

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

DISSERTAÇÃO DE MESTRADO

SUMARIZAÇÃO AUTOMÁTICA DE VÍDEOS
EM SESSÕES DE TESTE DE USABILIDADE

TIAGO OLIVEIRA DE ALBUQUERQUE

CAMPINA GRANDE – PARAÍBA

DEZEMBRO – 2012

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

SUMARIZAÇÃO AUTOMÁTICA DE VÍDEOS EM
SESSÕES DE TESTE DE USABILIDADE

TIAGO OLIVEIRA DE ALBUQUERQUE

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande - Campus I como parte dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação.

JOSÉ EUSTÁQUIO RANGEL DE QUEIROZ, DSC.
HERMAN MARTINS GOMES, PHD.
(ORIENTADORES)

ÁREA DE CONCENTRAÇÃO: CIÊNCIA DA COMPUTAÇÃO

LINHA DE PESQUISA: MODELOS COMPUTACIONAIS E COGNITIVOS

CAMPINA GRANDE - PARAÍBA

©TIAGO OLIVEIRA DE ALBUQUERQUE, JANEIRO DE 2012.



FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCCG

A345s

Albuquerque, Tiago Oliveira de.

Sumarização automática de vídeos em sessões de teste de usabilidade/Tiago Oliveira de Albuquerque – Campina Grande, 2012.
138f.: il. col.

Dissertação (Mestrado em Ciências da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática.

Orientadores: Prof. Dr. José Eustáquio Rangel de Queiroz, Profo. Dr. Herman Martins Gomes.

Referências.

1. Teste de Usabilidade. 2. Sumarização Automática. I. Título.

CDU 004.415.53(043)

**"SUMARIZAÇÃO AUTOMÁTICA DE VÍDEOS DE SEÇÕES DE TESTE DE
USABILIDADE"**

TIAGO OLIVEIRA DE ALBUQUERQUE

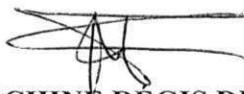
DISSERTAÇÃO APROVADA EM 06/12/2012



**JOSÉ EUSTAQUIO RANGEL DE QUEIROZ, D.Sc, UFCG
Orientador(a)**



**HERMAN MARTINS GOMES, Ph.D, UECG
Orientador(a)**



**JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO, D.Sc, UFCG
Examinador(a)**



**MARIA ELIZABETH SUCUPIRA FURTADO, Dr^a, UNIFOR
Examinador(a)**

CAMPINA GRANDE - PB

AGRADECIMENTOS

Inicio meus agradecimentos com a certeza de que falhei em agradecer a cada pessoa que, de uma forma ou de outra, me ajudou a concluir este trabalho. Minha falta não muda a profunda gratidão que tenho a cada uma delas.

Aos meus orientadores, José Eustáquio Rangel de Queiroz e Herman Martins Gomes, por me aceitarem como seu orientando de mestrado, pela dedicação e paciência que investiram e por cada ensinamento que me ofereceram.

A toda minha família pelo apoio. Minha mãe Vanja e meu pai Inaldo, por cada segundo de minha vida, pela educação, criação, amor, dedicação e apoio que sempre tive, em cada etapa de minha vida.

A minha namorada Shirley Targino Silva. Meu amor, você sabe que sempre fui ruim com palavras, mas é simplesmente por não haver palavras suficientes para descrever o quanto você é importante para mim.

A Thalysson Davyd Vieira, pela importante ajuda ainda no início do projeto da dissertação e aos amigos que fizeram parte do Laboratório de Visão Computacional (LVC).

A Saulo de Tarso Oliveira Rodrigues por ter me ensinado o caminho para ingressar no mestrado.

Aos membros da banca examinadora, pelas contribuições e sugestões de melhorias para este documento.

Aos professores e funcionários da UFCG, principais responsáveis pela minha formação, especialmente aos membros da COPIN Rebecka e Vera, pela paciência e ajuda.

A cada um dos estudantes e colegas que participaram voluntariamente dos experimentos realizados neste trabalho.

A Deus, Grande Arquiteto do Universo, criador de todas as coisas, por me dar a vida, força e iluminação para mais esta empreitada.

A todos os que mencionei e aos que esqueci, o mesmo muito obrigado!

RESUMO

No contexto de desenvolvimento de produtos, a etapa de avaliação da usabilidade é tida como complicada e onerosa. O tempo investido pelos avaliadores da usabilidade para analisar cada teste administrado tende a ser algumas vezes maior do que cada teste em si. Não só por isto, a avaliação da usabilidade ganha a atenção de um número cada vez maior de pesquisadores. A cada nova pesquisa, o processo de avaliação da usabilidade é modificado e, ou é adaptado a casos mais específicos, ou são criados moldes mais genéricos para o processo de avaliação. De uma forma ou de outra, algumas etapas e recomendações permanecem sendo propagadas. É o exemplo do registro em vídeo. Neste contexto, propõe-se no presente trabalho, a concepção de uma abordagem destinada ao auxílio ao processo de análise de vídeos de sessões de teste de usabilidade. Uma vez que o processo de análise de vídeos desta natureza é usualmente conduzido de forma manual ou com pouco grau de automação, o tempo necessário para finalizar o processo é demasiado longo. Assim, é constatada a necessidade de reduzir o tempo total associado à análise de vídeos de sessões de teste de usabilidade. A abordagem descrita nesse trabalho visa oferecer ao avaliador da usabilidade vídeos sumariados de sessões de testes de usabilidade. A abordagem adotada nesta pesquisa para a sumarização de vídeos fundamenta-se no uso de modelos ativos de aparência (AAM – *Active Appearance Models*) e de uma medida semanticamente adequada ao caso de comparação de expressões. A validação da técnica desenvolvida consistiu na condução de um teste de usabilidade realizado em condições de trabalho e na análise comparativa dos indicadores quantitativos de usabilidade considerados no teste e dos resultados dos vídeos sumarizados obtidos.

ABSTRACT

On the context of product development, the usability evaluation stage is considered a complex and time consuming one. The time required to the analysis of each conducted test tends to be some times larger than the time required for the conduction of the test itself. Nevertheless, usability evaluation gathers the attention of an ever growing number of researchers. In every new research, the process of usability evaluation is either customized for a more specific case or more generic casts are created to the evaluation processes. Anyhow, a set of stages and recommendations remain getting propagated. It is the example of video recording. In such context, this research proposes a novel approach to automatic summarization of usability test session videos. Considering the analysis of such videos is generally conducted with little to none automation. The time required to finish the process is too long. Therefore, it is found the need to reduce the time associated to the analysis of usability test session videos. The approach described in this research seeks to offer summarized videos to the usability evaluator from the usability test session videos. The summarization process of this research is founded upon Active Appearance Models (AAM), and on a semantically adequate measure for the comparison of facial expressions. The validation of the developed approach consisted on the conduction of a usability test. The test was conducted in real work conditions. From the usability tests were extracted quantitative indicators. More indicators are extracted automatically by a system developed with the proposed approach. A comparative analysis of such indicators is the stepping stone for the validation of the proposed approach.

LISTA DE FIGURAS

Figura 1	– Esquema de um laboratório de usabilidade típico (QUEIROZ, 2001)	25
Figura 2	– Visão de excertos, quadros e classes (Yahiaoui; Merialdo; Huet, 2001).	31
Figura 3	– Arquitetura básica para um sistema de reconhecimento de expressões faciais (TIAN; KANADE; KOHN, 2005)	35
Figura 4	– Teste de usabilidade com reconhecimento de expressões método invasivo (Branco; Encarnação; Marcos, 2006)	37
Figura 5	– Ferramenta eMotion (UNIVERSITEIT VAN AMSTERDAM, 2010)	48
Figura 6	– Diagrama do Processo de Sumarização de Vídeos de Teste de Usabilidade	52
Figura 7	– Vetor sofrendo rotação de 300	56
Figura 8	– Modelos de forma sendo otimizados com relação à escala e rotação: (a) ajuste isolado dos parâmetros; e (b) ajuste simultâneo dos parâmetros. Os pontos em vermelho são o modelo alvo e os pontos em verde são o resultado da aproximação	56
Figura 9	– Conjunto de modelos sendo alinhados: (a) modelos não alinhados; e (b) modelos alinhados.....	57
Figura 10	– Reconstrução do modelo de textura médio	64
Figura 11	– Fluxograma de funcionamento do sistema AAM em vídeos.....	66
Figura 12	– Exemplo de sequência de imagens de uma expressão facial da base de imagens de expressões faciais Cohn-Kanade ©Jeffrey Cohn	73
Figura 13	– Exemplo de expressão neutra e expressão final da base de imagens de expressões faciais Cohn-Kanade ©Jeffrey Cohn: (a) expressão neutra; e (b) expressão final	74

Figura 14	– Gráfico de evolução de uma expressão facial	75
Figura 15	– Gráfico de evolução normalizada de uma expressão facial.....	76
Figura 16	– Gráfico de distribuição de probabilidade do valor de distância entre modelo neutro pela medida de distância média ponto a ponto	78
Figura 17	– Gráfico de distribuição de probabilidade acumulada do valor de diferença entre modelo neutro pela medida de distância média ponto a ponto	78
Figura 18	– Gráfico de distribuição de probabilidade do valor de diferença entre modelo neutro pela distância de Procrustes	79
Figura 19	– Gráfico de distribuição de probabilidade acumulada do valor de diferença entre modelo neutro pela distância de Procrustes	80
Figura 20	– Gráfico de distribuição de probabilidade do valor de diferença entre modelo neutro pela média dos erros ao quadrado	81
Figura 21	– Gráfico de distribuição de probabilidade acumulada do valor de diferença entre modelo neutro pela média dos erros ao quadrado	81
Figura 22	– Imagens representativas dos agrupamentos de imagens do banco de imagens Cohn-Kanade. Abaixo de cada figura os rótulos g0, 1, ..., 6 indicam a nomenclatura adotada para as sequências de imagens visualmente semelhantes à expressão que está representada imediatamente acima de cada rótulo ©Jeffrey Cohn	82
Figura 23	– Análise do grupo g2 de expressões faciais	84
Figura 24	– Interface do sistema desenvolvido com base na abordagem proposta	92
Figura 25	– Interface de criação de modelo de comparação	92
Figura 26	– Gráfico representativo dos valores do Quadro 8.....	94
Figura 27	– Gráfico de soma de eventos e de opções erradas em cada tarefa	94
Figura 28	– Distribuição numérica do Grau de Instrução dos usuários de teste	126
Figura 29	– Distribuição numérica do Gênero dos usuários de teste.....	127
Figura 30	– Distribuição numérica dos usuários de teste que usam algum tipo	

de corretivo visual.....	127
Figura 31 – Distribuição numérica da Faixa Etária dos usuários	128
Figura 32 – Distribuição numérica dos Ambiente Operacionais mais utilizados pelos usuários de teste	128
Figura 33 – Distribuição numérica do nível de conhecimento em informática declarado pelos usuários.....	129
Figura 34 – Distribuição numérica da Frequência de Uso de Aplicativos de Composição multimídia.....	130
Figura 35 – Distribuição numérica da experiência com a suíte de aplicativos do LibreOffice.org	131
Figura 36 – Distribuição numérica da Experiência com o Aplicativo IMPRESS	131
Figura 37 – Distribuição numérica da frequência de Uso do Aplicativo IMPRESS.....	132
Figura 38 – Distribuição numérica da Habilidade de Fala da Língua Inglesa	133
Figura 39 – Distribuição numérica da Habilidade de Escrita da Língua Inglesa	133
Figura 40 – Distribuição numérica da habilidade de compreensão da Língua Inglesa	133
Figura 41 – Distribuição numérica da Habilidade de Leitura da Língua Inglesa	133

LISTA DE QUADROS

Quadro 1	– Valor geral da média, desvio padrão e variância para as medidas MP2PD, Distância de Procrustes e MSE	75
Quadro 2	– Média geral das medidas de correlação de Pearson, teste-F e teste-T entre as medidas MP2PD, Distância de Procrustes e MSE	75
Quadro 3	– Média geral das médias, desvios padrão e variâncias para as medidas MP2PD, Distância de Procrustes e MSE após normalização	76
Quadro 4	– Média geral das medidas de correlação de Pearson, teste-F e teste-T entre as medidas MP2PD, Distância de Procrustes e MSE após normalização	77
Quadro 5	– Taxa de acerto de cada grupo de expressão facial	84
Quadro 6	– Taxa de apresentação de cada grupo na área de erro.....	85
Quadro 7	– Síntese da análise de indicadores quantitativos.....	91
Quadro 8	– Sumário dos valores de comparação da face do usuário com o modelo neutro.....	93
Quadro 9	– Síntese dos resultados da sondagem do perfil do usuário	134

LISTA DE ABREVIACÕES

AAM	Active Appearance Models (Modelos Ativos de Aparência)
ABNT	Associação Brasileira de Normas Técnicas
ACM	Active Contour Models (Modelos Ativos de Contorno)
AFERS	Automated Facial Expression Recognition System (Sistema Automatizado de Reconhecimento de Expressões Faciais)
ASM	Active Shape Models (Modelos Ativos de Forma)
BD	Banco de Dados
BSD	Berkeley Software Distribution (Distribuição de Software da Berkeley)
CLM	Constrained Local Models (Modelos Locais Restritos)
DNMF	Discriminant Non-negative Matrix Factorization (Fatoração Discriminante de Matrizes Não negativas)
DR	Disturbance Ratio (Razão de Distúrbio)
DRUM	Diagnostic Recorder for Usability Measurement (Gravador de Diagnóstico para Mensuração de Usabilidade)
EICA	Enhanced Independent Component Analysis (Análise de Componentes Independentes Melhorada)
EMG	Electromyogram (Eletromiografia)
FACS	Facial Action Coding System (Sistema de Codificação de Ações Faciais)
FAU	Facial Action Unit (Unidade de Ação Facial)
FCP	Facial Characteristic Point (Ponto de Característica Facial)
FLDA	Fisher Linear Discriminant Analysis (Análise Linear

	Discriminante de Fisher)
HMM	H idden M arkov M odels (Modelos de Markov Ocultos)
ICA	I ndependent C omponent A nalysis (Análise de Componentes Independentes)
IHC	I nteração H umano- C omputador
ISO	I nternational O rganization for S tandardization (Organização Internacional de Padronização)
LBP	L ocal B inary P atterns (Padrões Binários Locais)
LDA	L inear D iscriminant A nalysis (Análise Linear Discriminante)
LFDA	L ocal F isher D iscriminant A nalysis (Análise Discriminante Local de Fisher)
LLVM	L ow L evel V irtual M achine (Máquina Virtual de Baixo Nível)
LPP	L ocality P reserving P rojection (Projeção Preservativa de Localização)
MCF	M odified C orrelation F ilters (Filtros de Correlação Modificados)
MP2PD	M ean P oint t o P oint D istance (Distância Média Ponto a Ponto)
MSE	M ean S quared E rror (Erro Médio Quadrático)
MU	M otion U nit (Unidade de Movimento)
MUSiC	M etrics for U sability S tandards in C omputing (Métricas para Padrões de Usabilidade em Computação)
NCC	N ormalizer C ut C riterion (Critério de Corte Normalizado)
OpenCV	O pen S ource C omputer V ision L ibrary (Biblioteca de Visão Computacional de Código Aberto)
OpenVULab	O pen V irtual U sability L ab (Laboratório Virtual Aberto de Usabilidade)
OSX	Apple's O perating S ystem 10 (Sistema Operacional 10 da Apple)

PAAM	Probabilistic Active Appearance Models (Modelos Probabilísticos Ativos de Aparência)
PCA	Principal Components Analysis (Análise de Componentes Principais)
PDI	Processamento Digital de Imagens
RGB	Red, Green, Blue (Vermelho, Verde, Azul)
SVM	Support Vector Machine (Máquina de Suporte de Vetores)
THA	Think Aloud (Pensar Alto)
ViSA	Video Segmentation and Annotation (Segmentação e Anotação de Vídeo)

SUMÁRIO

1	Considerações Iniciais	16
1.1	Contextualização do Problema	16
1.2	Motivações para a Pesquisa	19
1.3	Objetivos da Pesquisa	20
1.3.1	<i>Objetivo Geral</i>	20
1.3.2	<i>Objetivos Específicos</i>	20
1.4	Objetos da Pesquisa	20
1.5	Organização do Documento	22
2	Revisão Bibliográfica	23
2.1	Testes de Usabilidade com Registro em Vídeo	24
2.2	Sumarização Automática de Vídeos	29
2.3	Reconhecimento de Expressões Faciais	35
2.3.1	<i>Modelos Deformáveis</i>	40
2.4	Ferramentas Disponíveis para o Suporte a Atividades de Avaliação da Usabilidade	44
2.5	Considerações Finais	49
3	Abordagem Proposta	51
3.1	Visão Geral	52
3.2	Detalhes de Desenvolvimento	53
3.2.1	<i>Modelos Ativos de Aparência</i>	53
3.2.2	<i>Processo de Treinamento</i>	65
3.2.3	<i>Fluxo de Execução</i>	65
3.2.4	<i>Sumarização de Vídeos</i>	67
3.3	Detalhes de Implementação	68
4	Avaliação Experimental	70

4.1 Experimento 1 – Busca pela Melhor Medida de Similaridade de Modelos de Forma.....	70
4.1.1 Resultados do Experimento 1.....	75
4.2 Experimento 2 – Discriminação de Expressões Faciais	82
4.2.1 Resultados do Experimento 2	83
4.3 Experimento 3 – Relação entre Sumário e Dados Quantitativos	86
4.3.1 Ambiente	86
4.3.2 Produtos Selecionados para Condução das Sessões de Teste....	86
4.3.3 Elaboração dos Documentos Relacionados ao Teste.....	88
4.3.4 Análise do Perfil do Usuário	89
4.3.5 Condução das Sessões de Teste	90
4.3.6 Resultados do Experimento 3	91
5 Considerações Finais	95
5.1 Sumário da Dissertação	95
5.2 Conclusões	96
5.3 Contribuições.....	98
5.4 Proposição para Pesquisas Futuras.....	98
Referências Bibliográficas	100
Anexo A Termos e Condições para o Uso do Banco de Expressões Faciais Cohn-Kanade	108
Anexo B Banco de Imagens Cohn-Kanade.....	109
Apêndice A Imagens Anotadas para Treinamento	115
Apêndice B Amostras de Imagens Ajustadas após o Treinamento	118
Imagens com Ajuste Adequado	118
Imagem com Ajuste Inadequado.....	120
Apêndice C Formulário para o Cadastro de Participação.....	121
Apêndice D Documento de Aceitação das Condições de Teste	122
Apêndice E Termo de Confidencialidade	123
Apêndice F Questionário para o Delineamento do Perfil do Usuário	

.....	124
Apêndice G Análise dos Dados dos Questionários para o Delineamento do Perfil do Usuário	126
Apêndice H Síntese dos Resultados da Sondagem do Perfil do Usuário.....	134
Apêndice I Dados Coletados para a Mensuração de Desempenho dos Usuários.....	136
Apêndice J Roteiro das Tarefas de Teste	142

1 Considerações Iniciais

1.1 Contextualização do Problema

A temática central da presente dissertação é a sumarização de vídeos de sessões de testes de usabilidade. Assim sendo, a noção de *usabilidade* ocupa um papel importante no contexto da pesquisa, conforme será discutido adiante.

O termo *usabilidade*, conforme definido na Parte 11 do padrão ISO 9241 (1998), refere-se à extensão na qual um produto pode ser usado por usuários específicos para alcançar objetivos específicos com eficácia, eficiência e satisfação em um determinado contexto de uso. O termo usabilidade também faz referência a métodos de melhoria da facilidade de uso de um produto de *software/hardware*, os quais operam em todas as etapas do processo de desenvolvimento de um *software* (GALITZ, 2007).

Um sistema ideal oculta sua tecnologia de tal forma que o usuário não fica ciente de sua existência (NORMAN, 1998). O estudo da IHC (*Interação Humano-Computador*) insere-se no contexto do desenvolvimento de *software* com o objetivo de auxiliar projetistas a conceberem produtos mais usáveis, seguros e funcionais (ROCHA; BARANAUSKAS, 2003). Nielsen (1993) engloba esses mesmos aspectos no conceito de *aceitabilidade* de um sistema.

A usabilidade ocupa um papel decisivo na aceitação de produtos. Li, Shang e Slaughter (2010) compararam a taxa de sobrevivência das novas empresas de *software* com respeito às capacidades de pesquisa e desenvolvimento, propaganda e facilidade de uso dos produtos. Os autores concluíram que a propaganda e a inovação não são tão importantes quanto a facilidade de uso dos produtos.

Galitz (2007) já havia afirmado que a interface de um produto é, para a maior parte dos usuários, o produto em si e que a usabilidade é a característica

mais importante de sua interface. Uma interface de má qualidade e difícil de usar pode excluir categorias de usuários e comprometer a aceitação do produto pela comunidade usuária.

Um dos recursos que podem ajudar desenvolvedores a melhorar a usabilidade de um produto é a realização de ensaios de usabilidade, parte integrante do processo de desenvolvimento de qualquer produto, implicando usualmente a melhoria de sua qualidade.

Rubin (1994) destacou características de testes de usabilidade que considerou importantes tanto para o produto quanto para a equipe de avaliadores e de desenvolvimento. O autor afirmou que testes de usabilidade podem evidenciar os aspectos do processo interativo nos quais os usuários enfrentam dificuldades para executar suas tarefas, bem como quais os aspectos do produto usados de forma eficiente pelos usuários. No âmbito de ensaios de usabilidade, testes de usabilidade contribuem para remover prejulgamentos e suposições eventualmente assumidas pela equipe de desenvolvimento no início do processo. Os testes também possibilitam adotar ou refutar argumentos e dirimir dúvidas que a equipe possa ter com relação a decisões de projeto. Há diversas outras razões para as quais testes de usabilidade se mostram um recurso importante, mas a principal delas é que testes de usabilidade são poderosas ferramentas para o melhoramento de um produto pela equipe de desenvolvimento.

Testes de usabilidade podem ser conduzidos de diversas maneiras, sendo típica a observação das reações dos participantes, durante a execução das tarefas que lhes foram atribuídas. Durante a realização de testes de usabilidade, os avaliadores são responsáveis pela anotação de aspectos relacionados às reações dos usuários. Portanto, em tais cenários, o registro em vídeo das sessões de teste se faz importante para posterior análise, caso contrário, reações sutis ou reações que ocorram enquanto o avaliador faz anotações podem se perder (ROCHA; BARANAUSKAS, 2003).

Testes de usabilidade com registro em vídeo criam novas possibilidades para a equipe de avaliadores. A capacidade de manipulação do tempo de execução de um vídeo permite aos avaliadores análises mais precisas e minuciosas de eventos ocorridos durante os testes. Detalhes que possam ter escapado no processo de observação direta terão uma segunda chance de serem detectados pelos avaliadores (RUBIN, 1994).

Todavia, realizar testes de usabilidade com registro em vídeo requer tempo e atenção. Basicamente, a análise é feita retrospectivamente, revendo o vídeo da sessão de teste de usabilidade e identificando trechos que podem revelar falhas de usabilidade do sistema em condições de avaliação. Via de regra, estes trechos são identificados com base nas reações dos usuários de teste às diversas situações passíveis de ocorrer durante o uso do sistema. A tarefa do avaliador é usualmente muito árdua, uma vez que as reações dos usuários podem ser sutis e breves, tal como um franzir da testa, expressando confusão ou dúvida, requerendo, portanto, atenção, cuidado e paciência, em face da necessidade de possíveis retornos a determinados trechos do registro em vídeo da sessão de teste.

A motivação para o desenvolvimento desta pesquisa insere-se justamente no cenário da realização de testes de usabilidade de produtos com registro em vídeo. Sabendo que o tempo necessário para que se complete a avaliação de um vídeo de teste de usabilidade é, em geral, cerca de quatro a cinco vezes a duração do vídeo (NIELSEN, 1993; JACOB, 1994; HILBERT; REDMILES, 2000; GELLNER; FORBIG, 2003; HOLZINGER, 2005), a pesquisa realizada focaliza o uso de técnicas de sumarização de vídeos para a análise de sessões de teste de usabilidade, como forma de reduzir o tempo total do processo de análise dos vídeos.

A sumarização automática de vídeos produz sínteses de vídeos nas quais o conteúdo semântico não é alterado, reduzindo, em contrapartida, a duração do vídeo original (HUET; MERALDO, 2005). Técnicas de sumarização de vídeos representam instrumentos valiosos para o avaliador, uma vez que é de seu interesse reduzir o volume dos vídeos das sessões de teste de usabilidade realizadas, sem perder o conteúdo relevante para o processo de análise, via de regra, posterior à etapa de testes.

Sabendo que um usuário de teste usualmente reflete em suas expressões faciais diversas reações (positivas ou negativas) em face de diferentes situações passíveis de transcorrerem ao longo de uma sessão de teste de usabilidade (BRANCO; ENCARNAÇÃO; MARCOS, 2006), **para a presente pesquisa investigaram-se técnicas automáticas de reconhecimento de expressões faciais e de sumarização de vídeo, com o intuito de desenvolver uma abordagem destinada à sumarização de vídeos de sessões de teste de usabilidade.**

1.2 Motivações para a Pesquisa

Uma das razões para a realização desta pesquisa foi a criação de uma forma mais simples e rápida de executar uma tarefa que, em geral, é executada manualmente por avaliadores humanos e, portanto, susceptível a erros. A análise de usabilidade depende da observação, por um avaliador, de um usuário executando tarefas com o auxílio de um produto sob condições de teste. Conforme mencionado na seção anterior, o tempo gasto pelo avaliador para avaliar cada teste é, em geral, quatro a cinco vezes maior do que o tempo despendido na realização do teste em si (NIELSEN, 1993; JACOB, 1994; HILBERT; REDMILES, 2000; GELLNER; FORBIG, 2003; HOLZINGER, 2005).

O avaliador precisa observar a interação do participante do teste com o produto, ao longo de toda a sessão de teste, enquanto faz suas anotações e, ao finalizar o processo de observação, analisar o registro em vídeo várias vezes, verificando eventos de interesse ocorridos durante a sessão. Conforme supramencionado, tal experiência é usualmente tediosa e demorada sendo, conseqüentemente, mais suscetível a falhas de julgamento do avaliador.

Conforme será discutido no próximo capítulo (**Revisão Bibliográfica**), o processo de condução de um teste de usabilidade transcorre quase que inteiramente de forma manual. Apesar de já haver ferramentas destinadas ao suporte a atividades associadas a testes de usabilidade, tais ferramentas usualmente não dispõem de métricas automáticas que possibilitem acelerar o processo, sendo pouco melhores do que registros em papel.

Ao longo da condução desta pesquisa, construiu-se uma aplicação de *software*, baseada na técnica AAM, para auxiliar avaliadores da usabilidade de produtos na etapa de análise de vídeos de sessões de teste de usabilidade, com o intuito de sumarizar os vídeos produzidos durante as sessões. A abordagem desenvolvida nesta pesquisa produz excertos que sumarizam trechos da sessão de teste mais relevantes para o propósito do avaliador, de acordo com regras por ele previamente formuladas.

As regras de avaliação utilizadas pela abordagem desenvolvida consistem de exemplos de expressões faciais, que são fornecidos pelo avaliador. A análise dos vídeos segue uma análise de semelhança entre os modelos de expressões faciais fornecidos e os modelos obtidos da análise das expressões existentes no

vídeo de cada sessão de teste da usabilidade. A sumarização dos vídeos depende de um limiar de semelhança fornecido pelo avaliador que determina a taxa de semelhança que deve ser obtida para que a expressão facial avaliada seja, ou não, considerada como parte do sumário.

1.3 Objetivos da Pesquisa

Nesta seção, serão descritos, em linhas gerais, os objetivos geral e específicos desta dissertação.

1.3.1 Objetivo Geral

Conceber, desenvolver e validar uma abordagem que auxilie avaliadores em processos de análise e sumarização de vídeos de sessões de teste de usabilidade.

1.3.2 Objetivos Específicos

Para que o objetivo geral fosse atingido, fizeram-se necessários os seguintes objetivos específicos:

- Estudar e programar técnicas de Processamento Digital de Imagens (PDI) para extração de características visuais de faces;
- Estudar e programar uma técnica de sumarização de vídeos digitais; e
- Avaliar as técnicas empregadas mediante uma análise experimental objetiva.

1.4 Objetos da Pesquisa

As subseções 1.1 e 1.2 tratam principalmente da área de aplicação e do objetivo desta pesquisa, discutidos apenas em linhas gerais. Uma revisão mais detalhada da área de aplicação desta pesquisa e das técnicas relevantes para o desenvolvimento da abordagem constitui um dos objetos deste trabalho. Desta maneira, os aspectos descritos neste capítulo serão aprofundados no Capítulo 2, bem como serão descritas técnicas relevantes para o desenvolvimento da abordagem descrita nesta pesquisa. O desenvolvimento da abordagem constitui o objeto central desta pesquisa, que está retratado no Capítulo 3 deste trabalho.

De forma geral, o presente trabalho objetiva a aplicação de uma técnica de PDI para a extração de expressões faciais, alinhada ao caso de uso em questão, com o objetivo de desenvolver uma abordagem automática que auxilie a atividade de avaliação da usabilidade da interface de um aplicativo de *software*.

Neste contexto, foram investigados aspectos da técnica de aquisição de expressões faciais que poderiam servir como prova de conceito para o uso da referida técnica como mecanismo fundamental da abordagem desenvolvida. Finalmente, foi investigada uma relação entre medidas quantitativas obtidas de uma análise manual e da abordagem desenvolvida.

Nesta Pesquisa, foram empregados Modelos Ativos de Aparência (AAM) como meio para a extração de características de expressões faciais, assim como para o estudo de diferentes medidas de comparação entre características extraídas das faces. As primeiras questões que norteiam esta pesquisa podem ser assim enunciadas:

- P_1 : A eficácia da abordagem proposta irá variar de acordo com a medida de comparação adotada?
- P_2 : Utilizando esta abordagem, é possível diferenciar entre cada grupo de expressão facial estudado?

