior externa da pálpebra superior direita.
99	Extremidade superior externo da pálpebra inferior esquerda.
100	Extremidade superior externo da pálpebra inferior direita.
101	Extremidade inferior externa da pálpebra inferior esquerda.
102	Extremidade inferior externa da pálpebra inferior direita.
103	Extremidade superior interna da pálpebra superior esquerda.
104	Extremidade superior interna da pálpebra superior direita.
105	Extremidade inferior interna da pálpebra superior esquerda.
106	Extremidade inferior interna da pálpebra superior direita.
107	Extremidade superior interna da pálpebra inferior esquerda.
108	Extremidade superior interna da pálpebra inferior direita.
109	Extremidade inferior interna da pálpebra inferior esquerda.
110	Extremidade inferior interna da pálpebra inferior direita.
111	Extremidade inferior esquerda do nariz.
112	Extremidade inferior direita do nariz.

APÊNDICE D – CÓDIGO DE TRANSFORMAÇÃO GEOMÉTRICA DE TRANSLAÇÃO E ROTAÇÃO

A Figura 47 mostra o código de transformação geométrica de translação e rotação aplicadas nas imagens 3D (terceira dimensão) capturas da câmera do kinect.

```

void SingleFace::aplicaTransformacao(FT_VECTOR3D *pt, FT_VECTOR3D *pt1)
{
    float modulo, Ax, Ay, Az, Bx, By, Bz, Mx, My, Mz, Qx, Qy, Qz, Px, Py, Pz, x, y, z, A, B, C, K, t;
    unsigned i, fator = 500, constante = 300;

    // COORDENADAS DO CENTRO DO OLHO DIREITO
    Ax = (pt[20].x + pt[23].x)/2;
    Ay = (pt[19].y + pt[24].y)/2;
    Az = (pt[19].z + pt[24].z)/2;

    // COORDENADAS DO CENTRO DO OLHO ESQUERDO
    Bx = (pt[56].x + pt[53].x)/2;
    By = (pt[52].y + pt[57].y)/2;
    Bz = (pt[52].z + pt[57].z)/2;

    // COORDENADAS DO PONTO MÉDIO
    Mx = (Ax + Bx)/2;
    My = (Ay + By)/2;
    Mz = (Az + Bz)/2;

    // COORDENADAS DO PONTO A SER PROJETADO
    Qx = pt[6].x;
    Qy = pt[6].y;
    Qz = pt[6].z;

    // CÁLCULO DAS CONSTANTES
    K = Mx*Bx - Mx*Ay + My*By - My*Az + Mz*Bz - Mz*Az;
    A = (Bx - Ax)/2;
    B = (By - Ay)/2;
    C = (Bz - Az)/2;

    // CÁLCULO DO PARÂMETRO
    t = (-2*A*Qx - 2*B*Qy - 2*C*Qz + K)/(2*(A*A + B*B + C*C));

    // CÁLCULO DO PONTO P
    Px = Qx + A*t;
    Py = Qy + B*t;
    Pz = Qz + C*t;

    // CÁLCULO DO VETOR NB
    MBx = Bx - Mx;
    MBy = By - My;
    MBz = Bz - Mz;

    modulo = sqrt(MBx*MBx + MBy*MBy + MBz*MBz);

    MBx /= modulo;
    MBy /= modulo;
    MBz /= modulo;

    // CÁLCULO DO VETOR PM
    Pfx = Mx - pt[6].x;
    Pfy = My - pt[6].y;
    Pfz = Mz - pt[6].z;

    modulo = sqrt(Pfx*Pfx + Pfy*Pfy + Pfz*Pfz);

    Pfx /= modulo;
    Pfy /= modulo;
    Pfz /= modulo;

    PRODX = MBy*Pfx - MBz*Pfy;
    PRODY = MBz*Pfx - MBx*Pfy;
    PRODZ = MBx*Pfy - MBy*Pfx;

    TRANSx = pt[5].x - pt[6].x;
    TRANSy = pt[5].y - pt[6].y;
    TRANSz = pt[5].z - pt[6].z;

    matrizRT[0][0] = MBx;
    matrizRT[0][1] = MBy;
    matrizRT[0][2] = MBz;
    matrizRT[0][3] = TRANSx;

    matrizRT[1][0] = Pfx;
    matrizRT[1][1] = Pfy;
    matrizRT[1][2] = Pfz;
    matrizRT[1][3] = TRANSy;

    matrizRT[2][0] = PRODX;
    matrizRT[2][1] = PRODY;
    matrizRT[2][2] = PRODZ;
    matrizRT[2][3] = TRANSz;

    matrizRT[3][0] = 0;
    matrizRT[3][1] = 0;
    matrizRT[3][2] = 0;
    matrizRT[3][3] = 1;

    for(i=0; i<qtdPontos3D; i++)
    {
        x = pt[i].x*matrizRT[0][0] + pt[i].y*matrizRT[0][1] + pt[i].z*matrizRT[0][2] + TRANSx;
        y = pt[i].x*matrizRT[1][0] + pt[i].y*matrizRT[1][1] + pt[i].z*matrizRT[1][2] + TRANSy;
        z = pt[i].x*matrizRT[2][0] + pt[i].y*matrizRT[2][1] + pt[i].z*matrizRT[2][2] + TRANSz;
    }
}

```

Figura 47 – Código de transformação geométrica de translação e rotação de imagens 3D.

Fonte própria.