O último questionamento e o mais pertinente para o escopo desta pesquisa, diz respeito à utilidade da abordagem proposta em um cenário real de utilização, podendo assim ser enunciado:

- P_3 : A abordagem proposta é eficaz em evidenciar, em um vídeo de sessão de teste de usabilidade, os momentos em que ocorrem eventos relevantes para o diagnóstico da usabilidade do produto avaliado?

Os questionamentos que foram levantados servem de base para que sejam formuladas as hipóteses que irão nortear o desenvolvimento desta pesquisa. Para tanto, alinhadas aos questionamentos que foram levantados, formularam-se três hipóteses a serem investigadas:

- H_1 : Há diferenças estatisticamente significativas entre as diferentes medidas de similaridade de expressões faciais estudadas;
- H_2 : Há diferenças estatisticamente significativas entre os diferentes grupos de expressões faciais estudados; e
- H_3 : Há uma relação estatisticamente significativa entre o nível de dificuldade de uma tarefa e a taxa de apresentação de expressões faciais não neutras nos participantes.

1.5 Organização do Documento

Este documento está subdividido em cinco capítulos. Neste capítulo, é contextualizada a pesquisa ora documentada. No Capítulo **2**, é apresentado um panorama das técnicas computacionais, protocolos e sistemas já existentes que foram relevantes para a condução desta pesquisa. São descritos detalhes dos testes de usabilidade com registro em vídeo, bem como de técnicas de PDI que podem ser aplicadas ao processo de reconhecimento de expressões faciais, com fins à sumarização automática de vídeos de sessões de teste de usabilidade.

No Capítulo **3**, descreve-se a abordagem empregada para a resolução do problema de sumarização de sessões de teste de usabilidade com registro em vídeo, o processo de concepção da abordagem desenvolvida e o fluxo de processamento, incluindo uma formalização matemática para a descrição da abordagem desenvolvida. O funcionamento de cada etapa do fluxo de processamento da abordagem é descrito de forma a facilitar a replicação do mecanismo utilizado.

No Capítulo **4**, são descritos os experimentos conduzidos e discutidos os resultados obtidos. A apresentação dos experimentos contempla os testes realizados necessários para validar a abordagem utilizada. Este capítulo também contém uma avaliação exclusivamente quantitativa da abordagem desenvolvida.

Finalmente, no Capítulo **5**, são feitas considerações finais e apresentadas conclusões formuladas a partir da interpretação dos resultados dos testes descritos no Capítulo **4**, sendo também enumeradas as contribuições desta pesquisa para a avaliação dos resultados de sessões de teste de usabilidade e sugeridos temas para trabalhos futuros.

2 Revisão Bibliográfica

Neste capítulo, é apresentada e comentada uma série de publicações relacionadas às áreas da Engenharia da Usabilidade, da Sumarização Automática de Vídeo e do Reconhecimento de Expressões Faciais.

Na Seção 2.1 (**Testes de Usabilidade com Registro em Vídeo**), apresenta-se uma revisão da literatura relacionada a testes de usabilidade com registro em vídeo que, conforme anteriormente mencionado, constituem uma ferramenta poderosa para o avaliador da usabilidade de produtos de *software/hardware*.

Na Seção 2.2 (**Sumarização Automática de Vídeos**), são revisados trabalhos relacionados à sumarização automática de vídeos digitais, que vem sendo aplicada com sucesso em diversas áreas.

A Seção 2.3 (**Reconhecimento de Expressões Faciais**) contém uma revisão da literatura relacionada ao reconhecimento de expressões faciais que são, provavelmente, a melhor forma de diagnosticar automaticamente reações de um indivíduo durante o uso de um dado produto sob condições de avaliação, sem a necessidade de uso de técnicas invasivas.

Por fim, na Seção 2.4 (**Ferramentas Disponíveis para o Suporte a Atividades de Avaliação da Usabilidade**), há uma breve apresentação de algumas ferramentas concebidas com o objetivo de auxiliar avaliadores em sessões de testes de usabilidade, as quais foram relevantes para a pesquisa ora documentada.

2.1 Testes de Usabilidade com Registro em Vídeo

O estudo de processos interativos entre humanos e computadores busca desenvolver modelos teóricos do comportamento humano, bem como criar ferramentas e metodologias capazes de auxiliar o avaliador da usabilidade a diagnosticar a qualidade da interação humano-computador.

A Interação Humano-Computador (IHC) é uma área multidisciplinar que envolve aspectos da Psicologia, da Ergonomia, da Neurofisiologia e da Informática, dentre outras, cujo objetivo é estudar como usuários humanos e computadores interagem, de forma a desenvolver formas melhores de comunicação, a fim de que as necessidades humanas sejam satisfeitas da forma mais eficiente possível (GALITZ, 2007).

Galitz (2007) argumentou que a interface de um sistema é, para a maior parte dos usuários, o sistema em si, pois é a parte que pode ser vista, ouvida e tocada. Todo o volume de código-fonte presente nos programas é obscurecido por gráficos exibidos no monitor e por entradas de dados via teclado, *mouse* e inúmeros outros dispositivos de entrada que têm proliferado, sobretudo na última década. O objetivo da interface com o usuário é de tornar a experiência de se usar um sistema uma atividade simples, fácil, produtiva e agradável.

Testes de usabilidade permitem avaliar uma interface sob condições de uso real ou em ambientes controlados (GALITZ, 2007). Essa estratégia de avaliação envolve a participação de usuários representativos de um sistema, os quais auxiliam o avaliador a verificar o grau de conformidade de um produto a critérios predefinidos de usabilidade (RUBIN, 1994).

Testes de usabilidade podem ser realizados tanto no ambiente de trabalho real do usuário (testes de campo) quanto em um ambiente controlado que simula o realismo do ambiente de trabalho do usuário (teste laboratorial), no qual tarefas são executadas de forma semelhante e como parte de um fluxo de trabalho real. Isto possibilita ao avaliador identificar falhas de usabilidade em um dado produto e produzir um diagnóstico que norteará as ações da equipe de projeto, no sentido de melhorar o referido produto (KAIKKONEN ET AL., 2005).

Laboratórios de usabilidade normalmente usam duas salas, a saber: (i) a sala de testes, na qual são executados os testes; e (ii) a sala de observação, ocupada pelos observadores (Figura 1). A sala de testes deve simular o ambiente de uso do sistema sob avaliação, uma vez que é ocupado pelo participante do

teste, um usuário representativo do sistema. Em processos de observação direta com registro em vídeo, uma ou mais câmeras de vídeo, instaladas na sala de teste e conectadas à sala de observação, monitoram aspectos de interesse do avaliador (ROCHA; BARANAUSKAS, 2003; RUBIN, 1994; NIELSEN, 1993).

Figura 1 – Esquema de um laboratório de usabilidade típico (QUEIROZ, 2001)



As câmeras na sala de testes podem registrar imagens da face do usuário de teste, da tela do computador, dos dispositivos de interação e/ou da documentação (NIELSEN, 1993) mas, em geral, costuma-se produzir dois vídeos, um dos quais contendo a face do usuário e outro contendo um registro da tela do computador com a utilização de softwares de captura de tela.

Vídeos de sessões de teste de usabilidade são essenciais para diversos propósitos de pesquisa, pois permitem a análise minuciosa dos eventos ocorridos durante a interação do usuário com um dado produto, ao longo de uma sessão de teste (NIELSEN, 1993; JACOB, 1994; ROCHA; BARANAUSKAS, 2003).

Registrar em vídeo eventos relacionados à realização de tarefas de teste por usuários é sempre um recurso valioso para posterior revisão. Conforme anteriormente comentado, o processo de análise dos vídeos de sessões de teste de usabilidade é um trabalho difícil e tedioso. Portanto, sempre devem ser feitas anotações cuidadosas ou coletados durante o teste registros de eventos do sistema (*log files*), de modo a reduzir o tempo despendido em encontrar eventos críticos (ROCHA; BARANAUSKAS, 2003).

Nielsen (1993) argumentou que o tempo destinado à análise do vídeo de uma sessão de teste de usabilidade é maior do que o tempo da própria sessão, de modo que poderia ser mais vantajoso executar novos testes, se não fosse tão difícil recrutar participantes, sobretudo se forem voluntários. O tempo necessário para a análise completa do vídeo de uma sessão de teste de usabilidade é tipicamente de quatro a cinco vezes o tempo do teste em si, podendo chegar a ser dez vezes maior (NIELSEN, 1993; JACOB, 1994; HILBERT; REDMILES, 2000; GELLNER; FORBIG, 2003; HOLZINGER, 2005).

Além disto, uma análise completa de um produto requer que múltiplas sessões de teste de usabilidade sejam conduzidas e analisadas. Problemas de usabilidade devem ser corretamente identificados, comparados entre usuários e entre condições de teste (VERMEEREN, 2009). Tal tarefa de análise pode facilmente consumir meses; em contrapartida, pode gerar um fluxo de dados relevantes para a equipe de desenvolvimento do produto.

A análise de vídeos de sessões de teste de usabilidade pode ser bastante informal, baseando-se na experiência do avaliador e em suas impressões ao observar o vídeo de uma sessão de teste (KUSHNIRUK; PATEL; CIMINO, 1997).

Uma análise formal do vídeo de uma sessão típica de teste de usabilidade é feita isolando-se no vídeo eventos de interesse associados às atividades executadas durante o teste, cujas durações são registradas em uma tabela. No âmbito dos eventos associados a cada atividade, são separadas, da mesma forma, subeventos associados às tarefas individuais necessárias à finalização da atividade. De forma mais minuciosa, também deve ser marcado o momento em que cada ação individual é executada (HARRIS, 2004).

A tabela gerada a partir da separação das atividades ajuda a identificar momentos específicos do teste, além de servir como descrição formal da interação, bem como de referência para demais observações e anotações dos avaliadores.

Vermeeren (2009) sugeriu um conjunto de *marcadores* para descrever detalhes de cada momento da interação do usuário com o produto sob condições de avaliação, em um teste de usabilidade. Os marcadores sugeridos pelo autor são derivados de uma forma de marcação semelhante àquela introduzida por Sutcliffe et al. (2000).

O objetivo dos marcadores (VERMEEREN, 2009; SUTCLIFFE ET AL., 2000) é descrever formalmente detalhes da interação. Os marcadores de ambos os autores são separados em duas partes, a saber: (i) marcadores de *ações observadas*; e (ii) marcadores de *expressões* (verbais ou não). Os marcadores de ações de Vermeeren (2009) incluem detalhes tais como: (i) *ação incorreta* (ACT), quando o usuário executa uma ação que não faz parte da sequência de ações esperada; (ii) *ação descontinuada* (DISC), quando o usuário para de executar uma ação antes de seu término; e (iii) *ação corretiva* (CORR), quando o usuário executa uma ação corretiva ou uma sequência de ações corretivas. Dentre os marcadores de expressão do mesmo autor, estão: (i) *objetivo incorreto* (GOAL), quando o usuário formula um objetivo que não vai contribuir para a tarefa; (ii) *confuso* (PUZZ), usuário aparenta estar confuso; (iii) *ações aleatórias* (RAND), quando o usuário aparenta estar executando ações sem um propósito; e (iv) *dúvida, surpresa, frustração* (DSF), quando o usuário não sabe se uma ação foi executada corretamente, não entende os efeitos da ação ou frustra-se/sente-se insatisfeito com o efeito da mesma.

O modelo de avaliação MUSiC (*Metrics for Usability Standards in Computing*), (MACLEOD; BOWDEN; BEVAN, 1997; MACLEOD; RENGGER, 1993; BEVAN; MACLEOD, 1994) em composição com a ferramenta DRUM (*Diagnostic Recorder for Usability Measurement*), fornece diretrizes para a avaliação de vídeos de sessões de teste de usabilidade, com foco em obter dados numéricos, quantitativos e qualitativos consoantes com a engenharia da usabilidade de produtos de *software* (MACLEOD; BOWDEN; BEVAN, 1997; MACLEOD; RENGGER, 1993; BEVAN; MACLEOD, 1994), MUSiC define métricas de desempenho, detalhando técnicas como obter dados de: (i) *Eficácia*, eficácia total com a qual um usuário aplica uma ferramenta de *software*; (ii) *Eficácia de tarefa*, eficácia com que as funcionalidades de uma aplicação são usadas para completar uma tarefa; (iii) *Eficiência*, razão da energia produtiva pela energia despendida pelo *software*; e (iv) *Eficiência Temporal*, razão da eficácia pelo tempo gasto pelo usuário em uma determinada tarefa, dentre outras métricas.

Há diversos modelos para a avaliação de vídeos de usabilidade, cada um dos quais dependendo do método adotado na condução do teste. A verbalização de Ações – THA (*Thinking Aloud*) (NIELSEN, 1993) – não se enquadra

adequadamente ao modelo de avaliação de Vermeenen (2009), por fazer com que os usuários executem as ações mais lentamente e, conseqüentemente, cometam menos erros. Para THA, uma avaliação menos formal pode ser mais adequada (HOLZINGER, 2005).

A análise do vídeo de uma sessão de teste de usabilidade é uma tarefa complexa e, de acordo com Jacobsen, Hertzum e John (1998), mesmo que se inclua alguma formalização, ainda será uma tarefa subjetiva, na qual serão encontradas diferenças de julgamento de um avaliador para outro. Jacobsen, Hertzum e John (1998) atestaram que múltiplos avaliadores, analisando o mesmo vídeo, produzem resultados distintos e com pouca superposição dos problemas extraídos. Os autores denominaram tal efeito ‘Efeito Avaliador’ (*Evaluator Effect*) e o definiram como sendo a diferença da análise de um avaliador para outro, na detecção e avaliação da severidade de problemas de usabilidade.

Vermeeren (2009) expôs alguns problemas de inconsistência introduzidos, inadvertidamente, na análise dos resultados de um processo de avaliação: (i) o registro de eventos da interação, no que diz respeito *ao que* é registrado e *com que* nível de detalhamento; (ii) a transcrição de expressões verbais ou não verbais; (iii) a inferência dos objetivos que os usuários fazem; e, em geral, (iv) a identificação de problemas de usabilidade, a decisão se um ou outro problema de usabilidade é semelhante, ou se várias dificuldades comportamentais observadas na realidade são, de fato, instâncias do mesmo problema de usabilidade, ou não.

Queiroz (2001) propôs uma abordagem híbrida para a avaliação da usabilidade de produtos iterativos. A abordagem proposta permite avaliar a usabilidade do produto com base na: (i) *mensuração do desempenho* do usuário ao realizar tarefas com o um dado produto, o que sugere explicitamente a condução de testes de usabilidade; na (ii) *sondagem da satisfação subjetiva* do usuário com relação ao produto; e na (iii) *inspeção da conformidade* do produto a padrões. Queiroz (2001) avaliou a influência que cada enfoque de avaliação sobre a usabilidade da interface com o usuário.

A abordagem de Queiroz (2001) foi estendida por Ferreira (2007) à avaliação da usabilidade de interfaces de *software* com o usuário para dispositivos móveis. Barbosa (2009) fundamentou-se na abordagem proposta

por Queiroz (2001) para desenvolver uma abordagem direcionada à avaliação da usabilidade de interfaces de *software* com o usuário para dispositivos multimodais.

Na abordagem de Queiroz (2001), a inspeção de conformidade sugerida fundamenta-se nas recomendações do padrão internacional ISO 9241. O padrão apropriado deve ser adaptado de acordo com cada aplicação a ser avaliada. Em Ferreira (2007), a etapa de inspeção restringiu-se às partes 10, 11, 14, 16 e 17, do padrão ISO 9241, porém incorporou a recomendação E.161 (*Arrangement of digits, letters and symbols on telephones and other devices that can be used for gaining access to a telephone network*) (ITU, 2001). Barbosa (2009) incluiu às partes já utilizadas por Ferreira (2007) os padrões ISO 14754, ISO 18021 e ISO 24755.

Para a etapa de sondagem da satisfação subjetiva, Queiroz (2001) recomendou a utilização de questionários informatizados, a anotação de comentários feitos pelos usuários de teste e entrevistas não estruturadas.

Metas qualitativas de usabilidade são de difícil verificação ao final de um projeto. No geral, são especificadas de modo que abre margem a uma gama de conjecturas. Portanto metas quantitativas de usabilidade se mostram pertinentes (QUEIROZ, 2001).

Para a mensuração de desempenho do usuário, recomendou-se a adoção de métricas claras e objetivas de desempenho, como medidas de tempo, contagens de eventos ou verificação de frequências e proporções entre eventos, durante a execução de cada sessão de teste de usabilidade (QUEIROZ, 2001).

2.2 Sumarização Automática de Vídeos

Em testes de usabilidade com registro em vídeo, o volume requerido de usuários de teste se reflete, de forma proporcional, no volume de vídeos das sessões de teste.

Nielsen (1993) argumentou que 80% dos problemas de usabilidade podem ser identificados com até cinco usuários. Com este resultado, Nielsen tentou demonstrar que o rigor estatístico poderia ser negligenciado, no caso de testes de usabilidade reais. Por outro lado, Faulkner (2003) realizou experimentos com 100 usuários e observou amostras aleatórias de diferentes tamanhos. Faulkner concluiu que usar apenas cinco usuários pode resultar em

resultados muito deficientes, com a identificação de apenas 55% dos problemas de usabilidade da aplicação testada, além de que usar vinte usuários garantiria a identificação de um *mínimo* de 95% (98%, em média) dos problemas de usabilidade existentes.

Considerando, por exemplo, um universo amostral de vinte usuários de teste, atuando em sessões de 60 minutos, o tempo necessário para a análise dos vídeos – em geral de quatro a cinco vezes a duração da sessão de teste (NIELSEN, 1993; JACOB, 1994; HILBERT; REDMILES, 2000; GELLNER; FORBIG, 2003; HOLZINGER, 2005) – seria de 4800 a 6000 minutos. Tal exemplo ilustra que o tempo para a análise de vídeos é um fator crítico no processo de avaliação da usabilidade de um produto.

Vídeos são registros compostos por uma sequência volumosa de imagens redundantes e não estruturadas, exibidas ao longo de uma linha de tempo. Identificar uma sequência de interesse em um vídeo ou em uma série de vídeos sempre constitui uma tarefa difícil e não trivial (GONG; LIU, 2000).

A afirmação de Gong e Liu (2000) é particularmente verdadeira no contexto de vídeos de sessões de teste de usabilidade, nas quais há poucas mudanças substanciais na aparência dos vídeos. Vídeos de faces em sessões de teste de usabilidade caracterizam-se por apresentar uma estrutura na qual rostos humanos exibem, em geral, uma expressão neutra. A variação entre vídeos reside fundamentalmente na identificação do rosto que está sendo filmado e a variação ao longo do vídeo fica ao encargo dos curtos intervalos em que o usuário altera sua expressão facial. Destarte, a relevância de uma ferramenta para a sumarização de vídeos de sessões de testes de usabilidade.

Em geral, a demanda por ferramentas que facilitem a navegação em mídias digitais deve-se ao crescente volume de documentos multimídia, assim como às dificuldades de navegação em tais documentos (HUET; MERIALDO, 2005).

Nos anos 90, a sumarização de vídeos ganhou uma parcela mais expressiva do interesse da comunidade acadêmica (ARMAN ET AL., 1994; PFEIFFER ET AL, 1996; SMITH; KANADE, 1997; SMOLIAR; ZHANG, 1994; TONOMURA; AKUTSU, 1993). Desde então, surgiram as mais variadas técnicas, as quais são classificadas de acordo com o tipo de sumário produzido. Sumários de vídeos podem ser de natureza estática ou dinâmica. Sumários de natureza estática

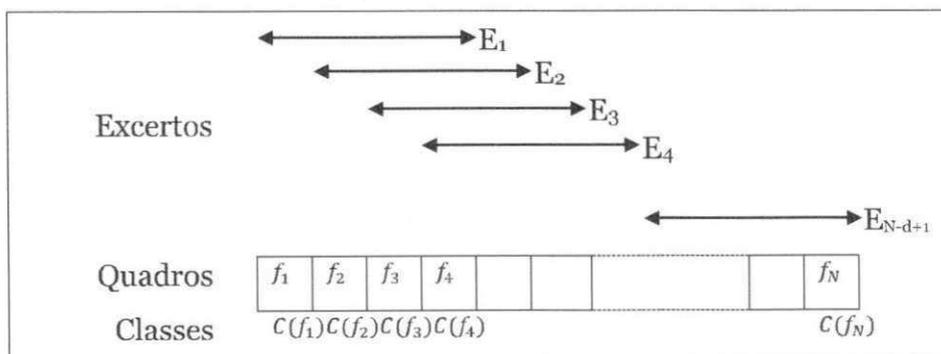
utilizam técnicas que possibilitam extrair uma pequena quantidade de quadros espacialmente separados. Este tipo de sumário produz algo semelhante a um *storyboard*. Sumários de natureza dinâmica, também denominados *video skims*, agrupam sequências de quadros do vídeo original, ao invés de quadros isolados, para compor um vídeo mais curto (HUET; MERIALDO, 2005; TRUONG; VENKATESH, 2007).

De uma forma ou de outra, o trabalho de um algoritmo de sumarização automática de vídeos é reter o conteúdo relevante e descartar o conteúdo desnecessário. Portanto, esta tarefa dependerá, integralmente, do propósito e do domínio da aplicação de sumarização.

Godil (2004) usou uma das formas mais simples de se obter um sumário estático de um vídeo. A ferramenta ViSA, desenvolvida pelo autor, analisa o histograma de cor, em vídeos de captura de tela e identifica como relevantes quadros nos quais há grandes diferenças no histograma. As diferenças entre quadros nos histogramas é decorrência de abertura ou fechamento de janelas. O sumário produzido pela ferramenta de Godil (2004) é uma sequência de quadros que caracterizam uma interação com um programa de janelas comum.

Uma abordagem relativamente simples para a construção de sumários estáticos é proposta por Yahiaoui, Merialdo e Huet (2001), de acordo com os quais busca-se atingir uma cobertura máxima. Primeiramente, é necessário que sejam criadas classes de similaridade C entre todos os quadros do vídeo, de forma que um quadro f_i é semelhante a outro quadro f_j se, e somente se, $C(f_i) = C(f_j)$. Em seguida, considera-se um excerto E de tamanho d . Se o tamanho do vídeo é de N quadros, então serão verificados $N - d + 1$ excertos diferentes do vídeo (Figura 2).

Figura 2 – Visão de excertos, quadros e classes (YAHIAOUI; MERIALDO; HUET, 2001).



Os autores definiram a cobertura $Cov(C)$ de uma classe C como sendo o número de excertos que contêm ao menos um quadro que pertença à classe C . Analogamente, a cobertura para um conjunto de classes $Cov(C_1, C_2, \dots, C_k) = |\{i: \exists j, l f_j \in E_i \text{ e } C(f_j) = C_l\}|$, é definida como sendo o número de excertos que contêm ao menos um quadro que pertença a qualquer das classes C_1, C_2, \dots, C_k .

Desta forma, em um sumário de vídeo composto por k quadros, haverá uma cobertura normalizada igual a $\frac{Cov(C(f_1), C(f_2), \dots, C(f_k))}{N-d+1}$. Assim sendo, o problema da sumarização com cobertura ótima usando k quadros resume-se a um problema de escolher k classes, de forma a maximizar a cobertura normalizada e, em seguida, selecionar os quadros representativos de cada classe (YAHIAOUI; MERALDO; HUET, 2001), conforme a Eq. 1.

$$S = \underset{f_1, f_2, \dots, f_k}{\operatorname{argmax}} \frac{Cov(C(f_1), C(f_2), \dots, C(f_k))}{N-d+1} \quad (1)$$

Para a criação de cada classe de similaridade são criados vetores de características de cada quadro do vídeo composto de um histograma de cor de nove regiões retangulares de mesmo tamanho. Quadros semelhantes em sequência são descartados de modo a simplificar o processamento (YAHIAOUI; MERALDO; HUET, 2001).

Os resultados experimentais dos autores mostram que a cobertura normalizada depende eminentemente da natureza do vídeo sendo utilizado. Nos vídeos em que o agrupamento de classes mostrou uma diversidade relativamente maior aos outros vídeos testados, houve uma diminuição na taxa de cobertura normalizada (YAHIAOUI; MERALDO; HUET, 2001).

Ngo, Ma e Zhang (2005) empregaram grafos dirigidos e *clusters* de tomadas para modelar sequências de vídeo. Cada *cluster* gerado pelo algoritmo corresponde a um nó do grafo que representa o vídeo, devendo o nó conter uma cena completa. O algoritmo proposto pelos autores separa tomadas do vídeo e as agrupa, em seguida, de acordo com a semelhança entre tomadas, a fim de identificar as cenas. O agrupamento (*clustering*) adota uma estratégia fundamentada no critério de corte normalizado – NCC (*normalized cut criterion*) – que leva em consideração a dissimilaridade entre *clusters*, assim como a similaridade dentro dos *clusters*. O grafo gerado pela abordagem de

Ngo, Ma e Zhang (2005) é um modelo sequencial do que ocorre no vídeo. A partir do uso direto do grafo, é possível obter um sumário estático do vídeo. Ao adicionar elementos temporais no grafo, é possível recuperar segmentos do vídeo e produzir um sumário dinâmico.

A sumarização dinâmica também pode ser obtida, mediante a verificação de características visuais nos vídeos. É comum observar-se a propagação, por diversos quadros de um vídeo, de um evento ou de uma ocorrência nele registrada. Há uma demanda por mecanismos para a verificação de eventos específicos, e.g., a detecção da presença de indivíduos ou lances perigosos em vídeos de partidas de futebol. Assim sendo, a utilização de tais mecanismos para solucionar um problema específico de sumarização dinâmica se afigura natural (HUET; MERIALDO, 2005).

Han, Hamm e Sim (2011) desviaram-se da análise de baixo nível para gerar sumarizações de vídeo e propuseram uma abordagem dependente da seleção de um conjunto de quadros-chave feita por um usuário. A técnica proposta pelos autores para geração de sumários verifica a semelhança de cada quadro tanto com os quadros próximos, no âmbito de uma janela de tempo pré-definida, quanto com os quadros selecionados pelo usuário.

De acordo com os autores, não é uma tarefa simples obter informações semanticamente úteis de vídeos, mas um sistema computacional pode ajudar a analisar quadros visualmente semelhantes e agrupá-los baseados em características de baixo nível. Uma vez que um conjunto de quadros é selecionado por um usuário, a tarefa de encontrar quadros semanticamente semelhantes àqueles selecionados se torna factível. É provável que quadros semanticamente semelhantes sejam, também, visualmente semelhantes e, portanto, podem ser identificados com uma análise de características de baixo nível (HAN; HAMM; SIM, 2011).

Maity, Chakrabarti e Bhattacharjee (2012) empregaram a detecção de objetos e de movimentos na análise de vídeos, com o intuito de criar uma técnica para a sumarização de vídeos genéricos. A técnica proposta pelos autores analisa cada quadro do vídeo de acordo com a quantidade de elementos presentes e a diferença entre cada quadro e o anterior. Os dados obtidos são combinados em um único valor que representa a quantidade de informação

contida no quadro. A etapa seguinte da técnica consiste em um agrupamento quadros do vídeo de forma sequencial até que a diferença de um dos quadros com relação ao primeiro quadro do grupo esteja acima de um limiar. Quando isso ocorre, o quadro em questão será o quadro inicial do grupo seguinte.

Na etapa de sumarização da técnica proposta por Maity, Chakrabarti e Bhattacharjee (2012), mensura-se a razão de distúrbio – *DR (Disturbance Ratio)*, um indicador que permite avaliar a dissimilaridade dentro de cada grupo. Quanto maior a dissimilaridade avaliada dentro do grupo, mais quadros deste grupo serão escolhidos. O sumário do vídeo é construído a partir do resultado da busca por quadros cuja medida da quantidade de informações nele contidas seja a maior de cada grupo.

Ejaz, Tariq e Baik (2012) estudaram a sumarização de vídeos empregando a combinação de correlação dos canais RGB, histogramas de cor e momentos de inércia. Aliado a um mecanismo de agrupamento, o objetivo do mecanismo criado foi extrair quadros chave de um vídeo. De acordo com os autores, uma única característica não é suficiente para descrever todos os detalhes de um quadro de vídeo, esta é a motivação que incita a utilização das características supracitadas.

A técnica de agrupamento utilizada por Ejaz Tariq e Baik (2012) permite computar valores que representam a similaridade entre quadros do vídeo e o último quadro chave selecionado. Esses valores são calculados para cada uma das características utilizadas na técnica. A cada característica está associado um limiar que representa uma tolerância máxima para cada valor computado, um valor computado menor que o limiar determinado para cada característica indica diferença relevante entre o quadro sendo avaliado e o último quadro chave selecionado. Os valores computados são combinados, de modo a resultar em um único valor que indica a diferença global do quadro sendo avaliado em comparação com o último quadro chave selecionado. Um novo quadro-chave é selecionado quando a diferença global for maior que um limiar estabelecido experimentalmente.

Ren, Liu e Wu (2010) estudaram formas para avaliar a qualidade de sumarizações de vídeos, tendo empregado quatro aspectos, a saber: (i) *Cobertura* – o conjunto de segmentos selecionados devem cobrir todos os trechos “importantes” do vídeo; (ii) *Concisão* – qualquer dos segmentos selecionados não deve conter nada além do necessário; (iii) *Coerência* – o fluxo

entre segmentos do sumário deve ser natural e fluido; e (iv) *Contexto* – os segmentos selecionados e seu sequenciamento deve ser tal que segmentos anteriores estabeleçam o contexto apropriado. Cada aspecto tem sua estratégia de avaliação descrita e avaliada separadamente. O resultado final da avaliação de cada aspecto ainda depende da multiplicação de cada valor obtido por um coeficiente definido pelo usuário.

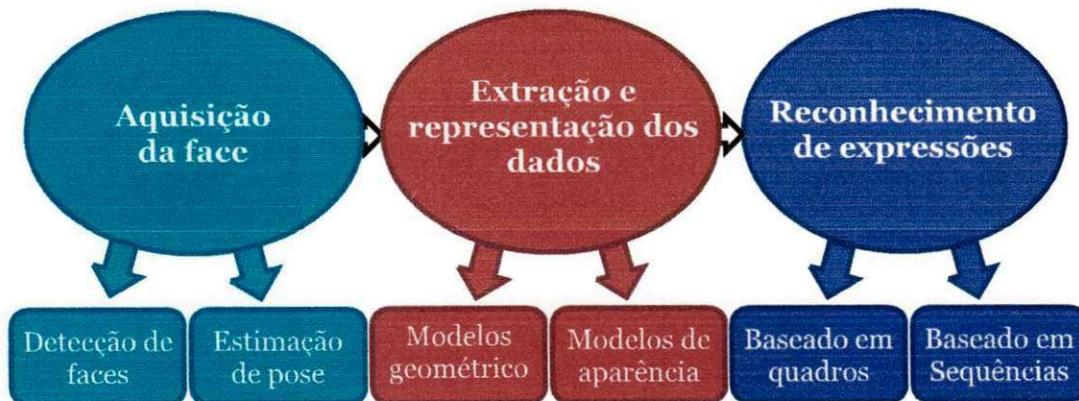
Os testes de validação da abordagem descrita por Ren, Liu e Wu (2010) foram baseados na comparação dos resultados obtidos pela ferramenta com valores definidos por avaliadores humanos. Os autores contaram com a colaboração de 10 voluntários para a classificação de 50 sumários, com base nos quatro aspectos supramencionados. Os resultados apresentaram uma correlação de cerca de 60% na avaliação da coerência e 90% para os demais aspectos considerados.

2.3 Reconhecimento de Expressões Faciais

A expressão facial é uma das principais formas de comunicação não verbal nas relações interpessoais, por apresentar grande potencial informativo e revelar estados emocionais. Eis porque a análise das expressões faciais pode ser uma poderosa ferramenta para avaliar a experiência de usuários em sessões de testes da usabilidade de produtos.

Tian, Kanade e Kohn (2005) descreveram uma arquitetura genérica para um sistema automático destinado ao reconhecimento de expressões faciais. Os autores afirmaram que as arquiteturas existentes para este fim são compostas de três etapas, a saber: (i) a aquisição da face; (ii) a extração e a representação de dados faciais; e (iii) o reconhecimento de expressões (Figura 3).

Figura 3 – Arquitetura básica para um sistema de reconhecimento de expressões faciais (TIAN; KANADE; KOHN, 2005)



A etapa de aquisição da face em vídeos corresponde à detecção quadro-a-quadro da face ou apenas à sua detecção no primeiro quadro, seguida do rastreamento das alterações em sua posição, nos quadros subsequentes (TIAN; KANADE; KOHN, 2005).

Há dois modos básicos de extrair e representar dados de expressões faciais, a saber, o: (i) uso de modelos geométricos; e (ii) uso de modelos de aparência. Modelos geométricos buscam as posições e formatos dos componentes da face (e.g., boca, olhos, sobrancelhas) para criar um vetor de características. Modelos de aparência utilizam filtros de imagem para extrair características, tais como wavelets de Gabor, em toda a imagem do rosto ou em regiões específicas (TIAN; KANADE; KOHN, 2005).

Expressões faciais podem ser reconhecidas estática ou dinamicamente. O reconhecimento *estático* (*frame-based*) compara as características faciais apresentadas em um único quadro com um conjunto padrão de expressões conhecidas sem levar em consideração nenhuma outra informação senão o vetor de características daquele quadro, conforme descrevem Kotsia, Nikolaidis e Pitas (2005, 2006, 2007). O reconhecimento *dinâmico* (*sequence-based*) utiliza a informação temporal de uma sequência de quadros para definir a expressão exibida em um quadro ou em uma sequência de quadros, conforme afirmam Uddin, Lee e Kim (2009).

Branco, Encarnação e Marcos (2006) conduziram um estudo destinado a observar expressões faciais de usuários de teste enquanto interagem com sistemas computacionais. O objetivo do estudo foi primeiramente correlacionar as expressões faciais dos usuários em função de sua dificuldade de interação com um sistema.

Os autores utilizaram, inicialmente, eletromiogramas da atividade facial dos usuários de teste, de forma a quantizar a *tensão* exibida pelos usuários ao encontrar tarefas mais complicadas de serem realizadas, tendo concluído que há alta correlação entre o grau de dificuldade das tarefas e a atividade muscular registrada.

Branco, Encarnação e Marcos (2006) restringiram-se a medir o nível de frustração dos usuários, criando um sistema reativo a tais níveis e iniciando sua pesquisa a partir do uso de sensores EMG (Figura 4), por considerarem que sistemas automáticos de reconhecimento de expressões faciais ainda não estavam completamente consolidados.

Figura 4 – Teste de usabilidade com reconhecimento de expressões método invasivo (BRANCO; ENCARNAÇÃO; MARCOS, 2006)



Diferentemente da pesquisa de Branco, Encarnação e Marcos (2006), a pesquisa ora documentada não se fundamenta em uma abordagem dessa natureza para a obtenção de dados de expressões faciais, mas considera que usuários expressam diferentes reações em diferentes momentos de um processo interativo com um sistema computacional, tais como neutralidade, dúvida, frustração, surpresa ou até raiva e não apenas frustração, conforme sugerido pelos autores. Tais alterações faciais dos usuários são causadas por eventos ocorridos ao longo do processo interativo usuário-computador (ROCHA; BARANAUSKAS, 2003).

Kobayashi e Hara (1992) afirmaram que o primeiro passo para a realização de uma interface ativa com um ser humano é a criação de um sistema capaz de reconhecer expressões faciais. Expressões faciais são inerentes à natureza humana e estão sempre presentes em quaisquer atividades humanas, mesmo que nenhuma reação esperada seja exibida.

Os autores adotaram um esquema de 60 pontos de características de expressões faciais – FCP (*Facial Characteristic Point*) – e uma rede neural treinada para categorizar uma dada expressão em um grupo de seis expressões previamente treinadas, a saber: surpresa, medo, nojo, raiva, alegria e tristeza. A primeira etapa do processo consistiu em obter as coordenadas normalizadas dos FCP nas imagens de faces, a normalização considerou que o espaço entre os olhos das pessoas não variaria, independentemente da expressão esboçada pelo indivíduo na imagem. Assim, todas as distâncias são calculadas assumindo que a distância entre os olhos do indivíduo na imagem era *unitária*. Recebendo com entradas as coordenadas normalizadas dos FCP, a rede neural, depois de treinada, foi capaz de discriminar as diferentes expressões faciais.

Kotsia e Pitas (2005, 2007) e Kotsia, Nikolaidis e Pitas (2006) utilizaram máquinas de suporte de vetores – *SVM (Support Vector Machines)* – para abordar o problema de reconhecimento de expressões faciais.

A abordagem de Kotsia e Pitas (2005, 2007) envolveu a detecção de unidades de ação da face – *FAU (Facial Action Units)* – conforme propostas originalmente por Ekman e Friesen (1978). As FAU representam um conjunto de movimentos musculares da face, os quais foram definidos com o objetivo de melhor padronizar o processo de reconhecimento de expressões faciais, dando origem ao sistema de codificação de ações da face – *FACS (Facial Action Coding System)*. Uma FAU independente pode ser expressa pelas pessoas, mas sozinha não caracteriza uma expressão facial, conforme sugerido por Kanade, Cohn e Tian (2000). Um sorriso, por exemplo, pode ser expresso pelo conjunto FAU (1 + 6 + 9 + 12 + (25 ou 26 ou 27)), variando em função da intensidade de expressão. Vale ressaltar, que nem todas as FAU envolvidas em uma determinada expressão facial precisam ser identificadas para se discriminar com sucesso as diferentes expressões faciais. Kotsia e Pitas (2005, 2007) usaram um subconjunto com apenas oito das 46 FAU especificadas para discernir entre seis emoções.

Para fazer o reconhecimento correto das expressões faciais, Kotsia e Pitas (2005, 2007) empregaram duas abordagens envolvendo máquinas de suporte de vetores, ambas fundamentadas em informações obtidas de 62 FCP, armazenadas em um vetor de dados relativos ao deslocamento geométrico dos pontos. Em uma das abordagens, denominada *direta* pelos autores, foi criada uma SVM multiclasse para mapear diretamente uma dada expressão facial em uma das seis classes. A segunda abordagem, chamada de *FAU-Based*, empregou uma SVM de duas classes para cada uma das oito FAU que caracterizam as seis expressões faciais.

A abordagem proposta por Kotsia, Nikolaidis e Pitas (2006) incluiu informações de textura, obtidas via algoritmo de fatoração discriminante de matrizes não negativas – *DNMF (Discriminant Non-negative Matrix Factorization)*. O algoritmo foi utilizado para obter um valor r_j que representa um tipo de distância da expressão dada e uma expressão neutra padrão. As informações de textura obtidas foram usadas em conjunto com informações geométricas de 104 FCP. SVM foram treinadas com os FCP para discriminar as

expressões faciais computando a semelhança s_j . As distâncias r_j e s_j foram normalizadas para o intervalo $[0, 1]$ e um novo vetor $c_j = [r_j, s_j]^T$ foi usado em uma SVM para classificar as expressões dadas em uma das seis classes de expressões.

Em um desenvolvimento mais recente, Uddin, Lee e Kim (2009) utilizaram sequências de vídeos para categorizar expressões faciais. Os autores propuseram uma combinação de técnicas envolvendo PCA (Principal Component Analysis), ICA (*Independent Component Analysis*), FLDA (*Fisher Linear Discriminant Analysis*), e HMM (*Hidden Markov Models*) em comparação com as técnicas empregadas independentemente.

Para produzir seu classificador, Uddin, Lee e Kim (2009) aplicaram EICA (*Enhanced Independent Component Analysis*) para extrair componentes independentes locais que são classificados a partir da técnica FLDA. De posse das características obtidas, foram criados modelos de Markov (HMM) para descrever as diferentes expressões faciais. Na primeira etapa, os autores utilizaram PCA para fazer uma primeira redução na dimensão do problema. A análise de ICA foi realizada sobre o espaço obtido. Tal sequência de passos produziu a EICA. Finalmente, aplicou-se a FLDA, de forma a condensar as classes semelhantes e separar as classes diferentes o máximo possível. A comparação com as técnicas empregadas de forma independente usou as mesmas estruturas de modelos de Markov sendo criados para cada técnica (modelos para PCA, ICA, EICA, PCA-FLDA e EICA-FLDA). Os modelos de Markov são normalmente usados para modelagem e reconhecimento (RABINER, 1989). O objetivo dos modelos de Markov é determinar que classe de modelo θ tem a maior probabilidade de semelhança $\Pr(O|\theta)$ enquanto observa o dado sequencial $O = \{O_1, O_2, \dots, O_t\}$.

Zhang, Zhao e Lei (2012) empregaram a característica LBP (*Local Binary Patterns*) em combinação com uma técnica de redução dimensional de problemas. O método proposto pelos autores emprega LBP para obter um descritor da textura da imagem sendo avaliada. A dimensão do vetor de características obtido com o LBP é então reduzida. A última etapa do método consiste do treinamento de um classificador SVM para distinguir entre as diferentes classes de expressões faciais treinadas.

O método descrito pelos autores foi testado com diferentes técnicas de redução dimensional. Foram empregados LFDA (*Local Fisher Discriminant Analysis*), PCA, LDA (*Linear Discriminant Analysis*) e LPP (*Locality Preserving Projection*). Os resultados experimentais mostraram que, para o mesmo nível de redução de dimensão, a técnica de LFDA mostra uma maior precisão de reconhecimento seguida por LDA (quando a redução de dimensão é abaixo de seis) e por PCA. A precisão do método de classificação quando está sendo empregada a técnica de LPP mostra os piores resultados para todos os casos.

Chew et al. (2012) propuseram uma nova técnica de classificação aplicada, em sua pesquisa, à classificação de expressões faciais. Os autores afirmaram que a utilização da técnica SVM, apesar de ter sido aplicada com sucesso em numerosas tarefas de classificação, apresenta diversas imperfeições. Os autores alegaram que a eficácia da SVM é negativamente afetada pela presença de ruído no conjunto de treinamento e pela presença de *outliers*. Os autores definiram ruído como sendo a variação natural observada nos dados entre uma observação e outra, cuja variação não pode ser correlacionada à nenhuma fonte, enquanto *outliers* constituem variações extremas nos dados observados, indicando erros de medição ou dados que se enquadram em uma distribuição de cauda longa, razão pela qual *outliers* não podem ser ignorados. Nessas circunstâncias, é necessária cautela no uso de ferramentas que podem ser sensíveis à sua presença.

Os filtros de correlação modificados – MCF (*Modified Correlation Filters*), técnica proposta por Chew et al. (2012), buscam um hiperplano único para criar um classificador binário do conjunto de treinamento. O hiperplano é otimizado, reduzindo a energia do conjunto de alvos e do conjunto de não-alvos. Esta abordagem reduz a influência de *outliers* e de ruído presente no conjunto de treinamento.

2.3.1 Modelos Deformáveis

Modelos deformáveis que são capazes de se ajustar a novos dados são de grande interesse em visão computacional (STEGMANN; ERSBØLL; LARSEN, 2003). Existem diversas abordagens de modelos deformáveis na literatura.

Kass, Witkin e Terzopoulos (1987) descreveram as *Snakes* ou Modelos Ativos de Contorno – *ACM (Active Contour Models)*, os quais definem curvas que se deformam de acordo com forças internas e externas. As forças internas do modelo tendem a manter a curva suave, enquanto que as forças externas empurram e puxam a curva na direção de características locais da imagem, como bordas ou linhas naturais.

Cootes et al. (1994) definiram os Modelos Deformáveis de Formas – *ASM (Active Shape Models)*. Semelhante ao modelo descrito em Scarloff e Isidoro (1998), os modelos descritos por Cootes et al. (1994) são descritos por malhas de triângulos, mas a variação do modelo de forma é descrita anteriormente em uma etapa de treinamento do modelo usando PCA. Cootes et al. (1994) argumentaram que se deveria criar modelos específicos para cada classe de objetos ao invés de usar técnicas genéricas para o rastreamento de objetos deformáveis. Em ASM, a textura é avaliada e combinada com conhecimento prévio de como o modelo poderá se deformar, levando o modelo a convergir com rapidez e precisão.

Scarloff e Isidoro (1998, 2002) propuseram uma abordagem denominada *Blobs* Ativos ou Modelos Deformáveis de Aparência Baseados em Regiões, segundo a qual a forma de um objeto é descrita por uma malha de triângulos deformável de acordo com a movimentação do objeto em uma sequência de quadros de vídeo.

A evolução natural dos ASM foram os Modelos Ativos de Aparência – *AAM (Active Appearance Models)* (EDWARDS; COOTES; TAYLOR, 1998), nos quais a textura dos objetos passou a receber o mesmo tratamento que a forma dos objetos recebe em ASM, no sentido de que também passou a ser descrita de forma dinâmica. Em AAM, o processo de treinamento irá guardar no modelo não só o conhecimento de como o modelo poderá se deformar, como também o conhecimento de como a textura do modelo irá se modificar.

Há uma gama de extensões para AAM na literatura. Em um trabalho posterior, Cootes e Taylor (2001) adotaram um *arcabouço* estatístico em seus algoritmos de ajuste do seu modelo AAM, de forma a possibilitar a inclusão de restrições extras. Os autores mostraram o impacto de tais restrições no número de iterações necessárias para aproximar modelos e como a interação com usuário pode ajudar a guiar o ajuste.

Modelos Diretos de Aparência (HOU ET AL., 2001) são modelos de previsão incorporados à AAM. Os modelos de previsão têm o propósito de modificar o modelo de forma com base no estado do modelo de textura. Os autores levaram em conta que, idealmente, um determinado estado do modelo de forma pode ser associado a múltiplos estados do modelo de textura. Entretanto, a recíproca não é verdadeira, de forma que um determinado estado do modelo de textura é associado a apenas um estado de forma.

Baker e Matthews (2001) argumentaram que, respeitadas algumas restrições, AAM podem ser reduzidos a problemas de gradiente descendente, i.e., o ajuste de AAM em uma nova imagem desconhecida é um problema no qual não há mínimos locais.

Na proposta original de AAM de Cootes e Taylor (1998), os parâmetros de forma são ajustados de forma aditiva enquanto os parâmetros de pose são ajustados de forma composicional. Baker e Matthews (2001) também mostraram que ambos os parâmetros de forma e pose de AAM podem ser ajustados de forma composicional.

Ryan et al. (2009) empregaram AAM e modelos locais restritos – *CLM* (*Constrained Local Models*) – em uma ferramenta para o reconhecimento de expressões faciais. Os autores modificaram o processo normal de construção do CLM de múltiplas maneiras: (i) simplificaram o algoritmo de otimização para que o processo fosse paralelizado; (ii) modificaram os especialistas de região com SVM, o que também reduziu o tempo de execução e aumentou a precisão do ajuste do modelo; e (iii) adotaram um ajuste de natureza composta, ao invés de aditiva, o que aumentou a robustez do algoritmo com relação à variação de escala. O reconhecimento de expressões faciais da ferramenta descrita no trabalho de Ryan et al. (2009) é obtido com a utilização de uma SVM treinada com as imagens de três bancos de imagens disponíveis publicamente.

Zhang et al. (2012) modificaram o processo de treinamento de modelos AAM adicionando elementos probabilísticos à etapa de PCA. Os autores denominaram a abordagem PAAM (*Probabilistic Active Appearance Model*). Em sua abordagem, os autores substituíram a matriz de covariância utilizada pela matriz descrita nas Eqs. 2 e 3.

$$Cov(k, l) = \sum_{i=1}^d \sum_{j=1}^d (k_i - \mu_k)(l_j - \mu_l)P(k_i)P(l_j) \quad (2)$$

$$P(k_i) = \frac{1}{2\pi\sigma} e^{-\frac{|k_i - \mu_k|}{2\sigma_k^2}} \quad (3)$$

Na Eq. 2, os elementos k e l representam elementos de dimensão d (pontos no modelo de forma ou pixels do modelo de textura), μ_k e μ_l representam a média dos valores em k e l , respectivamente. Na Eq. 3, o valor de σ_k^2 corresponde à variância dos elementos em k .

Em seu experimento de validação, os autores utilizaram as imagens de treinamento como plataforma de teste. Os modelos criados a partir da abordagem probabilística adotada e o AAM padrão foram ajustados às imagens de treinamento. A qualidade do ajuste foi avaliada de acordo com a distância média entre pontos e o erro médio quadrático entre os pixels.

Zhou et al. (2012) propuseram modificar a maneira como AAM se comporta com relação às texturas de imagens de teste. A medida utilizada combinou informações de textura e intensidade de bordas. Os autores empregaram uma transformada de *wavelets* invariante à translação para obter coeficientes de baixa frequência da imagem. A intensidade das bordas foi calculada a partir do resultado da transformada de *wavelets*. De acordo com os autores, esta medida é menos sensível à variação das condições da imagem.

Em sua proposta, Senechal et al. (2012) empregaram AAM e SVM *multikernel* para o reconhecimento de FAU. A abordagem proposta pelos autores emprega uma SVM específica para classificar cada FAU de interesse. A técnica de ajuste de AAM empregada pelos autores baseou-se na utilização de dois modelos 2,5-D locais, um aplicado a ambos os olhos, sobrancelhas e nariz e o outro aplicado à boca, base do nariz e contorno do rosto. Os modelos 2,5-D de AAM empregados pelos autores são compostos de um modelo de forma 3-D e de texturas 2-D da visão frontal das imagens de faces, mapeadas sobre a forma 3-D média.

O reconhecimento de expressões faciais vem evoluindo e mudando de curso de acordo com sua própria evolução. Atualmente, há um foco maior na concepção de sistemas que sejam capazes de reconhecer mais expressões, que

reconheçam as FAU com maior precisão, e que sejam mais robustas com relação ao movimento e à pose (DHALL ET AL., 2011, RUDOVIC; PANTIC; PATRAS, 2011; SENECHAL ET AL., 2012; VALSTAR; PANTIC, 2012). A pesquisa ora relatada se beneficiou das evoluções recentes, trazendo o reconhecimento de expressões faciais para um campo de aplicação novo.

2.4 Ferramentas Disponíveis para o Suporte a Atividades de Avaliação da Usabilidade

Há, no mercado, uma série de ferramentas destinadas a auxiliar a etapa de realização de testes de usabilidade. As ferramentas variam, mas há uma base comum de funcionalidades que são encontradas nas diversas ferramentas.

No âmbito da pesquisa ora relatada, foi realizado um levantamento de algumas das ferramentas disponíveis no mercado. São relatadas nesta seção as ferramentas que contêm descrições mais completas em suas páginas Web. As ferramentas foram encontradas de acordo com uma busca no engenho de busca do *google.com*, considerando as palavras-chave: *tool*, *software*, *usability* e *testing*.

A classe de ferramentas relevantes a esta pesquisa é aquela que possibilita gravar vídeos da execução de tarefas por usuários de teste. *UserTesting.com*¹ é um serviço via Internet que disponibiliza usuários de teste para seus clientes. O cliente do *UserTesting.com* requisita usuários segundo um perfil, informa uma página Web que quer que seja testada (qualquer página, inclusive a de concorrentes), a sequência de tarefas que o cliente quer que os usuários executem e o serviço repassa os detalhes para seus testadores. Os testadores são encorajados a falar o que estão pensando a cada passo do processo. Enquanto o usuário está executando as tarefas, vão sendo gravados vídeos da *webcam* do usuário e da tela.

O *Silverback*² (CLEARLEFT, 2010) é uma ferramenta semelhante a *UserTesting.com*, mas precisa ser instalada nas máquinas do usuário de teste e do avaliador. Além disto, o processo de avaliação fica todo ao encargo do avaliador. Trata-se de um sistema isolado do sistema em avaliação, podendo

¹ <http://www.usertesting.com>

² <http://clearleft.com>

ser utilizado para avaliar a interação de usuários com qualquer sistema computacional, não somente páginas da Internet. O *Silverback* é um sistema formado por duas partes. De um lado, encontra-se o programa que envia um fluxo de vídeo de captura de tela do computador do usuário de teste e de uma câmera direcionada para o rosto do usuário de teste. Do outro, encontra-se o programa que exibe o vídeo capturado. Com poucas funcionalidades adicionais, o *Silverback* é capaz de guardar anotações e inserir marcadores na linha de tempo do vídeo.

Apesar de mais complexa, a ferramenta *Morae*³ (TECHSMITH, 2010) apresenta a mesma base de funcionalidades do que o *Silverback*, com a vantagem de poder capturar a entrada do teclado e os movimentos do *mouse*. Os gráficos gerados pela ferramenta são obtidos com base em anotações que os avaliadores fazem nos vídeos de teste.

O *Open Virtual Usability Lab*⁴ (*OpenVULab*), desenvolvido por Kumar e Owston (2012), é uma ferramenta de código aberto destinada à captura de telas e sons, com o propósito de verificar a acessibilidade de sítios da Internet. A ferramenta coleta dados dos participantes antes, durante e depois da interação dos participantes com o alvo de cada avaliação. Antes de cada interação, um questionário pode ser administrado para obter dados demográficos dos participantes. Durante a interação do participante com o sítio sob avaliação, há a captura das telas durante o processo interativo. Caso o participante tenha um microfone ligado em seu computador, também poderá ser gravada a verbalização de suas ações. Após a conclusão da atividade, é apresentado ao participante um questionário pós-teste, com o intuito de obter sua opinião acerca do processo interativo.

O que diferencia o *OpenVULab* das outras ferramentas de natureza similar é que não há a necessidade instalar nenhuma aplicação de *software*, pois se trata de um serviço Web, o que permite que testes sejam realizados em qualquer computador, inclusive em sistemas nos quais o administrador da rede

³ <http://www.techsmith.com>

⁴ <http://leviathan.edu.yorku.ca/wordpress/>

não permita alterações nas configurações. Os participantes podem usar seus próprios computadores, utilizando o navegador e as configurações que preferir. Outra característica da ferramenta é que os dados obtidos são armazenados nos servidores disponibilizados pela ferramenta, com o intuito de aumentar a segurança dos dados e a privacidade dos participantes de cada teste. Por último, os autores alegam que o *OpenVULab* é a única ferramenta de código aberto dessa natureza.

O estudo conduzido no desenvolvimento do *OpenVULab* contou com a cooperação de 42 estudantes e mostrou que a ferramenta pode operar sem ser intrusiva no processo interativo. Um segundo teste contou com 7 estudantes com dificuldades mentais e de aprendizado, no qual se avaliou a acessibilidade de dois sítios da universidade de York e se obtiveram resultados que não puderam ser obtidos a partir de ferramentas destinadas à análise automática do código dos sítios.

Dentre as ferramentas que incorporam técnicas de visão computacional sobre os vídeos de sessões de teste está a ferramenta *ViSA (Video Segmentation and Anotation)*. A ferramenta foi concebida por Godil (2004) para analisar vídeos de captura de tela de testes de usabilidade, possibilitando a análise de um vídeo qualquer com o objetivo de segmentar o vídeo encontrando quadros-chave com base em alterações do histograma de cor de dois quadros consecutivos. *ViSA* cria uma forma diferente de interação com o vídeo, separando os quadros-chave e permitindo anotações sobre os referidos quadros. Considerando um vídeo de captura de tela, o processo de segmentação também representa uma sumarização do vídeo, uma vez que eventos importantes no vídeo de captura de tela ocorrem, exatamente, nos intervalos de tempo em que a tela sofre alterações.

Ryan, Cohn e Rossi (2009) desenvolveram uma ferramenta, baseada em AAM, para a análise de expressões faciais, com o intuito de auxiliar em interrogatórios. O *Automated Facial Expression Recognition System (AFERS)*, desenvolvido pelos autores, analisa e classifica a expressão facial do indivíduo sendo interrogado em um de sete grupos de expressões faciais.

Sendo uma ferramenta destinada ao auxílio a interrogatórios, *AFERS* foi desenvolvida de maneira a exibir gráficos semelhantes àqueles obtidos a partir de polígrafos, utilizados em cenários semelhantes. No caso desta ferramenta, o resultado da análise é obtido quadro a quadro por uma SVM que determina a probabilidade de que cada expressão analisada pertença ou não a cada um dos sete grupos de expressões.

Mesmo não tendo sido concebido com o intuito de analisar vídeos de sessões de teste de usabilidade, os cenários e a estrutura de utilização e dos vídeos analisados pela ferramenta são semelhantes. Portanto, a análise gráfica utilizada pela ferramenta pode ser estendida para o caso da utilização na análise de vídeos de sessões de teste de usabilidade.

*eMotion*⁵ (UNIVERSITEIT VAN AMSTERDAM, 2010) é uma ferramenta criada por pesquisadores da Universidade de Amsterdã para classificar expressões faciais em tempo real (Figura 5).

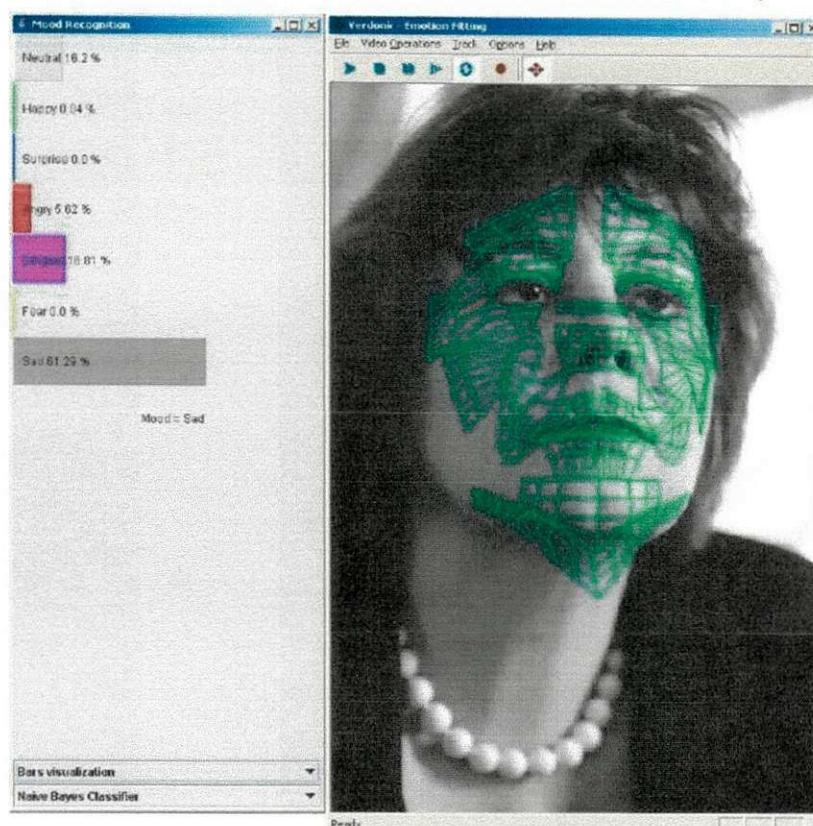
A ferramenta mantém um registro da evolução das expressões faciais dos usuários no tempo. Sendo pouco mais do que uma demonstração de tecnologia, a funcionalidade mais peculiar da aplicação é se conectar com o jogo de realidade alternativa '*Second Life*' e reproduzir, em tempo real, as expressões do usuário no *avatar* do jogo.

*Tobii*⁶ (ELVESJÖ; SKOGÖ; ESKILSSON, 2010). A ferramenta é capaz de rastrear o movimento dos olhos e inferir a direção para a qual o usuário está olhando. A partir das informações obtidas em uma sessão de teste, a ferramenta é capaz de gerar gráficos que indicam quais as regiões da tela que foram mais observadas pelo usuário (*Heatmaps*). Adicionalmente, podem ser gerados gráficos de barra e apresentadas tabelas contendo resultados estatísticos.

⁵ <http://www.visual-recognition.nl/index.html>

⁶ <http://www.tobii.com>

Figura 5 – Ferramenta eMotion (UNIVERSITEIT VAN AMSTERDAM, 2010)



Desenvolvida por Staiano et al. (2012), o *UX_Mate* faz análise automática da experiência do usuário empregando o rastreamento de movimentos da face. Diferentemente da ferramenta *eMotion*, o *UX_Mate* não se fundamenta na classificação de emoções, sendo a análise da face obtida a partir da identificação de FAU e MU (*Motion Units*).

O estudo conduzido no desenvolvimento da ferramenta relacionou a quantidade de erros observados em uma sessão de teste de usabilidade com a presença de cada MU. Os experimentos mostraram que há correlação entre determinadas MU e a taxa de ocorrência erros.

FaceReader™ é uma ferramenta desenvolvida pela *Noldus Information Technology*⁷. A empresa é especializada em aplicativos de *software* voltados para o estudo comportamental. A ferramenta *FaceReader™* foi concebida para a análise automática de expressões faciais em vídeos, sendo capaz de classificar as expressões encontradas em sete grupos de expressões faciais (Neutro, feliz, triste, assustado, enojado, surpreso, enraivado).

⁷ <http://www.noldus.com>

FaceReadertm também é capaz de determinar o estado de diferentes estruturas internas da face do usuário de teste, e.g., se o usuário está ou com a boca ou os olhos abertos ou se a sobrancelha está em uma posição neutra, levantada ou abaixada, bem como a orientação da cabeça. Outra funcionalidade da ferramenta permite determinar a direção geral do olhar, a partir da análise da distância entre a pupila e o canto dos olhos do usuário. A ferramenta também é capaz de estimar o sexo, a idade, a etnia, se o usuário tem barba e/ou bigode ou usa óculos.

2.5 Considerações Finais

Diante do exposto, constata-se que a análise de vídeos de sessões de teste de usabilidade pode beneficiar-se com os avanços descritos ao longo deste capítulo, no qual apresentaram-se técnicas de PDI, com a finalidade de extrair informações de faces humanas e facilitar a comunicação visual.

No que diz respeito à avaliação da qualidade de uma sumarização, nota-se que, no geral, para cada abordagem proposta os autores adotam uma forma própria para avaliar a qualidade da abordagem. No caso da abordagem de Ren, Liu e Wu (2010), o resultado depende dos pesos escolhidos para cada uma das medidas, o que a torna apropriada apenas para a avaliação de sumários de natureza intrínseca. Sumários de vídeos de sessões de teste de usabilidade influenciam a eficácia com que os avaliadores da usabilidade completam a avaliação de um produto. Desta maneira, determinados sumários só podem ser avaliados com relação ao impacto que exercem na eficácia com que os avaliadores completam a avaliação da usabilidade.

O estado da arte da classificação de expressões faciais encontra-se em um patamar de robustez no qual o resultado da classificação já se afigura confiável o bastante para sua aplicação em casos de uso reais, conforme se pôde observar na Seção 2.3.1 (**Ferramentas Disponíveis para o Suporte a Atividades de Avaliação da Usabilidade**). Entretanto, um sistema de classificação similar àqueles descritos, em que um conjunto de imagens de treinamento é a base para a geração de grupos de determinados padrões, pode não ser suficiente para capturar todas as nuances da interação humano-computador. Em vista do exposto, no presente trabalho, evita-se a classificação de expressões em grupos discretos e focaliza-se a similaridade entre expressões, de forma mais genérica.

Há uma riqueza de abordagens baseadas nos trabalhos de Edwards, Cootes e Taylor (1998), o que demonstra a versatilidade dos AAM propostos pelos autores. AAM são, no geral, empregados em conjunção com alguma técnica de classificação e agrupamento, de forma a permitir a classificação de novos modelos de expressões conforme exemplos predeterminados na etapa de treinamento da técnica. Nota-se a presença de duas etapas de treinamento nas abordagens descritas, sendo uma o treinamento do AAM, que fica responsável por extrair as características da face que está sendo avaliada, e a segunda é o treinamento da técnica de classificação empregada pela abordagem, que fica responsável por definir grupos de similaridade e de mapear novas observações em um dos grupos definidos.

Desta forma, percebe-se uma carência de abordagens que permitam a incorporação de novos modelos de expressões faciais sem a necessidade de novo treinamento da técnica de classificação, diminuindo assim a liberdade dos avaliadores de usabilidade. A abordagem de avaliação de expressões faciais empregada neste trabalho tenta minimizar esta limitação.

Na abordagem proposta nesta dissertação, empregou-se AAM para a análise de expressões faciais no contexto de vídeos de sessões de teste de usabilidade. A técnica foi escolhida levando-se em consideração a flexibilidade e o grau de precisão oferecidos pela técnica.

Na abordagem proposta nesta pesquisa não foi empregada uma técnica de classificação, em seu lugar foi utilizada uma medida de similaridade que permite que uma expressão seja avaliada de acordo com sua semelhança com outra expressão facial predeterminada, sem a necessidade de um conjunto de treinamento de expressões faciais semelhantes ou de um treinamento prévio.

No próximo capítulo, apresentam-se a abordagem proposta para a análise de expressões faciais e a técnica utilizada para a criação de sumários.

3 Abordagem Proposta

Neste capítulo, é descrita a abordagem desenvolvida para a sumarização automática de vídeos de sessões de teste de usabilidade. Inicialmente, é fornecida uma descrição geral de suas principais etapas, seguida do detalhamento do desenvolvimento de uma ferramenta de software que a materializa.

Considerando a revisão realizada no capítulo anterior, definiram-se como base da abordagem de sumarização desenvolvida nesta dissertação dois módulos responsáveis por: (i) extrair as informações das expressões faciais do participante de uma sessão de teste de usabilidade dos quadros do vídeo da sessão; e (ii) gerar gráficos representativos da evolução das expressões faciais no decorrer do vídeo.

O funcionamento da ferramenta fundamenta-se no vídeo de uma sessão de teste de usabilidade. De posse do vídeo, o avaliador informa à ferramenta que tipos de expressões faciais são de seu interesse, após o que a ferramenta realiza uma análise do conteúdo do vídeo, de acordo com a abordagem descrita, a fim de verificar os trechos em que as expressões faciais de interesse estão mais presentes.

Para extrair as informações das expressões faciais dos usuários de teste foram utilizados AAM. No Capítulo 2 (**Revisão Bibliográfica**), descreveu-se que AAM são flexíveis, tendo sido estendidos por diversos autores e aplicados em múltiplos cenários, incluindo o reconhecimento de expressões faciais. A técnica AAM é vantajosa por ser mais intuitiva e precisa que as outras técnicas discutidas.

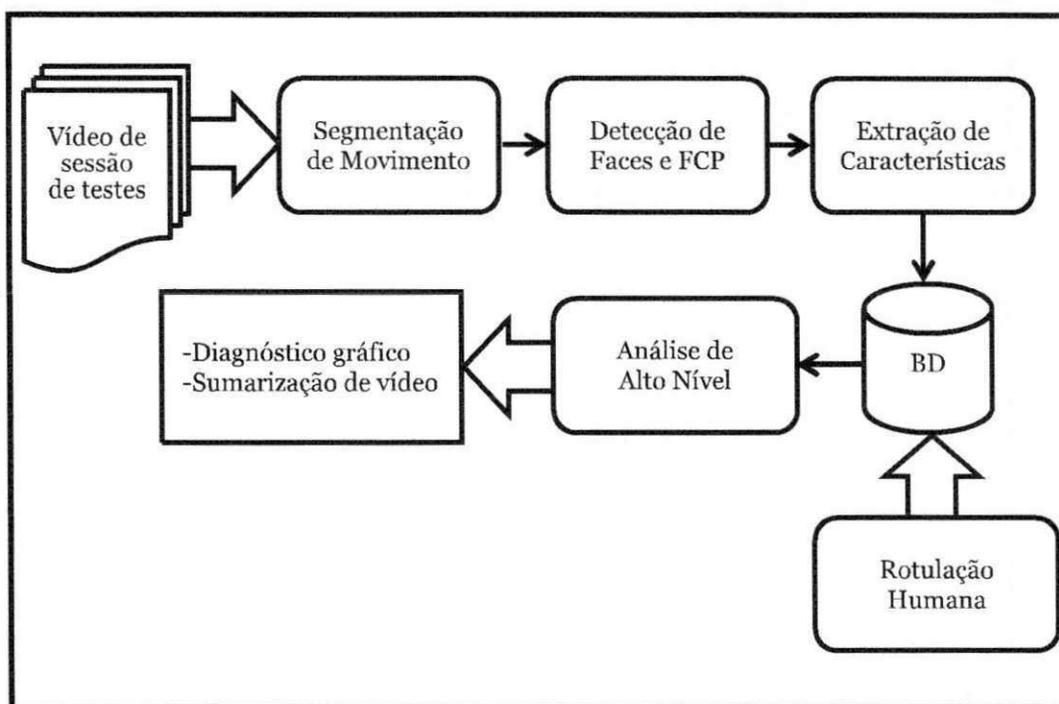
A abordagem para a avaliação experimental da ferramenta desenvolvida observou as hipóteses definidas no Capítulo 1 (**Considerações Iniciais**). Os objetivos da avaliação experimental conduzida nesta pesquisa foram (i) validar

o mecanismo da ferramenta desenvolvida no âmbito dos objetivos traçados para a pesquisa e (ii) relatar a utilidade da ferramenta. Para a realização do segundo objetivo se fez necessária a realização de testes de usabilidade. A abordagem híbrida proposta por Queiroz (2001) fundamentou a realização desta etapa dos experimentos, tendo sido focada apenas (i) a condução das sessões de teste de usabilidade, para a obtenção dos vídeos das referidas sessões e (ii) a análise objetiva dos vídeos.

3.1 Visão Geral

A arquitetura simplificada do sistema ora descrito é apresentada na Figura 6.

Figura 6 – Diagrama do Processo de Sumarização de Vídeos de Teste de Usabilidade



A etapa de 'Segmentação de Movimento' presente no diagrama compreende o processo de identificar a posição e orientação da face em cada quadro relativa a sua posição no quadro anterior. A etapa seguinte, de 'Detecção de Faces e FCP', ajusta o modelo de face desenvolvido nesta pesquisa à face presente na imagem. Durante a etapa de 'Extração de Características', o modelo de face ajustado à imagem é recuperado e armazenado.

A etapa de 'Rotulação Humana' é a única etapa da abordagem proposta que depende de intervenção humana. Durante a realização desta etapa o avaliador é responsável por criar modelos de forma, estes modelos serão comparados aos extraídos na etapa de 'Extração de Características'.

Na ‘Análise de Alto Nível’, realiza-se a comparação dos modelos armazenados de cada quadro do vídeo com modelos fornecidos pelos avaliadores na etapa de ‘Rotulação Humana’. Esta análise permite que sejam criados gráficos de similaridade e o sumário do vídeo.

3.2 Detalhes de Desenvolvimento

Nesta seção, são descritos os detalhes de cada etapa da abordagem desenvolvida nesta pesquisa. Na subseção 3.2.1 (**Modelos Ativos de Aparência**) há a descrição do problema que é abordado pelo AAM, bem como os detalhes do funcionamento de AAM. A subseção seguinte (3.2.2 - **Processo de Treinamento**) descreve o processo de treinamento do AAM e os recursos que foram utilizados para sua realização. Na subseção 3.2.3 (**Fluxo de Execução**) é descrito como as diversas etapas evidenciadas na subseção 3.2.1 interagem durante a execução do AAM na abordagem proposta. Na subseção 3.2.4 (**Sumarização de Vídeos**) é descrita a estratégia utilizada para a sumarização dos vídeos das sessões de teste de usabilidade.

3.2.1 Modelos Ativos de Aparência

Nesta seção, descreve-se o processo de desenvolvimento do módulo de reconhecimento e rastreamento de faces que compõe a ferramenta. A técnica escolhida para abordar o problema em questão foi a modelagem ativa da aparência – AAM (*Active Appearance Models*).

No caso geral, AAM soluciona dois dos problemas iniciais da ferramenta. O primeiro é o de encontrar, a cada novo quadro do vídeo f_n da sessão de teste de usabilidade, a região em que se encontra a face do usuário, descrito pela Eq. 4. O segundo problema é o de descrever o estado da face detectada.

No caso específico do primeiro quadro do vídeo, o problema de encontrar a face presente no vídeo é realizada pelo arcabouço de programação OpenCV utilizado na concepção da abordagem.

$$F(f_n) = (x, y, w, h, \theta) \quad (4)$$

$$E(f_n, x, y, w, h, \theta) = (c_1, c_2, \dots, c_m) \quad (5)$$

O resultado da Eq. 4 corresponde à posição geral da face encontrada descrita pelo par de coordenadas espaciais (x, y) , a largura e altura da face descrita no par (w, h) e a inclinação geral da face na imagem descrita na variável θ .

A Eq. 5 permite obter um conjunto de características c ajustadas à face encontrada no quadro f_n sob análise. A forma ideal da função que resolve o problema descrito na Eq. 5 possui uma quantidade mínima de valores obtidos e é uma função inversível.

Ambas as equações supradescritas são apenas generalizações dos problemas abordados. As equações foram formuladas de forma a se adaptar melhor ao caso de uso desta dissertação e não necessariamente retratam as soluções ou mesmo as descrições, de outros autores para os mesmos problemas.

I. Modelo de Forma

Nesta seção, descreve-se o processo de desenvolvimento do modelo de forma usado. Um modelo de forma é descrito como uma sequência ordenada de pontos no plano. Nesta pesquisa, a sequência de pontos deve coincidir com pontos de interesse da face humana considerada.

Uma forma definida por n pontos ou vértices, em um espaço de k dimensões, é representada por um vetor de tamanho nk , ou seja, um vetor com nk variáveis. Em imagens tradicionais, o universo tridimensional é capturado em uma matriz bidimensional. Assim, uma forma bidimensional descrita por n pontos pode ser descrita pela Eq. 6.

$$S = [x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n] \quad (6)$$

II. Alinhamento de Dois Modelos de Forma

O alinhamento de dois modelos de forma distintos S_1 e S_2 consiste em encontrar parâmetros de transformação que modifiquem um dos modelos com o objetivo de minimizar o erro ou maximizar a similaridade entre modelos. É importante executar transformações sem modificar as estruturas internas do modelo, de forma a preservar o aspecto de cada expressão facial que o modelo está representando.

Na seção 4.1, será discutida a diferença entre medidas de similaridade que podem ser utilizadas como método de avaliação de similaridade de modelos de forma. Tais medidas formam a base da avaliação de expressões faciais deste trabalho.

Os parâmetros de translação (t_x, t_y) , a escala s e a rotação θ são adequados para o caso de expressões faciais, por serem transformações lineares que não modificam a aparência interna da expressão.

O parâmetro de translação é o mais simples de se obter. Sejam \tilde{S}_1 e \tilde{S}_2 os centróides dos modelos de forma S_1 e S_2 , obtidos a partir da Eq. 7.

$$\tilde{S} = \left(\frac{Max(x_n) - Min(x_n)}{2}, \frac{Max(y_n) - Min(y_n)}{2} \right) \quad (7)$$

A translação de ambos os modelos até a origem do sistema de coordenadas a partir da Eq. 8 é uma forma de se calcular o parâmetro de translação.

$$S_c = S - \tilde{S} \quad (8)$$

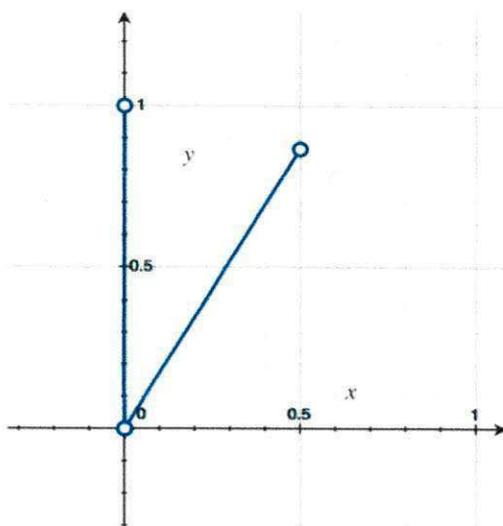
Uma vez que os modelos de forma S_{1c} e S_{2c} estão centralizados, modificar a escala dos modelos se torna uma tarefa trivial. Considerando cada par (x_n, y_n) de um modelo como um vetor, a multiplicação de cada um destes vetores por uma constante aumenta ou diminui a norma de cada vetor proporcionalmente ao valor da constante. Como o modelo está posicionado com o seu centróide na origem do sistema de coordenadas, a multiplicação de cada par (x_n, y_n) em um modelo de forma, de acordo com a Eq. 9, irá expandir ou comprimir o modelo inteiro em todas as direções.

$$S_s = s.S \quad (9)$$

De forma semelhante à estratégia que foi usada para descrever a transformação de escala, a transformação de rotação também considera cada par (x_n, y_n) de um modelo de forma como um vetor e resgata da geometria vetorial uma regra de transformação para a rotação de vetores. Uma matriz de rotação é uma matriz que, quando multiplicada por um vetor resulta em um novo vetor, de magnitude igual àquela do vetor original, mas com a direção alterada. A matriz de rotação descrita pela Eq. 10 segue as propriedades descritas. Os valores usados na Eq. 10 estão em radianos e valores positivos de θ representam rotação no sentido anti-horário com centro na origem do sistema de coordenadas (Figura 7).

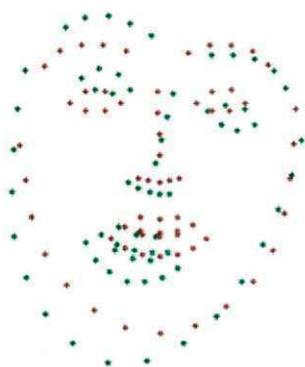
$$M(\theta) = \begin{bmatrix} \cos(\theta) & -\text{sen}(\theta) \\ \text{sen}(\theta) & \cos(\theta) \end{bmatrix} \quad (10)$$

Figura 7 – Vetor sofrendo rotação de 30°



Os parâmetros escala s e rotação θ requerem um maior esforço para o cálculo. Nesta pesquisa, adotou-se o gradiente descendente para aproximar ambos os parâmetros simultaneamente. Experimentos preliminares mostraram que a otimização isolada e sequencial destas variáveis não é suficiente para aproximar valores ótimos (Figura 8).

Figura 8 – Modelos de forma sendo otimizados com relação à escala e rotação: (a) ajuste isolado dos parâmetros; e (b) ajuste simultâneo dos parâmetros. Os pontos em vermelho são o modelo alvo e os pontos em verde são o resultado da aproximação



(a)



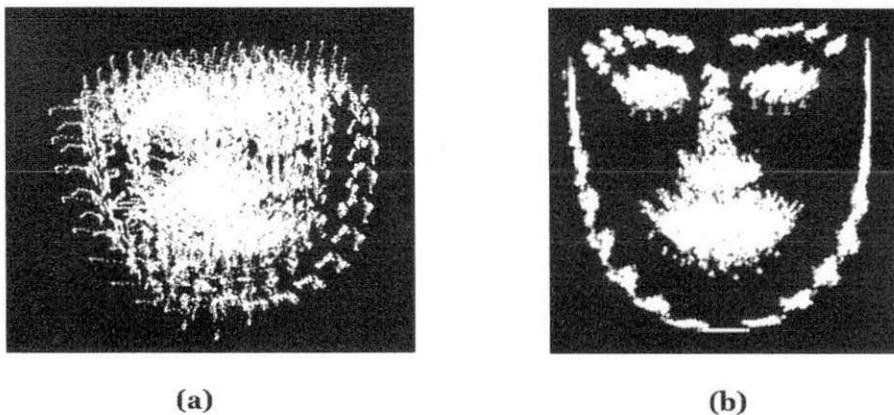
(b)

III. Treinamento do Modelo de Forma

O propósito do processo descrito na subseção II é minimizar a variação das coordenadas espaciais dos vértices dos modelos com relação a propriedades que podem ser replicadas de maneira confiável, e.g., translação, escala e rotação.

O processo descrito na subseção II deve ser executado para os N modelos de forma, a partir de um mesmo modelo-base que deve ser aproximado (Figura 9). Na teoria, qualquer modelo poderia ser adotado. Entretanto, é uma boa prática empregar um modelo que reduza sua diferença em relação a qualquer outro do espaço de treinamento, para evitar que o modelo médio final seja similar demais a um caso específico. O uso de um modelo inadequado faz com que o processo de convergência seja mais demorado, no caso geral.

Figura 9 – Conjunto de modelos sendo alinhados: (a) modelos não alinhados; e (b) modelos alinhados



Para garantir que não será usado um modelo inadequado, e.g., um modelo de expressão final, pode ser gerado o modelo médio com base nos modelos de que já se dispõe. Considerando-se um espaço amostral de N modelos de forma $S_i: i = 1, 2, \dots, N$, o modelo médio \bar{S} pode ser obtido pela Eq. 11.

$$\bar{S} = \frac{1}{N} \sum_{i=1}^N S_i \quad (11)$$

Se for subtraído o modelo médio de cada modelo do espaço amostral, o conjunto de modelos passa a ser descrito em função do modelo médio \bar{S} . É possível imaginar o conjunto total de modelos como uma grande matriz B_s de valores, na qual cada linha contém os valores de cada modelo subtraído do modelo médio, conforme descrito na Eq. 12.

$$B_s = \begin{bmatrix} x_1 1 & x_2 1 & \dots & x_n 1 & y_1 1 & \dots & y_n 1 \\ x_1 2 & x_2 2 & \dots & x_n 2 & y_1 2 & \dots & y_n 2 \\ \vdots & \vdots & & \ddots & \vdots & & \vdots \\ x_1 N & x_2 N & & \dots & & & y_n N \end{bmatrix} \quad (12)$$

Qualquer modelo original S_i pode ser obtido novamente, multiplicando o vetor $W_{S_i} = [0, \dots, 1, \dots, 0]$ (no qual o valor 1 está na posição i para corresponder ao modelo S_i) por B_s e, em seguida, somando com \bar{S} . A forma genérica do vetor W_{S_i} é dada na Eq. 13 (no caso sendo descrito $k = N$).

$$W_s = [w_{s1}, w_{s2}, \dots, w_{sk}] \quad (13)$$

Usando esta forma da equação e considerando que cada valor p_t pode assumir qualquer valor real, não só é possível obter qualquer modelo original, como também passa a ser possível aproximar novos modelos que se assemelhem a algum modelo do espaço amostral original. Na Eq. 14 sumaria-se o processo descrito neste parágrafo.

$$\hat{S} \approx \bar{S} + W_s B_s \quad (14)$$

Mesmo já sendo funcional, o procedimento descrito até este ponto ainda é muito lento em casos reais de uso. A quantidade de modelos necessários para que seja possível aproximar novos modelos com confiança aceitável é muito grande. Assim sendo, o processo de encontrar os melhores valores do vetor W_s cresce exponencialmente.

Ainda mais, em um espaço amostral grande, haverá modelos que podem ser obtidos, se não aproximados com muita precisão, dos outros modelos já incluídos no espaço amostral. Este excesso de modelos deve ser reduzido ao mínimo necessário para aproximar todos os modelos do espaço amostral.

PCA é uma técnica estatística que permite reduzir a complexidade de problemas que envolvem muitas variáveis possivelmente correlacionadas. PCA usa transformações ortogonais para identificar componentes linearmente não correlacionados, denominados **componentes principais**. Nesse procedimento, consideram-se os dados do espaço amostral como vetores multidimensionais e busca direções nos vetores que expliquem o máximo da variância observada no espaço amostral. A quantidade de componentes principais obtidos é menor ou igual à quantidade de observações originais no espaço amostral N . A redução na dimensão é obtida guardando-se os vetores que refletem a maior variação dos dados e expressando os componentes restantes em função dos componentes principais já encontrados. Cada novo vetor é, por sua vez, escolhido de forma a expressar o máximo da variância do espaço amostral, sob a condição de que seja ortogonal aos componentes anteriores.

IV. Modelo de Textura

O modelo de forma descrito nas seções I, II e III é suficiente apenas para ajustar o modelo de forma criado em novos modelos de forma. Entretanto, é necessário adaptar o modelo forma do AAM a uma imagem desconhecida, na qual não se sabe onde estão os pontos de interesse que compõem o modelo de forma. Nesta subseção, descreve-se o processo que resulta na construção de um modelo estatístico de textura que é usado como referência para o ajuste do modelo de forma.

Geralmente, uma textura define a qualidade táctil de uma superfície. Em Computação Gráfica, uma imagem passa a ser uma textura quando é carregada em um programa e aplicada sobre a superfície de um polígono.

Neste documento, uma textura é considerada como sendo uma sequência ordenada de intensidades de pixels aplicadas sobre o modelo de forma. Uma textura constituída de m pixels será representada pelo vetor expresso na Eq. 15.

$$T = [p_1, p_2, \dots, p_m] \quad (15)$$

De forma semelhante ao processo que foi usado para conceber o modelo de forma, o modelo de textura requer que sejam eliminadas diferenças entre os posicionamentos dos pixels que estão dentro dos limites do modelo de forma e compondo a textura do modelo AAM.

Na subseção II, foi descrito o processo de alinhamento de dois modelos de forma distintos, cujo intuito foi reduzir a variabilidade do modelo com relação ao posicionamento, tamanho e rotação do modelo. De forma semelhante, é preciso que o modelo de textura seja tolerante à variação nas posições dos pixels sob comparação, i.e., um pixel que em uma imagem seja parte de um olho humano esquerdo deve sempre ser comparado com o pixel correspondente do modelo de textura, independentemente da posição do pixel relativo à imagem.

V. Triangulação do Modelo de Forma

Na subseção I, omitiu-se uma noção familiar a toda forma geométrica: a noção de vértice, sem nenhuma menção à conexão entre vértices, uma vez que a noção de vértice era suficiente para explicar o processo de modelagem de formas.

Denomina-se *triangulação* ao processo de definir uma forma geométrica k -dimensional como um conjunto *maximal*⁸ de elementos k -simplex, de tal forma que quaisquer dois elementos k -simplex tenha apenas uma face em comum, ou seja, nenhum elemento k -simplex do conjunto cruze com outro daquele conjunto. Simplex é uma generalização do conceito de triângulo em espaços k -dimensionais, uma forma geométrica de $(k + 1)$ pontos independentes no espaço k -dimensional. No espaço bidimensional, a forma 2-simplex define um triângulo.

Triangulação de conjuntos de pontos $P \subset \mathbb{R}^{k>0}$ é um caso particular da triangulação geométrica, no qual os vértices dos elementos k -simplex que compõem a triangulação coincidem com P . Há dois algoritmos bem conhecidos que podem ser empregados para solucionar o problema da triangulação de conjuntos de pontos definidos no \mathbb{R}^2 : (i) a *triangulação de peso mínimo* (AICHHOLZER ET AL., 2009); e (ii) a *triangulação de Delaunay* (GUIBAS; RUSSEL, 2004).

No algoritmo de *triangulação de peso mínimo*, busca-se criar uma solução de tal forma que a soma dos comprimentos das arestas que formam a triangulação é mínima. Por outro lado, na *triangulação de Delaunay*, cria-se uma solução que maximiza o menor ângulo de todos os triângulos na triangulação.

A técnica de triangulação de peso mínimo é uma solução adequada, por exemplo, ao roteamento de redes de malha, pois haverá maior garantia de conectividade ou em redes de larga escala, conectadas por cabos, nas quais a conservação de materiais pode ser um fator crítico. O problema para a aplicação desta técnica na triangulação dos modelos, de forma AAM é a ocorrência de triângulos muito pequenos e com ângulos muito fechados. Nesses casos, o modelo perde precisão na região descrita pelo triângulo, haverá menos pixels para ajustar e manter o modelo na região correta. Esta propriedade é exatamente o que fez com que a triangulação de Delaunay fosse a técnica aplicada nesta pesquisa.

Triangulação é uma ferramenta útil em muitos cenários e foi extensivamente estudada, resultando na criação de vários algoritmos que aproximam a triangulação de Delaunay. A triangulação de Delaunay para um

⁸ Um conjunto é maximal em relação a uma determinada propriedade se ele não é subconjunto próprio de outro conjunto com a mesma propriedade.

conjunto de pontos $P \subset \mathbb{R}^2$ pode ser aproximada com uma triangulação $T(P)$, de tal forma que não existe $p \in P$ dentro de nenhum círculo circunscrito a nenhum dos elementos de $T(P)$ (GUIBAS; RUSSEL, 2004).

O determinante da matriz dada na Eq. 16 indica se a regra que aproxima a triangulação de Delaunay está sendo obedecida no triângulo $A, B, C \in P$, com relação ao ponto $D \in P$. Entretanto, é necessário que A, B e C estejam dispostos em ordem anti-horária no círculo circunscrito descrito pelos três pontos. Neste caso, o determinante positivo indica que o ponto D está ferindo a regra. A Eq. 17 ajuda a determinar se os pontos A, B e C estão em sequência horária, anti-horária ou são colineares. Quando o resultado da Eq. 17 for maior que zero, significa que os pontos estão em sequência anti-horária.

$$\det \begin{pmatrix} A_x & A_y & A_x^2 + A_y^2 & 1 \\ B_x & B_y & B_x^2 + B_y^2 & 1 \\ C_x & C_y & C_x^2 + C_y^2 & 1 \\ D_x & D_y & D_x^2 + D_y^2 & 1 \end{pmatrix} \quad (16)$$

$$= \det \begin{pmatrix} A_x - D_x & A_y - D_y & (A_x^2 - D_x^2) + (A_y^2 - D_y^2) \\ B_x - D_x & B_y - D_y & (B_x^2 - D_x^2) + (B_y^2 - D_y^2) \\ C_x - D_x & C_y - D_y & (C_x^2 - D_x^2) + (C_y^2 - D_y^2) \end{pmatrix}$$

$$\det \begin{pmatrix} B_x - A_x & B_y - A_y \\ C_x - A_x & C_y - A_y \end{pmatrix} \quad (17)$$

$$= (B_x - A_x)(C_y - A_y) - (B_y - A_y)(C_x - A_x)$$

O algoritmo utilizado nesta pesquisa é o modelo mais simples que aproxima a triangulação de Delaunay em tempo $O(n^3)$: $n = |P|$ (GUIBAS; RUSSEL, 2004). Esta etapa do processo só será executada uma única vez durante todo o processo de treinamento do modelo. Portanto, o tempo que este algoritmo leva para completar seu processamento não é crítico para o desempenho final da ferramenta ou do modelo.

VI. Mapeamento de Textura

Na subseção IV, foi dito que o modelo de textura é construído de forma semelhante ao modelo de forma, no sentido de que é preciso eliminar as diferenças causadas pela movimentação do modelo. A etapa de triangulação descrita na subseção V é crucial para que este objetivo seja alcançado. A etapa serve para criar os triângulos nos quais serão descritos todos os pixels do modelo.

Coordenadas baricêntricas (COXETER, 1989) oferecem uma maneira de descrever um ponto qualquer no espaço como função da posição relativa do ponto com relação aos vértices que definem um simplex qualquer do espaço. Cada pixel que está dentro do modelo de forma passa então a ser descrito pelas coordenadas baricêntricas do pixel com relação ao triângulo no qual o pixel está posicionado.

No plano, um ponto qualquer (x, y) pode ser descrito em função das posições dos vértices de um triângulo qualquer $\{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$ e das coordenadas baricêntricas α, β e γ do ponto com relação ao triângulo, de acordo com o sistema dado na Eq. 18.

$$\begin{cases} \alpha x_1 + \beta x_2 + \gamma x_3 = x, \\ \alpha y_1 + \beta y_2 + \gamma y_3 = y, \\ \alpha + \beta + \gamma = 1 \end{cases} \quad (18)$$

Dado qualquer ponto no espaço e um triângulo qualquer, definir as coordenadas baricêntricas do ponto com relação ao triângulo passa a ser uma questão de solucionar o sistema de equações da Eq. 18. A solução do sistema é dada na Eq. 19.

$$\begin{aligned} \alpha &= \frac{xy_1 + yx_3 + x_1y_3 - xy_3 - yx_1 - x_3y_1}{x_1y_3 + x_2y_1 + x_3y_2 - y_1x_3 - y_2x_1 - y_3x_2} \\ \beta &= \frac{xy_2 + x_1y + x_2y_1 - xy_1 - x_1y_2 - x_2y}{x_1y_3 + x_2y_1 + x_3y_2 - x_1y_2 - x_2y_3 - x_3y_1} \\ \gamma &= 1 - \alpha - \beta \end{aligned} \quad (19)$$

Definir se um ponto qualquer está dentro de um determinado triângulo também é um processo simples que depende das coordenadas baricêntricas do ponto com relação ao triângulo. Um ponto p está dentro do triângulo se, e somente se, $0 \leq \alpha, \beta, \gamma, \leq 1$, sendo α, β e γ as coordenadas baricêntricas do ponto p com relação ao triângulo.

Para cada pixel dentro do modelo de forma médio é registrado o triângulo no qual o pixel está contido e as coordenadas baricêntricas do pixel, com relação ao triângulo. Finalmente, o mapeamento criado é usado para retirar de todas as imagens do conjunto de treinamento os respectivos pixels correspondentes aos pixels do modelo médio. Desta forma, é criado um conjunto de modelos de textura de mesmo tamanho, nos quais cada elemento $p_i \in T$ tem exatamente um novo elemento $p'_i \in T'$ na mesma posição i de ambos os vetores T e T' .

VII. Normalização do Modelo de Textura

Os processos descritos nas seções V e VI são usados para definir vetores de pixels do mesmo tamanho, independentemente do modo como o modelo de forma esteja orientado. Neste estágio, é possível retirar cada pixel das imagens anotadas por modelos de forma e garantir que cada posição i em um modelo de textura corresponda à mesma posição em outro modelo.

Há, porém, um último problema no modelo de textura que precisa ser eliminado. A diferença de iluminação criada em vídeos de diferentes pessoas, sob diferentes condições de filmagem, pode ser reduzida a partir de um processo rápido de normalização. Mesmo sabendo que o processo de PCA aplicado ao conjunto de treinamento também é capaz de aproximar esta variação, é uma boa prática eliminar qualquer variação no conjunto de texturas que é explicado a partir de outras transformações.

Nesta pesquisa, a normalização de cada pixel p_i do modelo de textura $T = [p_1, p_2, \dots, p_m]$ é obtida a partir do resultado da função dada na Eq. 20. A partir deste processo, os valores em T' (resultado da normalização de T) estarão no intervalo fechado $[0, 1]$.

$$p'_i = \frac{p_i - \min(T)}{\max(T) - \min(T)} \quad (20)$$

VIII. Treinamento do Modelo de Textura

O treinamento do modelo de textura segue um processo análogo àquele descrito na subseção III para o treinamento do modelo de forma. Da mesma forma, as etapas descritas nas subseções V, VI e VII servem para reduzir a variação que a textura do modelo sofre sempre que o modelo de forma se altera e a variação introduzida por diferentes condições no momento da captura das imagens.

Os processos de mapeamento e normalização da textura são executados em todas as N imagens utilizadas no treinamento, tomando como base o modelo de forma médio (subseção III). Similarmente ao modelo de forma, o uso de um modelo diferente do modelo médio resulta em modelos de textura que necessitam de uma quantidade maior de iterações para aproximar a imagem sob avaliação pelo modelo.

Após a execução dos processos descritos nas subseção VI e VII, os modelos de textura retirados de cada imagem do conjunto de treinamento têm o mesmo tamanho e cada pixel do modelo pode ser comparados diretamente. Considerando um espaço amostral de N modelos de textura $T_i: i = 1, 2, \dots, N$,

obtidos dos mesmos N modelos de forma, conforme descrito na subseção III, o modelo de textura médio \bar{T} (Figura 10) é obtido pela Eq.21, analogamente à Eq. 11 que permite obter o modelo de forma médio.

$$\bar{T} = \frac{1}{N} \sum_{i=1}^N T_i \quad (21)$$

Figura 10 – Reconstrução do modelo de textura médio



O processo continua, conforme descrito na obtenção do modelo de forma. Se o modelo médio for subtraído de cada modelo no espaço amostral, o conjunto de modelos passa a ser descrito em função do modelo de textura médio \bar{T} . A matriz descrita na Eq. 12 tem sua contraparte para o modelo de textura dada pela matriz B_t (Eq. 22).

$$B_t = \begin{bmatrix} t_11 & t_21 & \dots & t_m1 \\ t_12 & t_22 & \dots & t_m2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1N & t_2N & \dots & t_mN \end{bmatrix} \quad (22)$$

Da mesma forma, é possível aproximar novos modelos de textura ou qualquer modelo original a partir das Eqs. 23 e 24. O mesmo processo de PCA, descrito na subseção III, é aplicado a B_t , de forma a reduzir a complexidade do problema de otimização do vetor W_t .

$$W_t = [w_{t1}, w_{t2}, \dots, w_{tL}] \quad (23)$$

$$\hat{T} \approx \bar{T} + W_t B_t \quad (24)$$

3.2.2 Processo de Treinamento

Nesta pesquisa, utilizou-se o banco Cohn-Kanade de expressões faciais (ver Anexo A). Para o treinamento do modelo AAM descrito neste capítulo foram manualmente anotadas 1950 imagens dentre as 8795 existentes no banco de imagens Cohn-Kanade. As imagens anotadas foram escolhidas de forma aleatória. Exemplos de imagens anotadas manualmente encontram-se no Anexo A.

A anotação das imagens que fizeram parte do treinamento do modelo foi realizada de forma totalmente manual. Para cada imagem de treinamento, fez-se a marcação de todos os pontos que o modelo iria aproximar. Este processo de anotação deve ser realizado com precisão. Para facilitar a identificação da posição correta dos pontos do modelo, é preciso que a posição de cada ponto seja descrita de forma clara e que sejam posicionados em locais de fácil discriminação.

De posse de um conjunto de imagens anotadas, dá-se início ao processo de treinamento do modelo. A primeira etapa do processo de treinamento é o alinhamento dos modelos anotados. Conforme descrito na subseção II, este processo remove a variação dos modelos associada à rotação, escala e translação do modelo na imagem.

Uma vez que todos os modelos tiverem sido alinhados a um modelo comum, tem início o processo descrito na subseção III, a partir do qual serão determinados os parâmetros de transformação de forma.

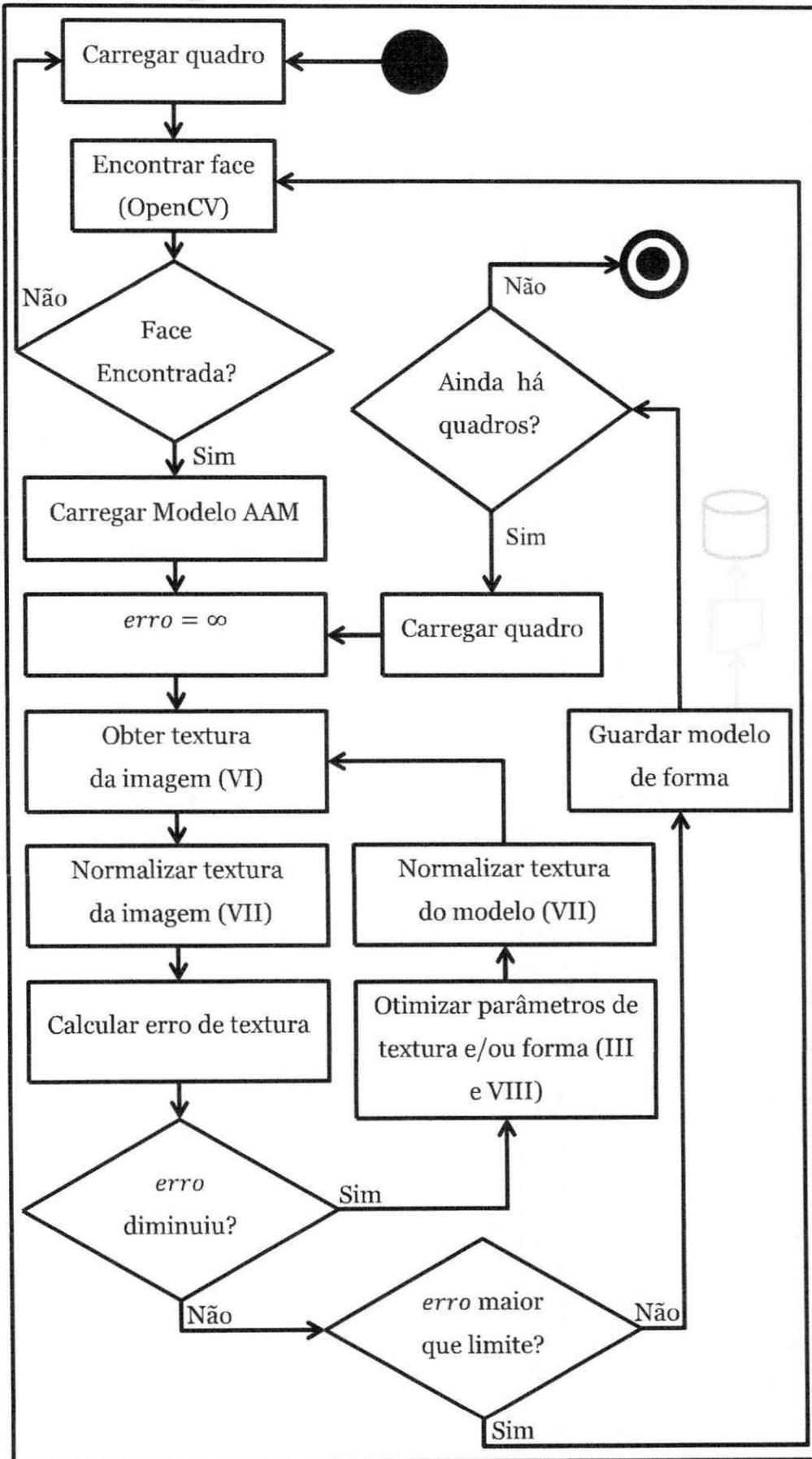
Em seguida, tem início o processo de treinamento do modelo de textura. O processo de mapeamento de textura, descrito na subseção VI, é análogo ao processo de alinhamento dos modelos de forma, descrito na subseção II. Assim, o processo de mapeamento de textura ordena os valores dos pixels, de acordo com uma posição universal, que não vai se alterar independentemente da variação do modelo de forma. O processo de normalização de textura (seção VII) é, então, aplicado a todas as texturas extraídas.

Finalmente, o processo de treinamento do modelo de textura (seção VIII) pode ser executado, resultando em um vetor de parâmetros de transformação para a textura.

3.2.3 Fluxo de Execução

Na Figura 11, representa-se o fluxo de execução da ferramenta desenvolvida durante esta pesquisa. No fluxograma exibido na Figura 11 é possível observar o processo de tomada de decisão e os passos executados pela ferramenta relacionados às seções descritas neste capítulo.

Figura 11 – Fluxograma de funcionamento do sistema AAM em vídeos



Em comparação com a Figura 6 (que exibe a arquitetura simplificada da abordagem), o fluxograma descrito na Figura 11 é responsável pelas três primeiras etapas do fluxo apresentado na Figura 6, a saber: (i) Segmentação de Movimento; (ii) Detecção de Faces e FCP ; e (iii) Extração de Características.

3.2.4 Sumarização de Vídeos

O processo de sumarização adotado nesta pesquisa é baseado na similaridade entre modelos de forma. Para a extração de um sumário dos vídeos de sessões de teste de usabilidade, a abordagem proposta nesta pesquisa exige que seja fornecido um modelo base para a comparação durante a análise do vídeo. Este modelo base pode representar qualquer expressão.

Uma vez fornecido ao menos um modelo de comparação, este será comparado a cada modelo extraído de cada quadro do vídeo da sessão de teste. As medidas de comparação adotadas nesta pesquisa serão descritas na seção 4.1 deste documento. O resultado da comparação entre quaisquer pares de modelos de forma, independentemente da medida adotada (dentre aquelas estudadas) é um número real que, teoricamente, varia no intervalo $[0, \infty[$.

Como resultado da comparação sequencial de um modelo de comparação com cada modelo extraído do vídeo, é gerado um gráfico, cujos valores representam a diferença entre os modelos. Para reduzir a complexidade do gráfico, é aplicado o processo de normalização dos valores obtidos para o intervalo $[0, 1]$.

O gráfico resultante pode ser usado como sumário estático do vídeo. O mesmo gráfico é também usado para a geração de um sumário dinâmico. A abordagem desenvolvida nesta pesquisa permite que o usuário defina um valor λ no intervalo $[0, 1]$ que é usado como limiar de corte no gráfico de sumarização. A abordagem permite ainda que o usuário escolha se o corte será das partes mais similares ($< \lambda$) ou mais distintas ($> \lambda$) do modelo de comparação.

Outra forma de sumarização estática é utilizada para a análise e comparação dos dados obtidos no Experimento 3, descrito na seção 4.3. Para representar a evolução geral das expressões faciais, de acordo com cada tarefa

executada no decorrer da sessão de teste, o gráfico obtido da normalização é separado em seções, de tal forma que cada seção inicia e termina junto com cada tarefa, então é extraída a média dos valores para a seção. Os valores obtidos mostram a distância média de todos os modelos de expressão para cada tarefa.

Em termos da descrição da Figura 6 (que exibe uma arquitetura simplificada da abordagem), a estratégia de sumarização descrita nesta seção utiliza a comparação dos dados armazenados da análise do vídeo (BD), com aqueles relativos a um modelo fornecido, produzidos por rotulação manual, para realizar a ‘Análise de Alto Nível’ e obter o diagnóstico gráfico e a sumarização.

3.3 Detalhes de Implementação

Para a validação da abordagem proposta apresentada neste capítulo, foi necessário desenvolver um sistema para a análise automática de expressões faciais em vídeos de sessões de teste de usabilidade. A implementação do sistema seguiu as etapas descritas e o fluxo de execução apresentado e discutido. Nesta seção, são apresentados detalhes de implementação do software desenvolvido para a validação da abordagem descrita, assim como para a realização dos testes descritos no Capítulo 4 (**Avaliação Experimental**).

Como ambiente de desenvolvimento, foi utilizado o sistema operacional *OSX* nas versões 10.6 e 10.8. O ambiente de desenvolvimento integrado empregado para o desenvolvimento foi o *Xcode*⁹ nas versões 4.3 e 4.5. O sistema é compilado com o *Apple LLVM*¹⁰ 4.1 (*Low Level Virtual Machine*).

A implementação foi realizada utilizando as linguagens de programação *C/C++* e *Objective C*¹¹. A linguagem *C/C++* foi escolhida para o desenvolvimento do arcabouço do sistema, dentre vários aspectos, por: (i) ser

⁹ <http://developer.apple.com/xcode/>

¹⁰ <http://llvm.org>

¹¹ <http://developer.apple.com/documentation/Cocoa/Conceptual/ObjectiveC/>

capaz de gerar programas com maior desempenho computacional em relação a outras linguagens (e.g. linguagens interpretadas ou semi-interpretadas); (ii) ser muito usada na área de Visão Computacional; e (iii) dispor de bibliotecas tais como a *Intel OpenCV*¹² (empregada nesta pesquisa), dotada de facilidades destinadas ao suporte no desenvolvimento deste trabalho.

A biblioteca *Intel OpenCV*, trata dos aspectos relacionados ao processamento de imagens digitais, oferecendo desde as estruturas de dados fundamentais que representam vídeos, imagens e cores até rotinas de mais alto nível (e.g. filtros de cor).

A linguagem de programação C/C++ foi utilizada em conjunto com a biblioteca *Intel OpenCV* para a implementação de todas as etapas descritas anteriormente neste capítulo. A linguagem de programação *Objective C* foi utilizada na implementação da interface gráfica com a utilização o arcabouço de desenvolvimento *cocoa*¹³ e *core Image*¹⁴ da *Apple*.

Há, no entanto, desvantagens no uso desta combinação de linguagens e bibliotecas, tais como a dependência em relação ao sistema operacional *OSX* (para a interface gráfica), problemas de gerenciamento de memória e apontadores entre linguagens e peculiaridades do desenvolvimento para o sistema operacional.

¹² <http://opencv.org>

¹³ <http://developer.apple.com/cocoa/>

¹⁴ <http://developer.apple.com/graphicsimaging/Conceptual/CoreImaging/>

4 Avaliação Experimental

Neste capítulo, são apresentados experimentos envolvendo os diversos aspectos estudados nesta pesquisa. O primeiro experimento teve como objetivo comparar 3 medidas de similaridade e definir qual das medidas estudadas é mais adequada para ser empregada pela abordagem proposta. Para o segundo experimento, a medida de similaridade selecionada foi empregada na comparação de modelos de expressões faciais, com o propósito de verificar a capacidade da abordagem proposta em discernir entre diferentes expressões faciais. Por último, foi realizado um experimento envolvendo os vídeos de sessões de um teste de usabilidade conduzido no decorrer desta pesquisa. O objetivo deste experimento foi correlacionar a taxa de aparição de expressões faciais em um vídeo de sessão de teste de usabilidade com a quantidade de eventos negativos observados no decorrer de cada sessão de teste.

4.1 Experimento 1 – Busca pela Melhor Medida de Similaridade de Modelos de Forma

O experimento inicial realizado nesta pesquisa serviu para dois propósitos. O primeiro foi verificar a qualidade média do sistema de comparação de modelos de forma. Conforme discutido no Capítulo 3, a medida de similaridade ou dissimilaridade entre modelos de forma é a base utilizada nesta pesquisa para obter a sumarização dos vídeos de sessões de usabilidade. O segundo propósito do experimento foi determinar qual a medida de similaridade que se afigurava mais adequada para a aplicação desenvolvida.

Neste experimento inicial, é apresentada uma comparação objetiva de três diferentes medidas de similaridade de vetores, aplicáveis a modelos de forma, a saber: (i) *Distância Média Entre Pontos* (ROGERS; GRAHAM, 2001); (ii) *Distância de Procrustes* (ANIKEENKO, 2006); e (iii) *Erro Médio Quadrático* (LEHMANN; CASELLA; OLKIN, 1998). As medidas foram escolhidas por ter sido verificado, na literatura, que as medidas são empregadas com alguma frequência. A análise e comparação entre as medidas consideradas é feita por meio de uma análise estatística descritiva.

Para a realização deste experimento, empregou-se a base de imagens de expressões faciais Cohn-Kanade (*Cohn-Kanade AU-coded Facial Expression Database*), a qual contém um total de 8795 imagens, caracterizando 490 sequências de quadros que descrevem a evolução de expressões faciais. Na Figura 12, ilustra-se uma das sequências de imagens contidas no banco. Sequências adicionais podem ser encontradas no Anexo A.

A *Distância Média Ponto a Ponto* (Eq. 25) é definida como sendo a razão do somatório das distâncias de todos os pares de pontos pela quantidade de pontos presentes no modelo.

$$MP2PD(A, B) = \frac{1}{n} \sum_{i=1}^n \sqrt{(ax_i - bx_i)^2 + (ay_i - by_i)^2}, \quad (25)$$

em que A e B são modelos de forma, conforme definidos na Eq. 6 (seção I). A quantidade de pontos descritos em ambos os modelos é definida pela variável n . Nos modelos A e B , ax_i e bx_i correspondem aos valores da coordenada x dos pontos no índice i dos modelos de forma A e B , respectivamente. De maneira semelhante, as variáveis expressas em ay_i e by_i representam os valores das coordenadas y dos mesmos pontos no índice i .

A *Distância de Procrustes* (Eq. 26) é definida como sendo a raiz quadrada do somatório dos quadrados das distâncias entre pares de pontos correspondentes.

$$Pro(A, B) = \sqrt{\sum_{i=1}^n (ax_i - bx_i)^2 + (ay_i - by_i)^2} \quad (26)$$

Ambas as medidas estão descritas no espaço bidimensional, mas podem ser expandidas para qualquer espaço k -dimensional.

O *Erro Médio Quadrático* (Eq. 27) é a medida padrão para a comparação de sequências pareadas de valores, sendo definida como a razão do somatório do quadrado das diferenças entre pares de valores pela quantidade de valores considerada. Na Eq. 27, a variável m representa a quantidade total de valores existente nos modelos de forma A e B , esta quantidade corresponde ao dobro da quantidade de pontos nos modelos. Os valores são representados nas variáveis a_i e b_i e correspondem às coordenadas espaciais dos pontos representados nos modelos sem distinção entre as coordenadas x e y .

$$MSE(A, B) = \frac{1}{m} \sum_{i=1}^m (a_i - b_i)^2 \quad (27)$$

Na Eq. 6 (seção I), define-se um modelo de uma forma a partir da sequência de coordenadas x da forma, seguida da sequência de suas coordenadas y . A medida de Erro Médio Quadrático não faz distinção entre os valores de x e y dentro do modelo de forma. Novamente, o valor de m na equação do Erro Médio Quadrático corresponde ao dobro da quantidade de pontos n no modelo de forma. Assim sendo, é possível reescrever a Eq. 27 em função de n como descrito na Eq. 28.

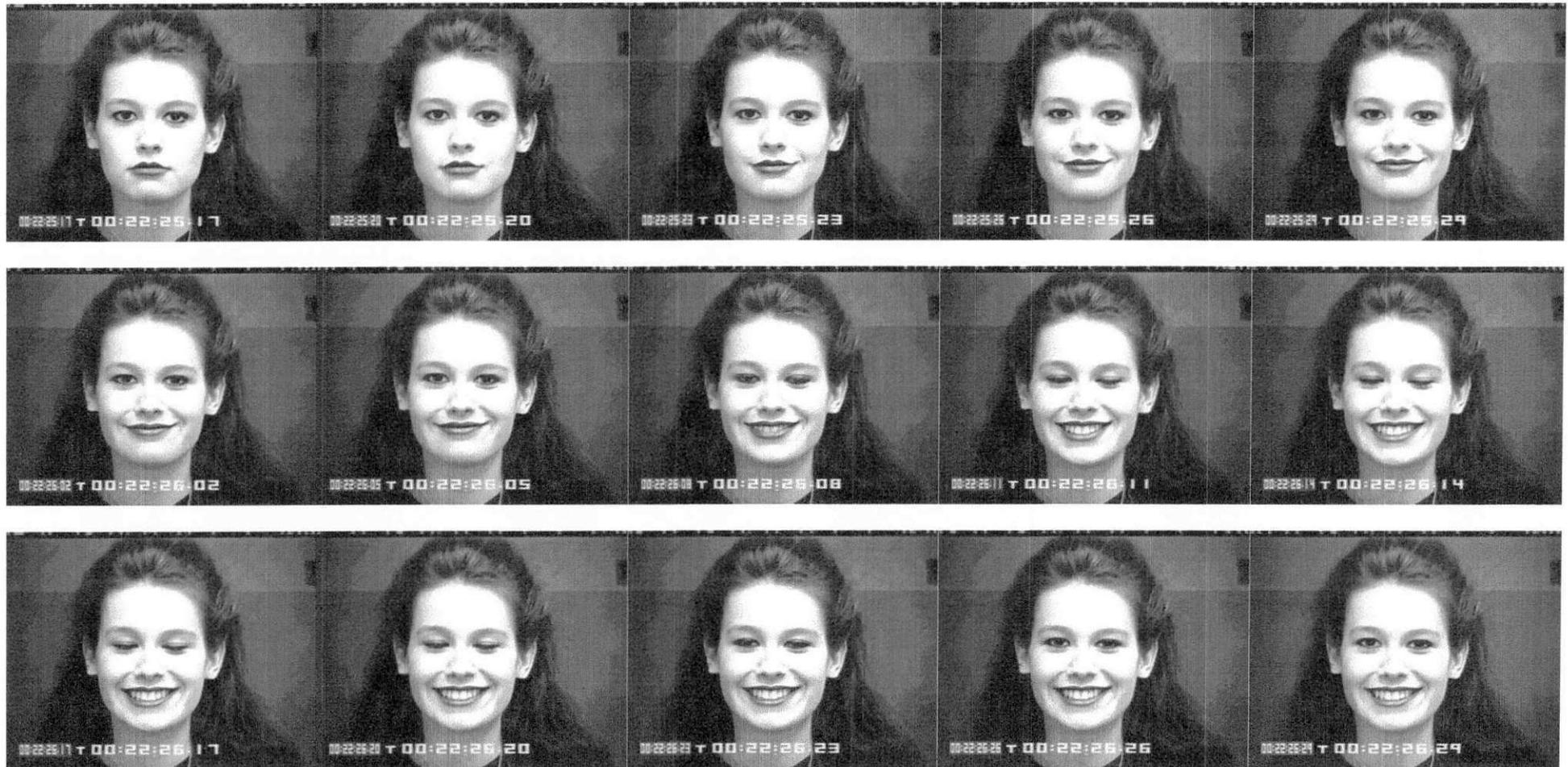
$$MSE(A, B) = \frac{1}{n} \left(\sum_{i=1}^n (ax_i - bx_i)^2 + \sum_{i=1}^n (ay_i - by_i)^2 \right) = \frac{1}{n} \left(\sum_{i=1}^n (ax_i - bx_i)^2 + (ay_i - by_i)^2 \right) \quad (28)$$

Há uma forte relação entre a medida da Distância de Procrustes e a medida do Erro Quadrático Médio, explicada na Eq. 29. Os termos do somatório de ambas as medidas resulta no mesmo valor. A única alteração está em se ter que obter a média do somatório ou extrair a raiz quadrada do valor do somatório.

$$MSE(A, B) = \frac{1}{n} Pro(A, B)^2 \quad (29)$$

A primeira etapa do Experimento 1 possibilitou verificar a relação entre as medidas descritas nas sequências de quadros que representam a evolução de uma expressão facial, a partir da diferença estatística entre as médias, variâncias e desvios padrão das medidas. Verificou-se a correlação de Pearson, o Teste-F de variância e o Teste-T de médias.

Figura 12 – Exemplo de sequência de imagens de uma expressão facial da base de imagens de expressões faciais Cohn-Kanade ©Jeffrey Cohn



Nesta etapa, todas as imagens do banco de expressões faciais Cohn-Kanade foram ajustadas utilizando-se o modelo AAM desenvolvido nesta pesquisa. Os modelos de comparação empregados na a geração dos gráficos desta etapa foram modelos neutros, i.e., modelos do primeiro quadro de uma sequência de expressão facial, comparados com outras sequências de expressões diferentes.

Na etapa final deste experimento, foram utilizados apenas os quadros iniciais e finais como referência para o teste (Figura 13). Para este experimento, a primeira imagem de cada sequência de imagens que descrevia a evolução de uma expressão facial foi considerada a imagem de uma expressão facial neutra. A imagem final de cada sequência foi considerada como uma expressão facial final de uma classe qualquer. Estas considerações estão alinhadas com as diretrizes de construção da base de imagens de expressões faciais de Cohn-Kanade. Não foi feita nenhuma distinção de alto nível quanto à semântica de cada expressão.

Figura 13 – Exemplo de expressão neutra e expressão final da base de imagens de expressões faciais Cohn-Kanade ©Jeffrey Cohn: (a) expressão neutra; e (b) expressão final

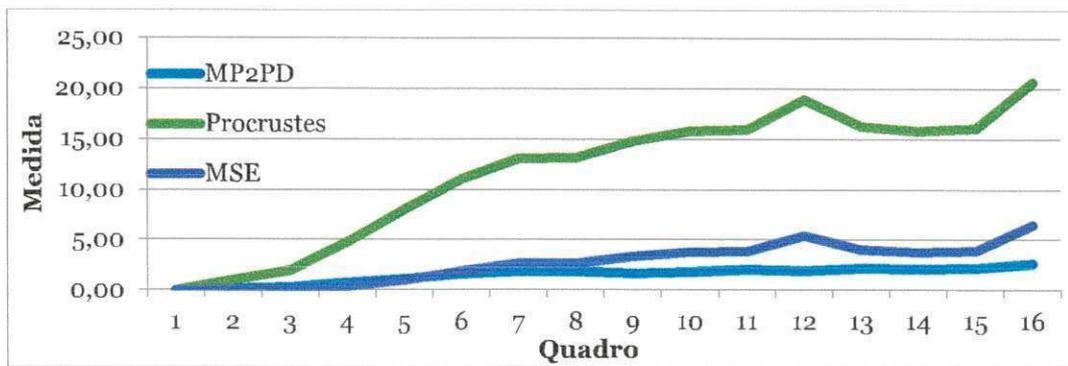


Nesta etapa, computaram-se as diferenças entre os modelos neutros e outros modelos neutros, bem como entre neutros e modelos finais de expressão. O objetivo foi gerar gráficos de distribuição de probabilidade que demonstrassem a distância entre modelos neutros e modelos de expressão final. Com estes gráficos, foi possível verificar a qualidade média da abordagem desenvolvida. A abordagem terá uma qualidade mínima aceitável se for observada a capacidade da ferramenta de distinguir a existência de uma expressão facial, qualquer que seja.

4.1.1 Resultados do Experimento 1

Na Figura 14, exemplifica-se uma evolução da expressão facial de uma das seqüências do banco de imagens utilizado neste experimento. O gráfico mostra a diferença existente na dimensão de cada uma das medidas de similaridade escolhidas para este experimento. O gráfico da Figura 14 foi criado a partir da comparação entre o modelo inicial da seqüência e cada quadro subsequente.

Figura 14 – Gráfico de evolução de uma expressão facial



Os Quadros 1 e 2 resultam dos dados obtidos para todas as seqüências existentes no banco de imagens utilizado neste experimento. Cada valor descrito em ambos os quadros é uma média dos valores observados em cada seqüência de quadros (um total de 8305 comparações, em 490 seqüências).

Quadro 1 – Valor geral da média, desvio padrão e variância para as medidas MP2PD, Distância de Procrustes e MSE

	MP2PD	Procrustes	MSE
Média	1.508093438	11.70187725	2.687799775
Desvio Padrão	0.864408443	6.965321724	2.036108358
Variância	0.747201957	48.51570672	4.145737247

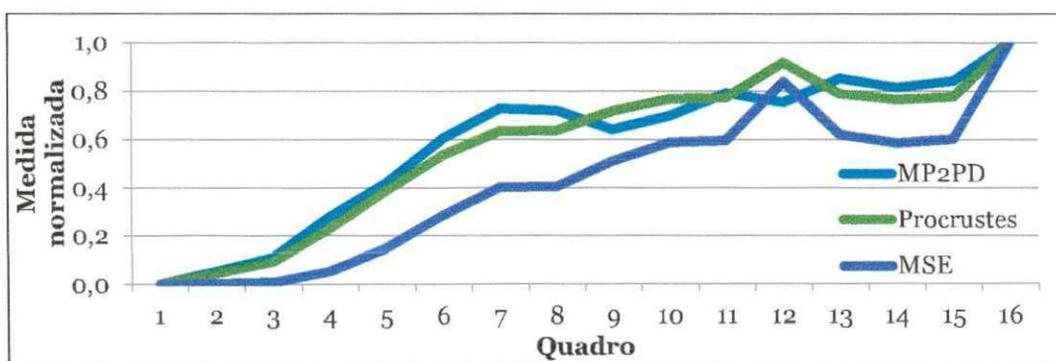
Quadro 2 – Média geral das medidas de correlação de Pearson, teste-F e teste-T entre as medidas MP2PD, Distância de Procrustes e MSE

	Procrustes MSE	MP2PD MSE	MP2PD Procrustes
Pearson	0.977349411	0.917150377	0.965403022
Teste F	0.001200059	1.23335E-10	3.51E-05
Teste T	1.14254E-05	0.036067373	8.91695E-07

No Quadro 1, apresenta-se, de forma numérica, a grandeza de cada medida de similaridade, já observada no exemplo gráfico da Figura 14. No Quadro 2, relacionam-se pares das medidas de similaridade adotadas. O objetivo do Quadro 2 é evidenciar o relacionamento entre cada medida de similaridade, de acordo com os testes de correlação de Pearson, Teste-T de variância e Teste-F de médias. O teste de correlação de Pearson permitiu verificar que há uma grande correlação na taxa de crescimento de todas as medidas, enquanto os testes F e T permitiram constatar que as variâncias e médias de todas as medidas são muito distintas, indicando que se tratam de medidas distintas, apesar de fortemente correlacionadas.

Na Figura 15, apresenta-se o mesmo exemplo da Figura 14, após a normalização dos dados que a geraram. O processo de normalização mapeou todos os valores dos gráficos para o intervalo [0, 1]. O processo de normalização altera os valores finais exibidos, mas preserva a taxa de crescimento geral dos gráficos. Uma vez normalizados, os dados passam a ser descritos no mesmo espaço e podem ser comparados diretamente.

Figura 15 – Gráfico de evolução normalizada de uma expressão facial



O Quadro 3 contém as médias dos valores observados em todas as sequências de imagens, após o processo de normalização.

Quadro 3 – Média geral das médias, desvios padrão e variâncias para as medidas MP2PD, Distância de Procrustes e MSE após normalização

	MP2PD	Procrustes	MSE
Média	0.582028265	0.566811362	0.416203375
Desvio Padrão	0.333606747	0.337383773	0.315289546
Variância	0.111293462	0.11382781	0.099407498

A partir do Quadro 4, verifica-se que o processo de normalização não altera o resultado obtido do teste de correlação de Pearson, mas há uma alteração significativa nos valores dos testes F e T. Os resultados do Teste-F mostram que, após a

normalização, a taxa de variação de todas as medidas passa a ser muito similar. Apesar da maior correlação entre as medidas de Procrustes e as medidas do Erro Médio Quadrático, o Teste-T para médias permitiu constatar que houve maior similaridade entre os valores da medida de Distância Média entre pontos e os valores da medida de Procrustes do que entre quaisquer outros pares de medidas considerados.

Quadro 4 – Média geral das medidas de correlação de Pearson, teste-F e teste-T entre as medidas MP2PD, Distância de Procrustes e MSE após normalização

	Procrustes MSE	MP2PD MSE	MP2PD Procrustes
Pearson	0.977349411	0.917150377	0.965403022
Teste F	0.963847605	0.905933885	0.941925787
Teste T	0.184186546	0.141171005	0.892304025

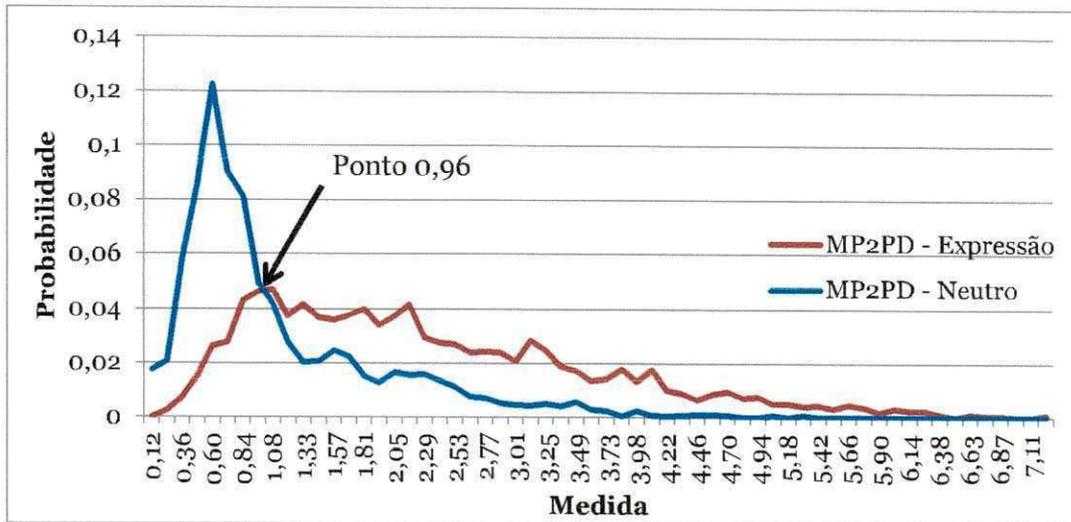
Em resumo, as medidas de similaridade adotadas neste estudo resultaram em medidas semelhantes da correlação de Pearson e do Teste-F de variância. As medidas estão no espaço de 5% umas das outras, o que significa que não há diferença estatisticamente significativa entre as medidas adotadas. Neste ponto, ainda é impossível formular uma conclusão definitiva sobre a diferença das medidas.

Na segunda etapa deste experimento, observam-se resultados mais práticos do uso das medidas de similaridade. Nesta etapa, foi realizada uma análise dos valores obtidos de cada medida ao comparar modelos neutros e modelos de expressão final. Esta análise permitiu a identificação de um valor de corte empregado para definir a partir de que valores observados se pode considerar que o modelo observado representa uma expressão facial ou se ainda representa um modelo neutro.

Os valores foram dispostos em gráficos de distribuição de probabilidade, os quais auxiliam a definir o valor de corte supramencionado. Aliados aos gráficos de distribuição de probabilidade foram empregados gráficos de probabilidade acumulada. Os gráficos de probabilidade acumulada auxiliam a definir a razão dos valores que estão abaixo de um determinado limiar.

Na Figura 16, ilustram-se os gráficos de distribuição de probabilidade de modelos neutros e de expressão sob a medida de distância média ponto a ponto. O eixo das abscissas contém os valores observados da medida de distância média entre pontos, enquanto no eixo das ordenadas são representadas as probabilidades associadas aos valores observados.

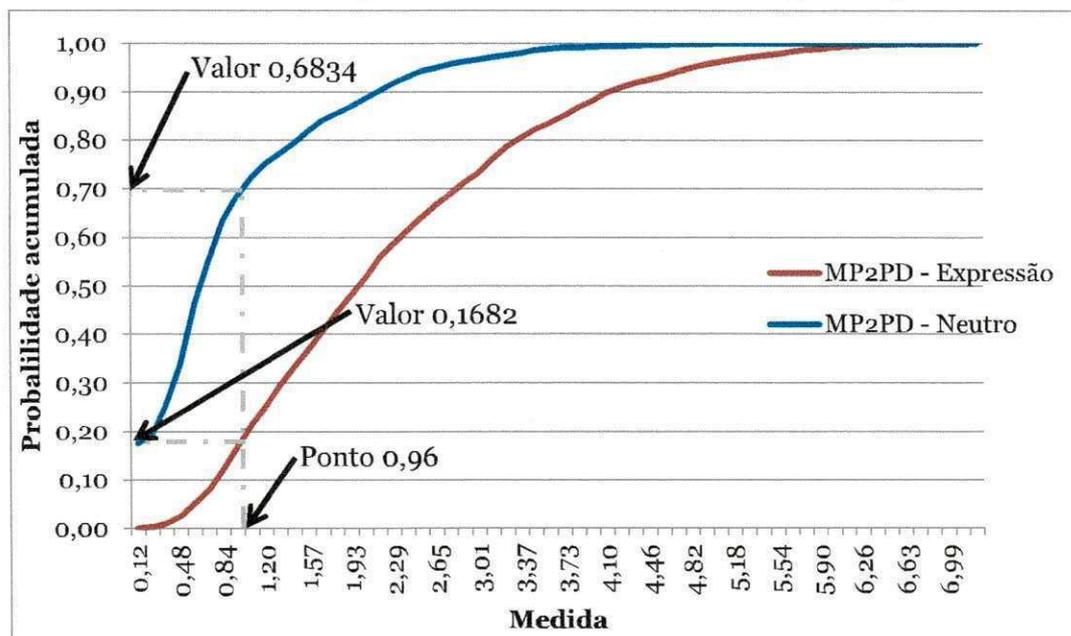
Figura 16 – Gráfico de distribuição de probabilidade do valor de distância entre modelo neutro pela medida de distância média ponto a ponto



Na Figura 16, evidenciou-se o ponto 0,96, no qual as curvas se cruzam. A partir deste ponto, a curva que descreve a distribuição de probabilidade de modelos de expressão ultrapassa a curva de modelos neutros. Qualquer valor medido a partir deste ponto terá maior chance de ser um modelo de expressão do que um modelo neutro.

O gráfico de distribuição de probabilidade acumulada (Figura 17) serve para verificar a quantidade total de observações de modelos que estão abaixo de qualquer valor. Na Figura 17, foram evidenciados os valores dos gráficos de distribuição de probabilidade acumulada correspondentes ao ponto 0,96 destacado na Figura 16.

Figura 17 – Gráfico de distribuição de probabilidade acumulada do valor de diferença entre modelo neutro pela medida de distância média ponto a ponto

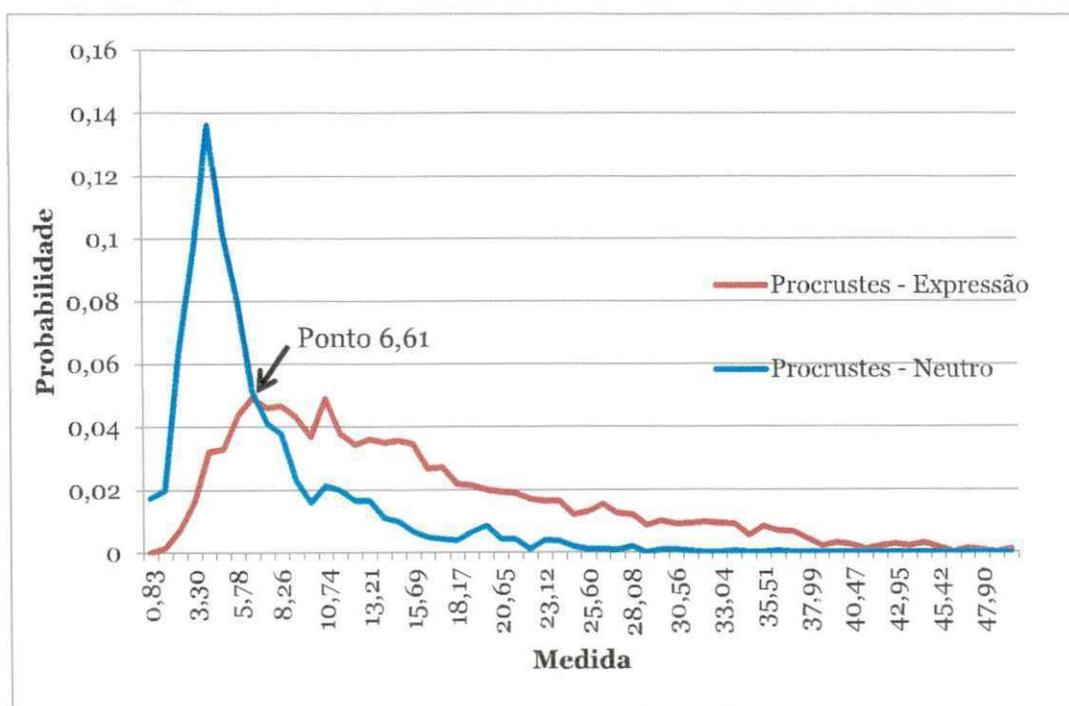


A curva de distribuição acumulada dos modelos neutros mostra que, na comparação de modelos neutros com outros modelos neutros, 68,34% das comparações resultam medidas de distância média ponto a ponto menores que 0,96.

Na curva de distribuição acumulada dos modelos de expressão, observa-se que 16,82% dos modelos de expressão possuem medidas menores que 0,96 na comparação de qualquer modelo de expressão com qualquer modelo neutro. A razão entre as porcentagens observadas é igual a **4,063**. Quanto maior for esta razão, menor será a taxa de confusão associada a esta medida de similaridade.

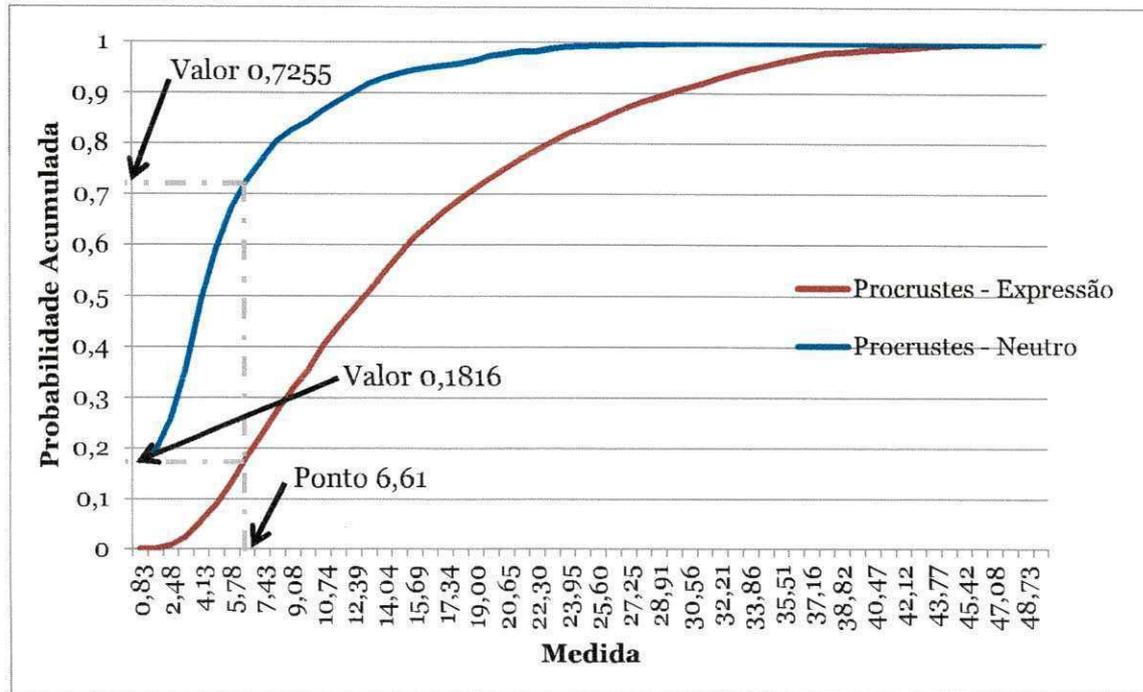
Os gráficos representados nas Figuras 18 e 19 descrevem a distribuição de probabilidade e a distribuição de probabilidade acumulada para a medida de Procrustes. Na Figura 18, está evidenciado o ponto 6,61, que corresponde ao limiar de corte associado à medida de Procrustes.

Figura 18 – Gráfico de distribuição de probabilidade do valor de diferença entre modelo neutro pela distância de Procrustes



No gráfico de distribuição de probabilidade acumulada da medida de Procrustes (Figura 19), o ponto 6,61 permite verificar que 72,55% dos modelos neutros possuem medidas de distância de Procrustes menores que 6,61, quando comparados a outros modelos neutros.

Figura 19 – Gráfico de distribuição de probabilidade acumulada do valor de diferença entre modelo neutro pela distância de Procrustes



Neste ponto, a medida da distância de Procrustes exibe um valor superior àquele exibido pela a medida de distância média entre pontos. Entretanto, também há uma proporção superior de modelos de expressão que possuem medidas menores que 6,61, verifica-se uma taxa de 18,16% dos modelos de expressão exibindo medidas menores que 6,61 em comparação com modelos neutros, usando a medida de distância de Procrustes. A razão resultante entre as porcentagens observadas é de **3,995**.

A Figura 20 contém o gráfico de distribuição de probabilidade para a medida de erro médio quadrático. Para a medida em questão o limiar de corte foi identificado no ponto 1,24. Na Figura 21, o gráfico de distribuição acumulada para os modelos neutros mostra que 82,65% dos modelos neutros possuem um valor de erro médio quadrático menor do que 1,24 em comparação com outros modelos neutros. A curva de distribuição de probabilidade acumulada para os modelos de expressão permite observar que 31,6% dos modelos de expressão apresentam um valor de erro médio quadrático menor que o limiar de corte desta medida. A razão resultante entre as porcentagens observadas nos gráficos de distribuição de probabilidade para a medida de erro médio quadrático é de **2,615**.

Figura 20 – Gráfico de distribuição de probabilidade do valor de diferença entre modelo neutro pela média dos erros ao quadrado

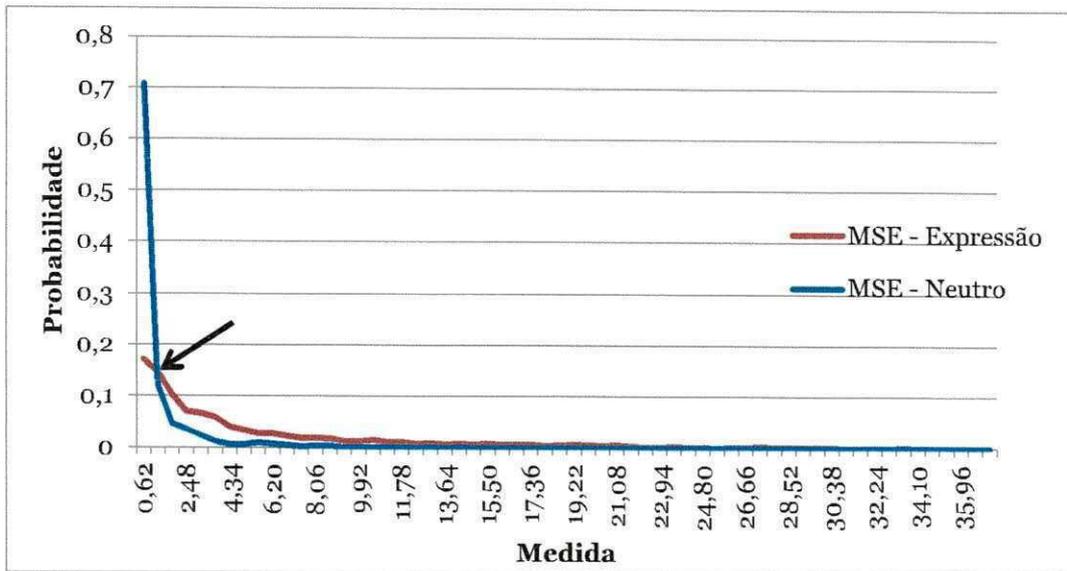
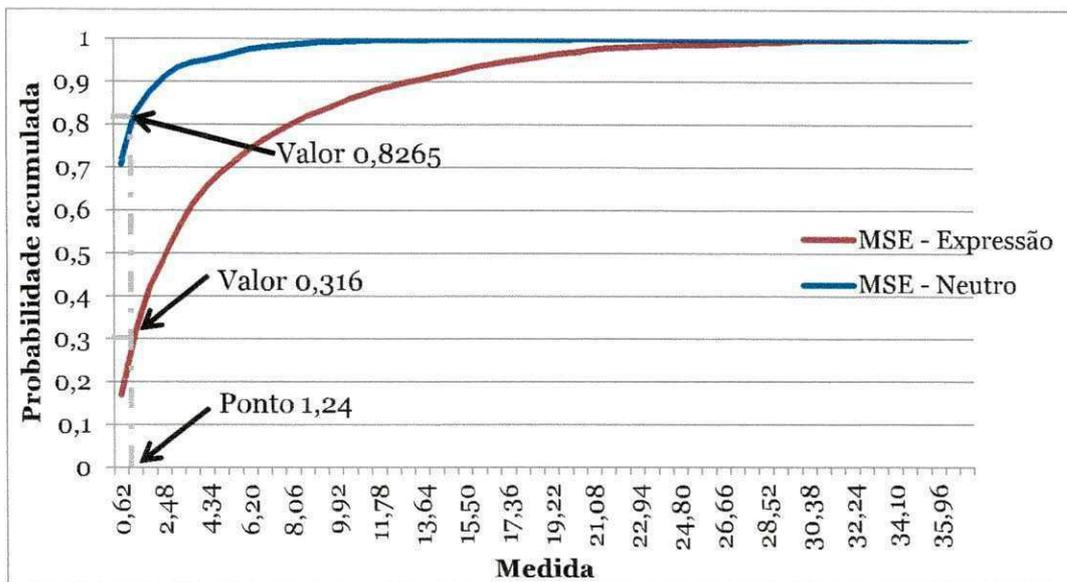


Figura 21 – Gráfico de distribuição de probabilidade acumulada do valor de diferença entre modelo neutro pela média dos erros ao quadrado



A diferença entre as razões observadas das medidas de distância média entre pontos e a distância de Procrustes não é suficiente para que seja possível discernir a qualidade de qualquer das medidas com segurança estatística.

Por outro lado, a medida de erro médio quadrático se comportou de forma diferente. A medida de erro médio quadrático exibiu uma razão entre os valores observados de **2,615** que, estando a mais de 5% de desvio de ambas as medidas de distância de Procrustes e distância média entre pontos, pode ser considerada de qualidade inferior às medidas mencionadas.

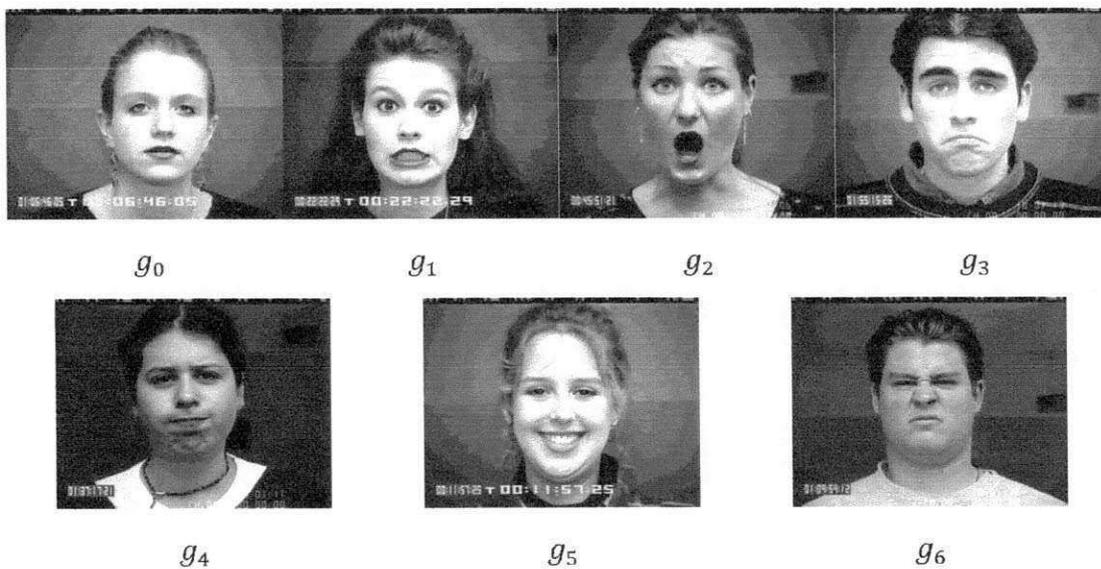
Mesmo não sendo estatisticamente superior, em comparação com a medida de distância de Procrustes, a medida de distância média entre pontos mostrou um resultado marginalmente superior e foi, portanto, escolhida como medida padrão da ferramenta, tendo sido a medida utilizada em todos os experimentos subsequentes.

4.2 Experimento 2 – Discriminação de Expressões Faciais

O experimento descrito nesta seção expande a análise feita no Experimento 1, de forma a verificar a qualidade da abordagem desenvolvida mais especificamente no tocante às expressões faciais individuais. Na seção 4.1, o experimento descrito permitiu verificar a qualidade do classificador de expressões de forma geral com foco nas possíveis medidas de similaridade, buscando diferenciar apenas expressões neutras de quaisquer outras expressões. Neste experimento, verificou-se a qualidade do classificador quanto à capacidade de identificação de expressões faciais específicas.

Para este propósito, o banco de imagens Cohn-Kanade foi dividido em 6 grupos de expressões semelhantes (do ponto de vista de aparência visual), além das expressões neutras presentes em todas as sequências de imagens (Figura 22).

Figura 22 – Imagens representativas dos agrupamentos de imagens do banco de imagens Cohn-Kanade. Abaixo de cada figura os rótulos $g_{0,1,\dots,6}$ indicam a nomenclatura adotada para as sequências de imagens visualmente semelhantes à expressão que está representada imediatamente acima de cada rótulo ©Jeffrey Cohn



Os grupos foram rotulados de g_0, g_1, \dots, g_6 , representados na Figura 22. O grupo rotulado com g_0 representa as expressões neutras exibidas no quadro inicial de todas as 490 sequências de imagens do banco de imagens. Os demais rótulos foram definidos sem nenhuma ordem em particular, contudo cada rótulo indica sequências de imagens visualmente semelhantes à expressão que está representada nas imagens acima dos rótulos. Em cada grupo restante representado na Figura 22 há entre 48 e 122 sequências de imagens, sendo que o grupo g_1 representa 60 sequências de imagens (1079 imagens), o grupo g_2 representa 122 sequências de imagens (2329 imagens), o grupo g_3 representa 83 sequências de imagens (1456 imagens), o grupo g_4 representa 58 sequências de imagens (1000 imagens), o grupo g_5 representa 119 sequências de imagens (2143 imagens) e o grupo g_6 representa 48 sequências de imagens (782 imagens).

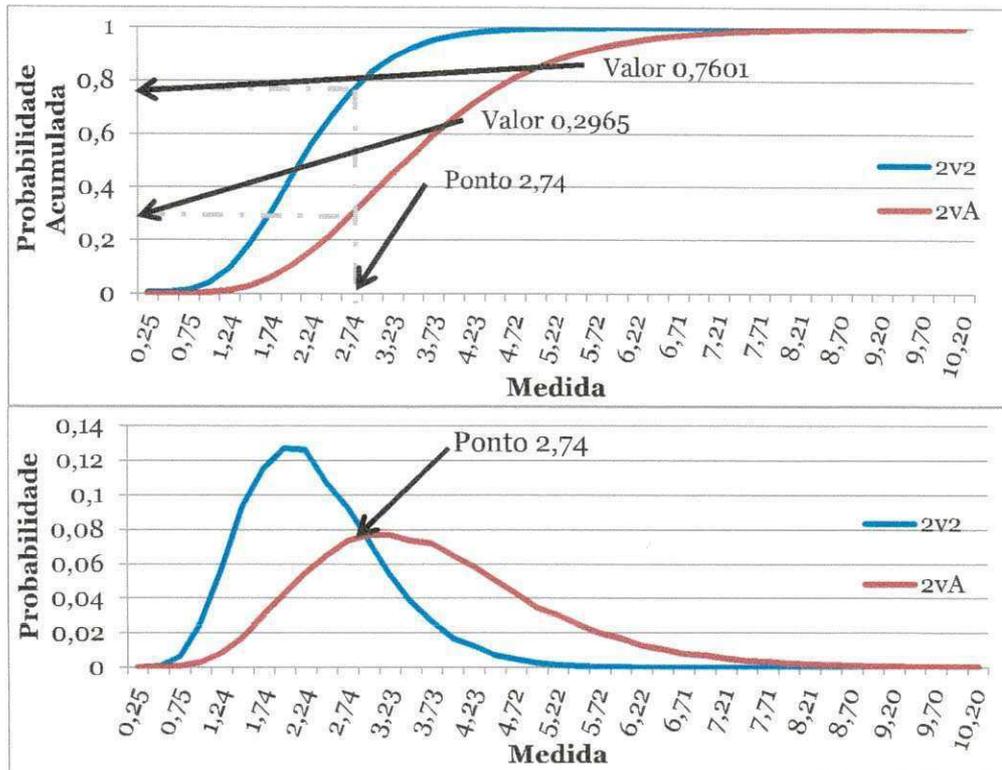
De maneira semelhante à estratégia adotada no Experimento 1, no Experimento 2 não se faz nenhuma distinção de alto nível a respeito de cada grupo de expressões (não há nenhuma classificação da reação facial apresentada nas imagens). Este experimento foi conduzido de maneira semelhante à forma como foi conduzida a segunda etapa do Experimento 1. Cada grupo de imagens de expressões faciais é comparado com as outras imagens dentro do grupo e com as imagens fora do grupo.

4.2.1 Resultados do Experimento 2

Embora o processo executado na segunda parte do Experimento 1 tenha sido repetido no Experimento 2, este diferiu do Experimento 1 por realizar isoladamente a análise de cada grupo de expressões faciais.

Na Figura 23, exemplifica-se a análise do grupo g_2 de expressões faciais. Os gráficos expostos na Figura 23 são os mesmos tipos de gráficos, de distribuição de probabilidade e de probabilidade acumulada, usados para a análise das medidas de similaridade na seção anterior.

Figura 23 – Análise do grupo g_2 de expressões faciais



A análise realizada nos demais grupos de expressões faciais produz gráficos e análises similares àqueles exemplificados para o grupo g_2 . Os resultados das análises de todos os grupos estão condensados no Quadro 5.

Quadro 5 – Taxa de acerto de cada grupo de expressão facial

	g_0	g_1	g_2	g_3	g_4	g_5	g_6	Média
Dentro do grupo	68,34%	72,43%	76,01%	63,89%	63,8%	71,57%	65,37%	68,77%
Fora do grupo	16,82%	34,12%	29,65%	46,64%	40,27%	33,92%	40,75%	34,6%
Razão de acerto	4,063	2,122	2,563	1,369	1,584	2,110	1,604	2,202

A baixa taxa de acerto exibida no Quadro 5 tem relação com a falta de uma técnica mais robusta de classificação de expressões. Na abordagem empregada nesta dissertação, a única forma de diferenciação entre expressões é o valor observado da comparação entre os modelos de forma de acordo com a medida de comparação adotada. Esta técnica confia que haverá maior variação na comparação entre modelos de expressões de grupos diferentes do que a comparação de modelos de forma do mesmo grupo, que nem sempre será verdade.

Com relação aos resultados obtidos, o grupo g_0 (que corresponde às expressões neutras) repete o resultado que foi obtido no Experimento 1 com a medida de distância média ponto a ponto, por se tratar do mesmo espaço amostral utilizado para a geração dos gráficos. Coincidentemente, o resultado obtido na comparação do grupo g_0 é o melhor resultado do experimento. O pior resultado foi apresentado pelo grupo g_3 , com uma razão de apenas 1,369. Há, no entanto, uma característica comum aos três grupos com as piores razões, foi observado que os grupos g_3 , g_4 e g_6 , apresentam pouca variação na boca, enquanto os grupos g_1 , g_2 e g_5 , que correspondem aos melhores resultados, apresentam uma variação relativamente superior no movimento da boca.

É importante observar a grande diferença que existe na razão de acerto entre os diferentes grupos de modelos e o resultado obtido para os modelos neutros. O principal motivo da diferença está na quantidade de modelos de fora de cada grupo cujo erro se assemelha ao erro dos modelos de dentro do grupo. Em comparação com o grupo de expressões neutras g_0 , o grupo g_2 de expressões foi aquele que apresentou o melhor resultado, porém apresentou apenas 63% da taxa de acerto do grupo g_0 .

O Quadro 6 contém a proporção de modelos cujo erro é semelhante ao erro observado em cada grupo. No exemplo do grupo g_2 , cada coluna do Quadro 6 apresenta a proporção de modelos de outros grupos cujo erro medido, em comparação com modelos do grupo g_2 , foi menor do que 2,74 (o valor de corte).

Quadro 6 – Taxa de apresentação de cada grupo na área de erro

	g_0	g_1	g_2	g_3	g_4	g_5	g_6
g_0		5,65%	1,21%	32,40%	17,12%	0,71%	18,67%
g_1	5,40%		35,08%	10,00%	9,46%	35,51%	6,60%
g_2	1,02%	37,23%		2,24%	13,28%	36,54%	6,43%
g_3	37,16%	5,20%	5,88%		26,85%	15,90%	31,03%
g_4	30,68%	5,68%	6,69%	30,09%		5,36%	35,48%
g_5	8,68%	30,61%	36,41%	4,98%	2,47%		1,79%
g_6	17,06%	15,64%	14,74%	20,29%	30,82%	5,98%	

Não se deve confundir o Quadro 6 com uma matriz de confusão. Nesta pesquisa, não objetivou-se a geração de uma técnica de classificação de modelos de forma. Nas estatísticas exibidas, um modelo qualquer do grupo g_2 , por exemplo, cujo erro em comparação com outro modelo do grupo foi inferior a 2,74 não está isento de aparecer como um valor superior em outra comparação.

Conforme anteriormente comentado, a boca é uma estrutura interna da face que parece ter um impacto relativamente maior na taxa de acerto ou erro da técnica desenvolvida. De fato, os grupos g_1 , g_2 e g_5 exibiram maior taxa de aparição dentre os modelos errados dos mesmos grupos, bem como os grupos g_0 , g_3 , g_4 e g_6 exibiram maior taxa de confusão entre si, reforçando a noção de que o movimento da boca é um fator impactante no erro observado.

4.3 Experimento 3 – Relação entre Sumário e Dados Quantitativos

Se evidencia, nas seções 4.1 e 4.2, que a abordagem desenvolvida é capaz de discernir expressões faciais diversas, no experimento descrito nesta seção buscou-se verificar a qualidade da sumarização produzida pela abordagem desenvolvida. Com este objetivo, uma bateria de sessões de teste de usabilidade foi realizada.

4.3.1 Ambiente

Não foi possível, no âmbito desta pesquisa, realizar sessões de teste laboratoriais. Foram realizadas sessões de teste de campo, em diferentes ambientes de trabalho do Departamento de Sistemas e Computação (DSC) e, do Centro de Engenharia Elétrica e Informática (CEEI) da Universidade Federal de Campina Grande (UFCG).

A maior parte dos testes foi realizada no Laboratório de Visão Computacional (LVC). Entretanto, houve exceções para usuários que preferiram realizar os testes em seus próprios ambientes de trabalho, o que caracteriza os testes realizados como de campo.

4.3.2 Produtos Selecionados para Condução das Sessões de Teste

Os produtos selecionados para a realização das sessões de teste de usabilidade foram: (i) o MacBook Pro, da Apple Inc.; e (ii) o IMPRESS, programa para a composição de apresentações multimídia do pacote de aplicativos de escritório LibreOffice.org. O sistema operacional escolhido foi o Linux Ubuntu, versão 12.4, executado em uma máquina virtual VirtualBoxVM, da Oracle.

A versão do MacBook Pro utilizada no teste é internacional, o que significa que o padrão utilizado no teclado é diferente do padrão ABNT2 usado no Brasil. Esta diferença em uma das interfaces de entrada passa despercebida até o momento em que uma letra precisa ser acentuada ou a letra ‘ç’ se faz necessária no texto.

A suíte de aplicativos para escritório LibreOffice.org é uma solução de código aberto, licenciado com a *GNU Lesser General Public License* (GNU LGPLv3), que oferece as funções básicas esperadas de um pacote de aplicativos de escritório, processador de texto, editor de planilhas numéricas, editor de apresentações multimídia, gerenciador de banco de dados e editor de imagens.

Os aplicativos contidos no pacote do LibreOffice.org estão disponíveis para os principais ambientes operacionais, Windows, OSX, Linux e BSD, mas foi notado que o ambiente Linux é mais estável para a execução dos aplicativos do LibreOffice.org. Por essa razão foi adotado o ambiente Linux com uma distribuição popular, o Ubuntu, na versão 12.04.

O *VirtualBoxVM* é uma solução de virtualização de sistemas operacionais de código aberto, licenciado sob os termos da GNU General Public License (GNU GPLv2). O *VirtualBoxVM* está disponível para ambientes operacionais Windows, OSX, Linux e Solaris. O aplicativo não tem limites teóricos quanto à quantidade de máquinas virtuais que é capaz de iniciar, os únicos limites práticos são o espaço de disco e de memória principal.

Tendo em vista que as sessões de teste deste experimento não foram laboratoriais, o registro em vídeo de cada sessão ocorreu no próprio sistema de teste. Para reduzir o impacto dos aplicativos de captura da tela e o registro em vídeo da micro câmera, durante cada sessão de teste, o ambiente Linux, utilizado no teste pelos usuários, foi logicamente separado do ambiente OSX que hospeda o ambiente de teste.

Utilizar uma técnica de virtualização traz benefícios à realização das sessões de teste. Um benefício supracitado é a separação lógica do ambiente de teste. Outro benefício em é a possibilidade de uso do recurso de “fotografia instantânea (*snapshot*)”. Este recurso possibilita capturar o estado de uma máquina virtual e reiniciar a sua execução a partir do estado gravado. O estado da máquina virtual é caracterizado por tudo o que está sendo executado (detalhes do estado de execução do processador), tudo o que está na memória principal (programas em execução, itens na área de transferência) e pelo estado do disco da máquina virtual (o que está instalado, em quais arquivos e onde se encontram).

No contexto deste experimento, esta funcionalidade garantiu que cada usuário de teste esteve diante da mesma situação inicial, independentemente de qualquer ação possivelmente destrutiva que os usuários de teste anteriores possam ter executado. Antes de cada nova sessão de teste, a máquina virtual é carregada a partir da mesma *snapshot*, que foi mantida constante para todas as sessões realizadas. Após o iniciada a máquina virtual a partir da *snapshot*, cada usuário se deparou com a máquina virtual exibindo exatamente o mesmo estado de execução.

4.3.3 Elaboração dos Documentos Relacionados ao Teste

As tarefas de teste para o aplicativo IMPRESS foram planejadas de forma a contemplar algumas das funções básicas do aplicativo. As tarefas contemplaram desde a ação de criação de slides de uma apresentação até a criação de gráficos.

Após a realização dos testes-piloto, constataram-se dois problemas que foram a motivação para a reestruturação das tarefas de teste. O primeiro problema foi a extrapolação do tempo das sessões, com relação ao limite pré-estabelecido de 60 minutos (NIELSEN, 1993; BARBOSA, 2009).

O segundo problema foi a constante mudança de foco, por parte do usuário, do aplicativo sob avaliação para o roteiro de teste impresso. Este problema, em particular, foi de grande impacto para a abordagem desenvolvida. A variação de pose da cabeça de cada usuário, no momento em que estava com a atenção voltada para o roteiro de testes, resultou em grande variação no modelo de forma e, por consequência, grande variação no valor medido. Tal variação foi muito grande, em comparação com a variação exibida por expressões faciais, chegando a superar a variação exibida por expressões faciais. Tal variação não foi considerada no escopo desta pesquisa.

Foram realizadas 5 sessões de teste com a estrutura inicialmente descrita. Os vídeos obtidos com estas sessões de teste não puderam ser utilizados para a análise da abordagem descrita neste trabalho. Os voluntários de teste não puderam mais ser chamados para uma segunda sessão por já terem conhecimento do propósito do teste e desta pesquisa. Tal conhecimento poderia influenciar o comportamento natural dos usuários, o que também diminuiria a confiança nos resultados obtidos do vídeo das novas sessões.

Diante do exposto, as tarefas de teste foram reestruturadas, de forma a compor 8 tarefas. O roteiro de testes foi modificado, tendo sido criada uma versão sob a forma de slides no próprio aplicativo sob avaliação (Apêndice J).

Em seguida, foi redigido o formulário para o cadastro de participação (Apêndice C), cujo propósito principal foi criar a base de participantes potenciais para experimentos futuros.

Foram elaborados documentos descrevendo as condições de teste às quais os participantes seriam submetidos. O *Documento de Aceitação das Condições de Teste* (Apêndice D) trata das questões relacionadas ao registro em vídeo da sessão de teste que o candidato irá participar. O *Termo de Confidencialidade* (Apêndice E) é relacionado à divulgação de informações confidenciais relacionadas ao teste.

Em seguida, foi elaborado o material necessário para a análise do perfil do usuário, a qual fundamentou-se em um questionário inspirado na pesquisa de Oliveira

(2005). O questionário para o delineamento do perfil do usuário foi distribuído juntamente com o formulário para o cadastro de participação. O questionário produzido pode ser visualizado no Apêndice F.

4.3.4 Análise do Perfil do Usuário

No Apêndice G, são resumidas as respostas dos usuários de teste ao questionário para o delineamento do perfil do usuário de teste (Apêndice F). A síntese descrita no Apêndice G reproduz a sequência de questões e a ordem das respostas e exibe a quantidade de usuários que selecionaram cada opção do questionário.

O questionário para o delineamento do perfil do usuário de teste deste experimento compôs-se de 15 itens, com o propósito de obter informações acerca de características físicas (itens 2, 3 e 4), nível de conhecimento geral (itens 1 e 15) e experiência do usuário (itens 5 a 14). As questões de 5 a 14 sondam desde a experiência geral do usuário com sistemas computacionais e com aplicativos semelhantes ao aplicativo testado até sua com o próprio aplicativo testado, a partir de itens incrementalmente mais restritivos. O questionário facilitou a caracterização do espaço amostral de usuários de teste.

A partir dos dados obtidos mediante a administração do questionário para o delineamento do perfil do usuário (Apêndice F), foi possível descrever o usuário típico do experimento pelas seguintes características:

- Possui como grau de instrução ensino superior incompleto (18 voluntários, 60% do espaço amostral);
- É predominantemente do sexo masculino (20 voluntários, dois terços do espaço amostral);
- Usa corretivos visuais (18 voluntários, 60% do espaço amostral);
- Pertence a faixa etária de 18 a 23 anos (20 voluntários, dois terços do espaço amostral);
- É experiente no uso de sistemas computacionais, utilizado computadores há mais de um ano e atualmente utiliza computadores diariamente (30 voluntários, 100% do espaço amostral);
- É usuário do ambiente operacional Windows (21 voluntários, 70% do espaço amostral);
- Possui nível elevado de conhecimento de informática (25 voluntários, 83% do espaço amostral no nível de programação e avançado)
- Possui experiência prévia com aplicativos de composição de apresentações multimídia (30 voluntários, 100% do espaço amostral);

- Possui experiência prévia com aplicativos de suíte de aplicativos para escritório do *LibreOffice.org* (25 voluntários, 83% do espaço amostral);
- Não possui experiência prévia com o *IMPRESS* (23 voluntários, 77% do espaço amostral);
- Possui nível intermediário de habilidade com a língua inglesa (entre 46% e 57% do espaço amostral);

A análise do perfil do usuário foi capaz de caracterizar um usuário padrão para o teste. Tal uniformidade é fundamental para que os resultados obtidos da realização do teste possam ser analisados corretamente. Um perfil comum aumenta as chances de que os usuários entrem em acordo com relação aos problemas de usabilidade encontrados, assim como também significa que haverá menor diversidade de problemas de usabilidade encontrados. Entretanto, não é do interesse deste experimento identificar o máximo de falhas de usabilidade do produto sob avaliação, mas verificar a corretude dos resultados obtidos, que só é possível após constatada a uniformidade no espaço amostral, relevante tanto para a contextualização dos dados coletados quanto para as conclusões que ocorrerão dos dados coletados.

4.3.5 Condução das Sessões de Teste

A condução de cada sessão de teste foi realizada o mais similar e consistentemente possível. A primeira atividade realizada na sessão de teste foi a introdução do usuário no ambiente de teste. Nesta atividade, foram esclarecidos tópicos sobre a usabilidade de produtos e sobre os procedimentos a serem seguidos durante a condução do teste. O usuário foi informado de que o teste seria registrado em vídeo, tendo o avaliador destacado seu comprometimento com o sigilo e anonimato do usuário, quanto aos dados pessoais e resultados de seu teste.

Durante a sessão de teste, não foi informado ao usuário o propósito da pesquisa, uma vez que tal informação poderia comprometer seu comportamento natural, favorecendo a explicitação ou supressão de expressões faciais artificiais. Apenas após a conclusão de cada sessão de teste é que o usuário foi informado sobre o propósito da pesquisa. De acordo com o termo de confidencialidade (Apêndice E), os usuários de teste concordaram em não divulgar informações do teste, este termo do teste foi reforçado verbalmente.

Em seguida, ocorreu a execução das tarefas pelos usuários de teste, de acordo com o roteiro das tarefas apresentado no Apêndice J. Durante a condução dos testes, a face do usuário e a tela do computador foram registrados em vídeo. Os vídeos dos usuários foram registrados a partir de uma micro câmera, embutida na tela do

computador e o registro da tela deu-se por meio de uma aplicação de *software* para a captura de vídeo.

Após a finalização dos testes, foi realizada a extração manual dos dados quantitativos selecionados para a mensuração do desempenho neste teste. Esta etapa foi realizada com uma observação e anotação dos vídeos de captura de tela. Os resultados detalhados da coleta dos dados de mensuração de desempenho estão apresentados no Apêndice I.

4.3.6 Resultados do Experimento 3

No Quadro 7, apresenta-se uma síntese dos dados coletados a partir da mensuração do desempenho dos usuários de teste (Apêndice I). No quadro em questão, em cada coluna estão somados os eventos observados em cada tarefa executada pelos usuários, a linha **OI** corresponde à quantidade de opções incorretas feitas por todos os usuários em cada tarefa, a linha **AI** representa a quantidade de ações incorretas executadas por todos os usuários em cada tarefa, a linha **ER** diz respeito à quantidade de erros repetidos, enquanto **Ajuda** indica o número de total de vezes em que os usuários acessaram a guia de ajuda do produto sob avaliação.

Quadro 7 – Síntese da análise de indicadores quantitativos

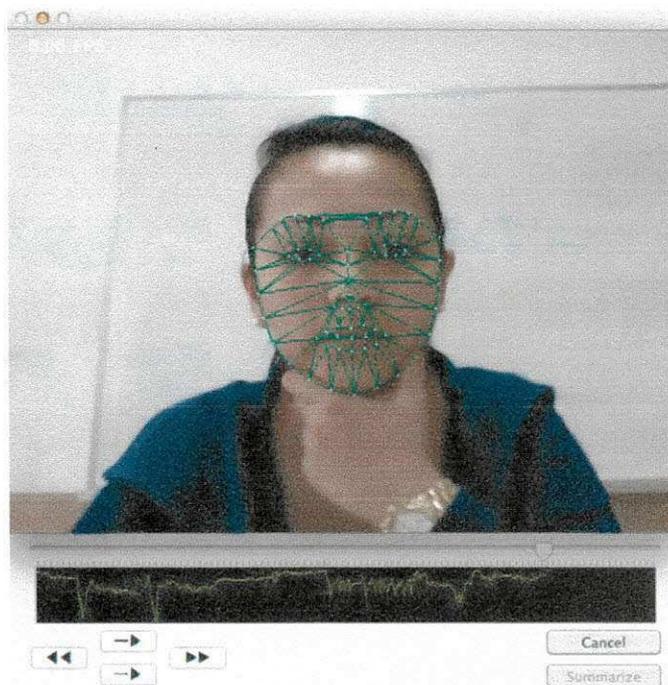
	To	T1	T2	T3	T4	T5	T6	T7	T8
OI	0	45	20	15	22	104	14	3	0
AI	0	10	0	0	0	19	0	0	0
ER	0	3	0	0	0	7	0	0	0
Ajuda	1	3	0	0	0	10	0	0	0
Soma	1	61	20	15	22	140	14	3	0

Os dados exibidos no Quadro 7 oferecem uma base de comparação para os dados obtidos dos vídeos das sessões de teste conduzidas para este experimento. Os dados exibidos no Quadro 7 foram comparados com dados obtidos de forma automática pela ferramenta desenvolvida. Para obter os dados do Quadro 7, foram usados os vídeos obtidos a partir da captura de tela do computador, os quais contêm as imagens do que o usuário visualizou na tela do computador de teste durante a execução das tarefas descritas no Apêndice J.

A abordagem proposta é responsável pela análise dos vídeos obtidos com a micro câmera do computador. Cada vídeo foi analisado pela ferramenta desenvolvida e a cada quadro foi associado um modelo de forma ajustado à face e expressão facial exibida pelo usuário de teste.

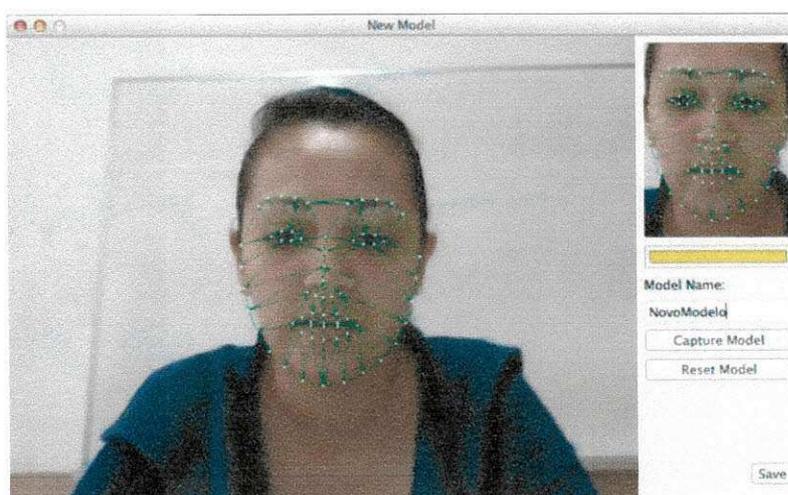
Na Figura 24, apresenta-se a interface da ferramenta desenvolvida baseada na abordagem proposta. A figura exemplifica a etapa de análise de um dos vídeos compostos durante a realização deste teste. O gráfico exibido logo abaixo da janela de vídeo na interface é utilizado para a composição do sumário do vídeo.

Figura 24 – Interface do sistema desenvolvido com base na abordagem proposta



A abordagem necessita de um modelo de comparação para obter valores referentes à distância média entre pontos do modelo de comparação e do modelo ajustado em cada quadro. Foi escolhido um modelo neutro retirado de um quadro de cada vídeo. Na Figura 25, apresenta-se a interface de criação de modelos de comparação necessários para o funcionamento do sistema.

Figura 25 – Interface de criação de modelo de comparação



O processo definido pela abordagem proposta requer que cada quadro do vídeo sendo avaliado seja comparado com o modelo de comparação. De acordo com os resultados do Experimento 1, foi utilizada a medida de distância média entre pontos para comparar os modelos de forma. A comparação do modelo de comparação com os modelos obtidos do decorrer do vídeo resulta em uma lista ordenada de valores reais. A lista ordenada resultante possui um único valor para cada quadro do vídeo e cada vídeo tem sua própria lista. Cada lista é dividida em seções que correspondem ao mesmo espaço de tempo de cada tarefa, ou seja, cada seção que é separada da lista tem início no mesmo quadro no qual uma nova tarefa é iniciada pelos usuários e término no último quadro de cada tarefa. De cada seção, é obtida a média dos valores da distância média entre pontos presentes em cada seção.

Neste estágio do processo, 30 listas foram obtidas, correspondentes aos 30 usuários de teste, cada uma das quais contendo nove valores. Os valores exibidos no Quadro 8 são uma média dos valores dos 30 vídeos para cada seção. Assim, é possível observar a variação média de todos os usuários para cada tarefa. O desvio padrão observado no Quadro 8 é um subproduto da natureza da composição dos dados, baseados em média de médias (Lindgreen, 1968). Todas as médias obtidas tem a mesma natureza, uma vez que foram obtidas duas médias de conjuntos de dados similares as médias finais obtidas, exibidas no Quadro 8, terão a tendência de estarem quantitativamente próximas. Por este motivo, a existência de variação entre as médias finais ganha força estatística, mesmo considerando que os respectivos desvios padrão indicam baixa variação dos dados.

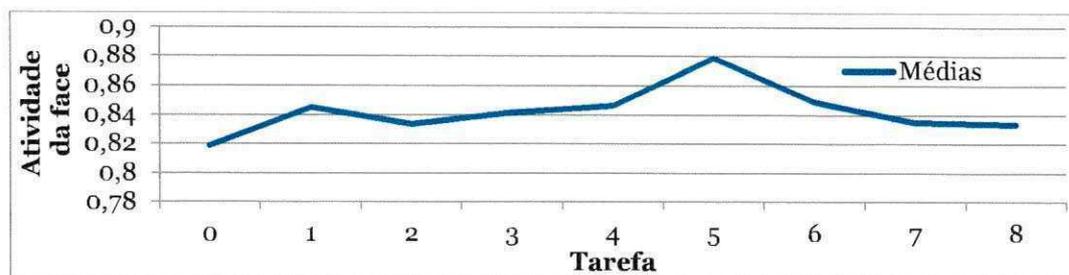
Na Figura 26, representa-se graficamente a variação média das expressões faciais de todos os usuários no decorrer do experimento, de acordo com os dados exibidos no Quadro 8.

Quadro 8 – Sumário dos valores de comparação da face do usuário com o modelo neutro

	T0	T1	T2	T3	T4	T5	T6	T7	T8
Média	0,8187	0,8450	0,8338	0,8413	0,8463	0,8789	0,848	0,8350	0,8338
Desvio	0,044	0,0454	0,0398	0,0414	0,0422	0,0481	0,041	0,0405	0,0391

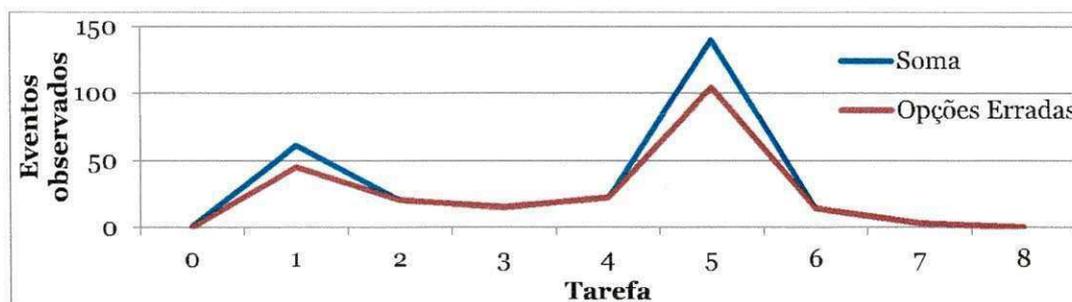
O desvio padrão observado nos dados exibidos no Quadro 8, pode ser considerado alto, quando observado fora de contexto, estes valores se devem ao fato de que há uma grande diferença entre um valor obtido da comparação do modelo neutro utilizado e um modelo de expressão que tenha ocorrido durante a condução da sessão de teste.

Figura 26 – Gráfico representativo dos valores do Quadro 8



A partir do gráfico exibido na Figura 26, em comparação com os gráficos exibidos na Figura 27, é possível verificar a forte correlação existente entre a quantidade de eventos anotados manualmente e os valores médios obtidos com o auxílio da ferramenta desenvolvida. De fato, o valor do coeficiente de correlação de Pearson entre os valores relatados no Quadro 8 e a quantidade total de eventos, relatados na linha **Soma** do Quadro 7, é aproximadamente 0,891.

Figura 27 – Gráfico de soma de eventos e de opções erradas em cada tarefa



O valor da correlação é ainda levemente maior se a comparação for feita apenas observando a quantidade de opções erradas. Neste caso, o valor resultante da medida de correlação de Pearson é aproximadamente 0,898.

Este resultado indica que existe alta probabilidade de ocorrência de uma dependência linear entre as medidas relacionadas o que sugere que a medida das médias de valores de comparação da face do usuário com o modelo neutro pode ser usada para inferir conclusões acerca da taxa de aparição de eventos de usabilidade.

5 Considerações Finais

Neste capítulo, apresentam-se as discussões finais referentes às análises dos resultados obtidos a partir da aplicação da abordagem proposta. Inicialmente, resgata-se o contexto de desenvolvimento deste trabalho, apresentando as motivações e justificativas para seu desenvolvimento. Em seguida, são apresentadas as conclusões advindas dos resultados obtidos a partir da aplicação da abordagem desenvolvida. Por fim, apresentam-se algumas proposições para trabalhos futuros.

5.1 Sumário da Dissertação

A motivação para o desenvolvimento deste trabalho surgiu da necessidade de criação de uma forma mais simples de realizar a tarefa de analisar vídeos de sessões de teste de usabilidade. Conforme discutido nos Capítulos 1 e 2, a análise de vídeos de sessões de teste de usabilidade é uma tarefa onerosa e neste contexto, a presente dissertação propôs uma abordagem de sumarização para os referidos vídeos.

Na abordagem proposta nesta pesquisa, descrita no Capítulo 3, empregou-se como mecanismo básico a técnica de AAM. Os vídeos das sessões de teste de usabilidade empregados na avaliação experimental foram analisados automaticamente pela abordagem proposta e implementada neste trabalho. Como resultado, a cada quadro de cada vídeo foi associado um modelo AAM ajustado à face do usuário de teste do vídeo. Os modelos de cada uma destas sequências de AAM, associadas a cada vídeo de sessão de teste de usabilidade, foram comparados com um modelo específico para obter um gráfico de similaridade. Este gráfico de similaridade foi utilizado para a sumarização do vídeo.

O sistema computacional desenvolvido com base na abordagem proposta neste trabalho foi submetido a experimentos de validação, detalhados no Capítulo 4 desta dissertação, para a avaliação de sua viabilidade. Os experimentos levaram em consideração os questionamentos e hipóteses levantados no Capítulo 1. Considerações a respeito da realização de tais experimentos e as conclusões desta pesquisa são apresentadas na próxima seção.

5.2 Conclusões

Conforme descrito na seção 1.4 (**Objetos da Pesquisa**) deste documento, a presente pesquisa teve o propósito de verificar as seguintes hipóteses:

- H_1 : Há diferenças estatisticamente significativas entre as diferentes medidas de similaridade de expressões faciais estudadas;
- H_2 : Há diferenças estatisticamente significativas entre os diferentes grupos de expressões faciais estudados; e
- H_3 : Há uma relação estatisticamente significativa entre o nível de dificuldade de uma tarefa e a taxa de apresentação de expressões faciais não neutras nos participantes.

Os experimentos descritos no Capítulo 4 foram norteados pelas hipóteses formuladas na seção 1.4. A confrontação dos dados experimentais e considerações resultantes com as hipóteses descritas possibilitaram as seguintes conclusões:

Argumento 1: De acordo com a análise feita das Figuras 16 a 19, dentre as medidas de similaridade estudadas, a *distância média entre pontos* e a medida de *distância de Procrustes* não são estatisticamente diferentes. De acordo com a análise das Figuras 20 e 21, a medida de erro médio quadrático é estatisticamente diferente de ambas as medidas.

Conclusão 1: **Há diferenças estatisticamente significativas entre as diferentes medidas de similaridade adotadas na comparação de modelos de forma.** (Hipótese H_1 Aceita)

Argumento 2: No Quadro 5, registra-se a taxa de acerto de cada grupo de expressões faciais com relação aos demais. De acordo com o Quadro 5, a medida de similaridade utilizada é suficiente para exibir diferenças estatisticamente significativas entre grupos de expressões faciais de tal forma que a abordagem desenvolvida foi capaz de discernir entre cada grupo de expressões faciais com uma taxa mínima aceitável de acerto.

Conclusão 2: **Há diferenças estatisticamente significativas na comparação de diferentes modelos de forma.** (Hipótese H_2 Aceita)

Argumento 3: A correlação observada entre os valores apresentados nos Quadros 7 e 8 permite inferir que há forte correlação entre as medidas quantitativas observadas durante a condução do teste e as medidas de variação de expressões faciais obtidas com o auxílio da aplicação desenvolvida nesta pesquisa.

Conclusão 3: **Há uma relação estatisticamente significativa entre a quantidade de eventos negativos de usabilidade observados em uma tarefa e a taxa de apresentação de expressões faciais no decorrer da tarefa na qual tais eventos ocorreram.** (Hipótese H_3 Aceita)

Adicionalmente, a partir da aplicação da técnica de AAM no contexto de sumarização de vídeos de sessões de teste de usabilidade, dos dados experimentais coletados e da análise estatística realizada, conclui-se que: **o mecanismo de AAM, como caso específico de aplicação, é eficaz na detecção de variações de expressões faciais para ser utilizado como mecanismo fundamental de uma aplicação de sumarização de vídeos de sessões de teste de usabilidade.**

5.3 Contribuições

As principais contribuições desta pesquisa foram:

1. Um estudo da área de aplicação desta pesquisa e das técnicas relevantes para o desenvolvimento da abordagem, compilado no Capítulo 2 (**Revisão Bibliográfica**);
2. A proposição de uma abordagem para a sumarização automática de vídeos de sessões de teste de usabilidade, fundamentada no estudo realizado no Capítulo 2 (**Revisão Bibliográfica**) e compilada no Capítulo 3 (**Abordagem Proposta**);
3. Um aplicativo de *software* que implementa a abordagem proposta neste trabalho;
4. A realização de experimentos objetivos para a validação da abordagem proposta, utilizando um banco de imagens de expressões faciais que foram rotuladas para o propósito deste trabalho e um conjunto de vídeos de sessões de teste de usabilidade; e
5. A realização de testes estatísticos que respaldam a abordagem proposta para a sumarização automática de vídeos de sessões de teste de usabilidade.

A seguir, são apresentadas sugestões para a melhoria e/ou para a complementação dos resultados obtidos nesta pesquisa.

5.4 Proposição para Pesquisas Futuras

É importante, para a contínua evolução científica, que pesquisas sejam postas à prova, revalidadas, refutadas, estendidas e aprimoradas. Durante o desenvolvimento de qualquer ideia, é normal que novas ideias relacionadas surjam para contornar alguma barreira. Entretanto, nem sempre podem ser postas em prática. Nesses casos, é obrigação do pesquisador oferecer as linhas para o desenvolvimento de pesquisas futuras. Nesta seção, portanto, propõem-se seis sugestões para pesquisas futuras que poderão consolidar os fundamentos da pesquisa ora relatada.

A avaliação estatística realizada neste documento teve caráter estritamente quantitativo. Entretanto, foi observado o potencial da abordagem desenvolvida como ferramenta de auxílio à atividade de análise de vídeos de sessões de teste de usabilidade. Portanto, uma avaliação qualitativa da abordagem, em que um grupo de avaliadores da usabilidade de produtos emite sua opinião sobre a utilidade dos sumários produzidos com a abordagem proposta, se afigura apropriada.

O fluxo de tarefas da abordagem desenvolvida exige que seja criado pelo menos um modelo de forma, a ser utilizado como modelo de comparação para o restante do vídeo da sessão de teste de usabilidade. Os testes realizados nesta pesquisa utilizaram um quadro do próprio vídeo para a criação do modelo de comparação. O modelo inicial poderia ser derivado da própria execução da etapa de análise do vídeo. Tal estratégia eliminaria uma etapa de interação do avaliador que é potencialmente desnecessária.

De acordo com o que foi observado na seção 4.2, as diferentes estruturas internas da face (sobrancelhas, olhos, nariz e boca) contribuem de forma diferente para a taxa de erro medida na comparação com qualquer modelo. Este fenômeno se deve a duas características. Primeiro, à quantidade de pontos usados para descrever cada estrutura. Segundo, à quantidade de movimento médio de cada estrutura em relação às outras. Descrever uma técnica capaz de minimizar tais problemas pode trazer benefícios para o funcionamento da abordagem desenvolvida.

Uma das dificuldades encontradas e não solucionadas nesta pesquisa diz respeito à variação de pose da cabeça dos usuários de teste. A abordagem foi descrita e desenvolvida supondo que o usuário não desvia sua atenção da tarefa em execução. Alinhados a este fato, os testes descritos na seção 4.3.3 foram concebidos de forma a minimizar as chances de o usuário ser obrigado a variar a pose da cabeça durante a condução do teste. Contudo, ainda ocorreram casos de desvio da atenção. Replicar esta pesquisa utilizando uma técnica de reconhecimento de expressões faciais que seja invariante à pose é importante, a fim de consolidar a robustez da abordagem proposta.

Mesmo sendo uma dificuldade da abordagem proposta com relação aos objetivos delineados para esta pesquisa, o fato da técnica utilizada ser sensível à pose assumida pelo rosto dos usuários pode ser explorada em outros cenários, e.g. medição do nível de atividade ou agitação do usuário. Deste modo, a sensibilidade da técnica utilizada, que é uma característica destrutiva para os propósitos desta pesquisa, pode ser útil em outros cenários.

Embora a medida de similaridade utilizada pela abordagem tenha permitido constatar variações no estado das faces dos usuários de teste e correlacioná-las às medidas quantitativas anotadas manualmente, observou-se que eventos mais sutis por parte dos usuários de teste (e.g. sobrancelhas levemente baixadas) não estavam sendo capturados nos sumários de vídeo. Ainda assim, o módulo de captura de expressões faciais (Módulo AAM) foi preciso o suficiente para registrar graficamente a apresentação de referidos eventos. Esta observação leva a crer que o sistema de sumarização ainda pode ser refinado.

Referências Bibliográficas

AICHHOLZER, O; AURENHAMMER, F; HACKL, T; SPECKMANN, B. On minimum weight pseudo-triangulations. In: COMPUTATIONAL GEOMETRY: THEORY AND APPLICATIONS, 2009, Amsterdam, **Proceedings...** Amsterdam: Elsevier, 2009, p. 627-631.

ANIKEENKO, A. V. Application of Procrustes Distance to Shape Analysis of Delaunay Simplexes. In: VORONOI DIAGRAMS IN SCIENCE AND ENGINEERING, 2006, Alberta, **Proceedings...** Alberta: IEEE, 2006, p. 148-152

ARMAN, F; DEPOMMIER, R; HSU, A; CHIU, M.Y. Content-based video indexing and retrieval. In: ACM MULTIMEDIA (ACMMM'94), 1994, San Francisco, **Proceedings...** São Francisco: ACM, 1994, p. 97-103.

BARBOSA, A. E. V. **Abordagem híbrida para avaliação de interfaces multimodais.** 2009. 346f. Dissertação (Mestrado em Ciência da Computação) – Centro de Engenharia Elétrica e Informática, Universidade Federal de Campina Grande, 2009.

BEVAN, N; MACLEOD, M. Usability Measurement in Context. In: BEHAVIOUR AND INFORMATION TECHNOLOGY, 1994, London, **Proceedings...** London: Taylor & Francis, 1994.

BRANCO, P; ENCARNAÇÃO, L. M. E; MARCOS, A. F. It's all in the Face: Studies on Monitoring Users' Experience. In: THIRD IBEROAMERICAN SYMPOSIUM IN COMPUTER GRAPHICS, 3, 2006, Santiago de Compostela. **Proceedings...** Santiago de Compostela: Eurographics Association, 2006.

CHEW, S.; LUCEY, S; LUCEY, P; SRIDHARAN, S; COHN, J. Improved facial expression recognition via uni-hyperplane classification. In: COMPUTER VISION AND PATTERN RECOGNITION, 2012, Rhode Island. **Proceedings...** Rhode Island: IEEE, 2012.

CLEARLEFT. Silverback. Versão 2.0. Brighton, Reino Unido. Disponível em: <<http://www.silverbackapp.com/>>. Último acesso em: 16 dez. 2010.

COOTES, T. F; EDWARDS, G. J; TAYLOR, C. J. Active Appearance Models In: IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2, 1998, **Proceedings...**, :Springer, 1998, p. 484-498.

COXETER, H. S. M. Introduction to Geometry. Second edition. New York: John Wiley & Sons, inc. 1989. 496. (Wiley Classics Library)

EJAZ, N.; TARIQ, T.; BAIK, S. Adaptive Key Frame Extraction for Video Summarization Using a Aggregation Mechanism In: JOURNAL OF VISUAL COMMUNICATION AND IMAGE REPRESENTATION, 2012, **Proceedings...** , :Elsevier, 2012, p. 1031–1040.

EKMAN, P; FRIESEN, W.V. Facial Action Coding System: A technique for the measurement of facial movement. In: CONSULTING PSYCHOLOGIST PRESS, 1978, Palo Alto. **Proceedings...** Palo Alto: , 1978.

Elvesjö, J.; Skogö, M.; Eskilsson, H. Tobii Eye Tracker. Estocolmo, Suécia. Disponível em: < http://www.tobii.com/corporate/eye_tracking/what_is_eye_tracking.aspx>. Último acesso em: 16 dez. 2010.

FAULKNER, L. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. In: BEHAVIOR RESEARCH METHODS, INSTRUMENTS, & COMPUTERS, 2003, **Proceedings...** 2003, p. 379-383.

FERREIRA, D. S. **Abordagem Híbrida para Avaliação da usabilidade de Dispositivos Móveis**. 2007. 227f. Dissertação (Mestrado em Informática) – Pós-graduação em Informática, Universidade Federal de Ampina Grande, Campina Grande, 2007.

GALITZ, W. O. **The Essential Guide to User Interface Design**. Indianapolis:Wiley Publishing Inc, 2007.

- GELLNER, M; FORBRIG, P. ObSys – a Tool for Visualizing Usability Evaluation Patterns with Mousemaps. In: HCI INTERNATIONAL, Crete, 2003, **Proceedings... Crete: Lawrence Erlbaum Associates, Inc., 2003.**
- GODIL, A. ViSA: Video Segmentation and Annotation. In: UPA2004 conference, 2004, Minneapolis, **Proceedings... Minneapolis: NIST, 2004.**
- GONG, Y; LIU, X. Video Sumarization Using Singular Value Decomposition. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2000. Hilton Head Island, **Proceedings... Hilton Head Island: IEEE, 2000.**
- GUIBAS, L; RUSSEL, D. An empirical comparison of techniques for updating Delaunay triangulations. In: SCG '04 Proceedings of the twentieth annual symposium on Computational geometry, 2004. New York, **Proceedings... New York: ACM, 2004, p. 170-179.**
- HAN, B; Hamm, J; Sim, J. Personalized Video Summarization with Human in the Loop. In: APLICATIONS OF COMPUTER VISION (WACV), 2011, Pohang. **Proceedins... Pohang: IEEE, 2011, p. 51-57**
- HARRIS, S. R. Systemic-Structural Activity Analysis of Video Data: A Practical Guide. In: WORKSHOP ON ACTIVITY THEORY BASED PRACTICAL METHODS FOR IT DESIGN, 2004, Copenhagen. **Proceedings... Copenhagen: , 2004, p. 48-63**
- HILBERT, D. M; REDMILES, D. F. Extracting Usability Information from User Interface Events. In: ACM COMPUTING SURVAYS, 2000, New York. **Proceedings... New York: ACM, 2000, p. 384-421.**
- HOLZINGER, A. Usability Engeneering Methods for Software Developers. In: COMMUNICATIONS OF THE ACM, 2005, Graz. **Proceedings... Graz: ACM, 2005.**
- HUET, B; MERIALDO, B. **Signals and Communication Technology: Automatic Video Summarization.** Kokomo: Springer, 2005, p. 27-42.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, ISO 9241 Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on Usability, 1998.
- JACOB, R. J. K. New Human-Computer Interaction Techniques. In: ACM TRANSACTIONS ON COMPUTER-HUMAN INTERACTION, 1994, New York, **Proceedings... New York: ACM p. 258-291, 1994.**

JACOBSEN, N.E; HERTZUN M; JOHN, B.E. The evaluator effect in usability tests. In: CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 1998, Los Angeles. **Proceedings...** New York: ACM, 1998, p. 255-256.

KAIKKONEN, A. et al. Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing. In: JOURNAL OF USABILITY STUDIES, 2005, Sonera, **Proceedings...** Sonera: TeliaSonera, 2005, p. 4-17.

KANADE, T; COHN, J.F; TIAN, Y. Comprehensive Database for Facial Expression Analysis. In: PROCEEDINGS OF THE 4TH IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, 2000, Grenoble, **Proceedings...** Grenoble: IEEE, 2000, p. 46-53.

KASS, M; WITKIN, A.; TERZOPOULOS, D. Snakes: Active Contour Models, In: INTERNATIONAL JOURNAL OF COMPUTER VISION, 1988, Boston, **Proceedings...** Boston: Kluwer Academic Publishers, 1988, p. 321-331.

KOBAYASHI, H; HARA, F. Recognition of Six Basic Facial Expressions and Their Strength by Neural Network. In: IEEE INTERNATIONAL WORKSHOP ON ROBOT AND HUMAN COMMUNICATION, 1992, Tokyo, **Proceedings...** Tokyo: IEEE, 1992, p. 381-386.

KOTSIA, I; NIKOLAIDIS, N; PITAS, I. Fusion of Geometrical and Texture Information for Facial Expression Recognition. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, 2006, Atlanta, **Proceedings...** Atlanta: IEEE, 2006, p. 2649-2652.

KOTSIA, I; PITAS, I. Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines. In: IEEE TRANSACTIONS ON IMAGE PROCESSING, 16, 2007, **Proceedings...** 2006, p. 172-187.

KOTSIA, I. Real Time Facial Expression Recognition From. Image Sequences Using Support Vector Machines. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, 2005, **Proceedings...** 2005, p. 42-45.

KUMAR, K; OWSTON, R. The Open Virtual Usability Lab: A New Tool for Research on Learning Technologies, Disponível em: <http://www.yorku.ca/irit/reports/TechReport2012-1.pdf>. 2012.

KUSHNIRUK, A. W; PATEL, V. L; CIMINO, J. J. Usability testing in medical informatics: cognitive approaches to evaluation of information systems and user interfaces. In: AMIA ANNUAL SYMPOSIUM, 1997, **Proceedings...** Nashville: AMIA, 1997.

LEHMANN, E.L; CASELLA, G; OLKIN, I. Theory of Point Estimation. Second Edition. London: Springer, 1998. p. 616. (Springer Texts in Statistics)

LI, S; SHANG, J; SLAUGHTER S. Why do Software Firms Fail? Capabilities, Competitive Actions and Firm Survival in the Software Industry from 1995 to 2007, In: INFORMATION SYSTEMS RESEARCH JOURNAL, 2010, **Proceedings...** Maryland: Institute for Operations Research and the Management Sciences (INFORMS), 2010, p. 631-654.

LINDGREEN, B. W. Statistical Theory, London: Collier-Macmillan Limited, 1968. 521 p.

MACLEOD, M; BOWDEN, R; BEVAN, N. The MUSiC Performance Measurement Method. In: BEHAVIOUR AND INFORMATION TECHNOLOGY, 1998, Loughborough, **Proceedings...** Loughborough: Cambridge University Press, 1998.

MACLEOD, M; RENGGER, R. The Development of DRUM: A Software Tool for Video-assisted Usability Evaluation. In: PROCEEDINGS OF HCI, 1993, London, **Proceedings...** London: Taylor & Francis, 1993.

MAITY, S; CHAKRABARTI, A; BHATTACHARJEE, D. Feature Based Information Extraction for Genetic Video Summarization. In: IJCA PROCEEDING ONINTERNATIONAL CONFERENCE ON RECENT ADVANCES AND FUTURE TRENDS IN INFORMATION TECHNOLOGY, 2012, **Proceedings...** New York: Foundation of Computer Science, 2012, p. 29-34.

NGO, C; MA, Y; ZHANG, H. Video Summarization and Scene Detection by Graph Modeling. In: IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, 2005, **Proceedings...** Gaithersburg: IEEE, 2005.

NIELSEN, J. **Usability Engineering**. Morgan Kaufmann, 1993.

NOLDUS. FaceReader Methodology, Disponível em: <http://info.noldus.com/free-white-paper-on-facereader-methodology>. 2012.

NORMAN, D. A. **The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex, and Information Appliances Are the Solution.** The MIT Press, 1998.

OLIVEIRA, R. C. L; **WebQuest: Uma Ferramenta Web Configurável para a Sondagem da Satisfação do Usuário.** 2005. 217f. Dissertação (Mestrado em Informática) – Pós-graduação em Informática, Universidade Federal de Ampina Grande, Campina Grande, 2005.

PFEIFFER, S; LIENHART, R; FISCHER, S; EFFELSBURG, W. Abstracting digital movies automatically. In: JOURNAL OF VISUAL COMMUNICATION AND IMAGE REPRESENTATION, 1996, **Proceedings...** Mannheim: ELSEVIER, 1996 p. 345–353.

QUEIROZ, J. E. R. **Abordagem Híbrida para a Avaliação da Usabilidade de Interfaces com o usuário.** 2001. 410f. Tese (Doutorado em Engenharia Elétrica) – Pós-graduação em Engenharia Elétrica, Universidade Federal da Paraíba, Campina Grande, 2001.

RABINER, L. R. A Tutorial on Hidden Markov Models and selected application in speech recognition. In: Proceedings of IEEE, 1989, **Proceedings...** vol. 77, 1989, p. 257-286.

REN, T; LIU, Y; WU, G. Video Summary Quality Evaluation Based on 4C Assessment and User Interaction. In: SHAO, L; CAIFENG, S; LUO, J; ETOH (ed), M. *Multimedia Interaction and Intelligent User Interfaces: Principles, Methods and Applications.* Londres: Springer, 2010. p. 243-269.

ROCHA, H. V; BARANAUSKAS, M. C. C. **Design e Avaliação de Interfaces Humano-Computador.** Campinas: Unicamp, 2003.

ROGERS, M; GRAHAM, J. Structured Point Distribution Models: Modelling Intermittently Present Features. In: 12th British Machine Vision Conference, 2001, **Proceedings...** Manchester. 2001, p. 33-42.

RUBIN, J. **Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests.** New York: John Wiley e Sons, Inc., 1994.

RYAN, A; COHN, J.F; LUCEY, S; SARAGIH, J; LUCEY, P; DE LA TORRE, F; ROSSI, A. Automated Facial Expression Recognition System. In: INTERNATIONAL CARNAHAN CONFERENCE ON SECURITY TECHNOLOGY, 2009, **Proceedings...** Zurich: IEEE, 2009, p. 172-177.

Senechal, T; Rapp, V; Salam, H; Seguier, R; Bailly, K; Prevost, L. Facial Action Recognition Combining Heterogeneous Features via Multikernel Learning. In: IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, 2012, **Proceedings...** :IEEE, 2012, p. 993-1005.

SMITH, M.A; KANADE, T. Video skimming and characterization through the combination of image and language understanding. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 1997, **Proceedings...** Puerto Rico: IEEE, 1997, p. 775-781.

SMOLIAR, S; ZHANG, H. Content-based video indexing and retrieval. In: IEEE MULTIMEDIA MAGAZINE, 1994, **Proceedings...** : IEEE, 1994, p. 62-72.

STAIANO, J; MENÉNDEZ, M; BATTOCCHI, A; DE ANGELI, A; SEBE, N. Ux_Mate: from facial expression to UX evaluation. In: PROCEEDINGS OF THE DESIGNED INTERACTIVE SYSTEMS CONFERENCE, 2012, **Proceedings...** New York: ACM, 2012, p. 741-750.

STEGMANN, M. B; ERSBØLL B. K; LARSEN, R. FAME – A Flexible Appearance Modelling Environment. In: IEEE Transactions on Medical Imaging, 2003, **Proceedings...** : IEEE, 2003, p. 1319-1331.

SUTCLIFFE, A; RYAN, M.; DOUBLEDAY, A; SPRINGETT, M. Model mismatch analysis: towards a deeper explanation of users' usability problems. In: BEHAVIOUR AND INFORMATION TECHNOLOGY, 2000, **Proceedings...** London: Taylor and Francis, 2000.

TECHSMITH CORPORATION. Morae. Versão 3.2. Okemos, Michigan, USA. Disponível em: <<http://www.techsmith.com/morae.asp>>. Último acesso em: 16 dez. 2010.

TIAN, Y; KANADE, T; COHN, J. F. **Handbook of face recognition**. Berlin: Springer, 2005.

TONOMURA, Y; AKUTSU, A; OTSUJII, K; SADAKATE, T. Videomap and videospaceicon: Tools for anatomizing video content. In: INTERCHI 93, 1993 **Proceedings...** New York: ACM, 1993, p. 131–141.

TRUONG, B. T.; VENKATESH, S. Video abstraction: A systematic review and classification. In: ACM TRANSACTION ON MULTIMEDIA COMPUTING, COMMUNICATIONS, AND APPLICATIONS., 2007, **Proceedings...** New York: ACM, 2007.

UDDIN, Z.M.D; LEE, J. J; KIM, T. S. An Enhanced Independent Component-Based Human Facial Expression Recognition from Video. In: IEEE TRANSACTIONS ON CONSUMER ELECTRONICS, 2009, **Proceedings...** 2009, p. 2216-2224.

Universiteit van Amsterdam, eMotion, Amsterdã. Disponível em: <<http://www.visual-recognition.nl/index.html>>. Último acesso em 15 jan. 2011.

VERMEEREN, A. **What's the problem?** Studies on identifying usability problems in user tests. 2009. 186f. Tese (Doutorado em Computação) – Delft University of Technology, Delft, 2009.

YAHIAOUI, I; Merialdo, B; HUET, B. Automatic Video Summarization. In: MMCBIR 2001 - Indexation et Recherche par le Contenu dans les Documents Multimedia, 2001, **Proceedings...** INRIA: Rocquencourt, 2001.

ZHANG, S; ZHAO, X; LEI, B. Facial Expression Recognition based on Local Binary Patterns and Local Fisher Discriminant Analysis. In: WSEAS TRANSACTIONS ON SIGNAL PROCESSING, 2012, **Proceedings...** Wisconsin, 2012.v. 8, p. 21-31.

ZHANG, Y; BENHAMZA, Y; IDRISSE, K; GARCIA, C. Probabilistic Active Appearance Models. In: DANS COMPRESSION ET PRÉSENTATION DES SIGNAUX AUDIOVISUALES, 2012, **Proceedings...** Lille, 2012.

ZHOU, T; WU, X; WU, T; FENG, Z. An Improved AAM Method for Extracting Facial Features. In: J. APPLIED MATHEMATICS, 2012, **Proceedings...** 2012.

Anexo A Termos e Condições para o Uso do Banco de Expressões Faciais Cohn-Kanade

Neste anexo encontra-se o documento de aceitação das condições de uso do banco de imagens de expressões faciais Cohn-Kanade.

AGREEMENT ON USE OF IMAGE DATA

Cohn-Kanade Facial Expression Database

I agree

- to cite Kanade, Cohn, & Tian (2000) in any paper of mine or my collaborators that makes any use of the database. The reference is:

Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, Grenoble, France, 46-53.

- to use the images for research purposes only.
- not to provide the images to second parties.
- if I reproduce images in electronic or print media, to use only those from the following subjects and include notice of copyright (©Jeffrey Cohn).

S52	S55	S74	S106	S111	S113
S121	S124	S125	S130	S132	

Signature: _____

Name: _____

Title: _____

Institution: _____

Date: _____

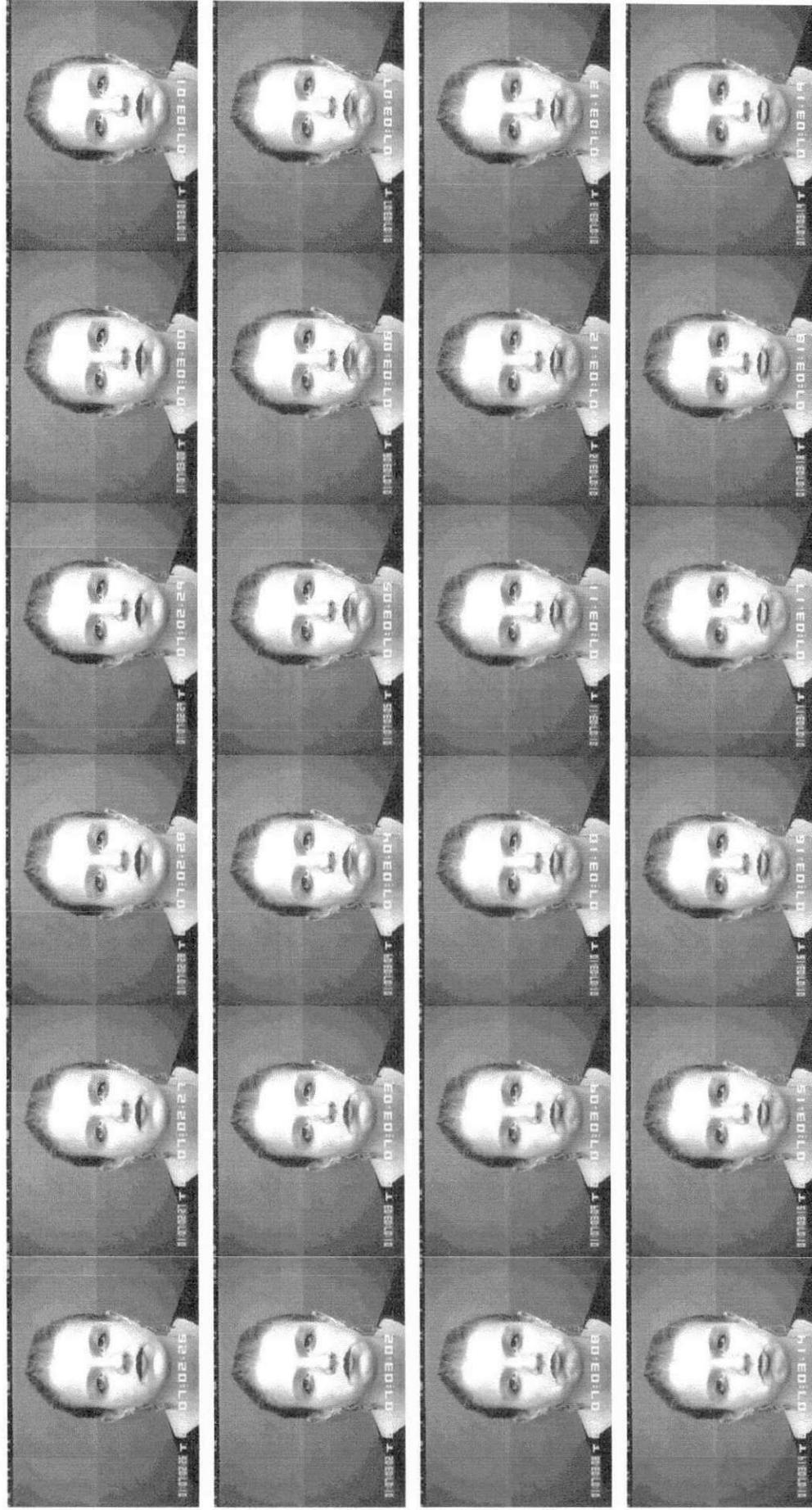
Email address: _____

Anexo B Banco de Imagens Cohn-Kanade

Neste anexo está exibida parte das seqüências de imagens que compõem o banco de imagens Cohn-Kanade. O banco de Imagens de expressões faciais construído por Cohn-Kanade é composto de 490 seqüências de imagens, totalizando 8789 imagens, que representam a evolução de expressões faciais exibidas por voluntários.

S052 ©Jeffrey Cohn

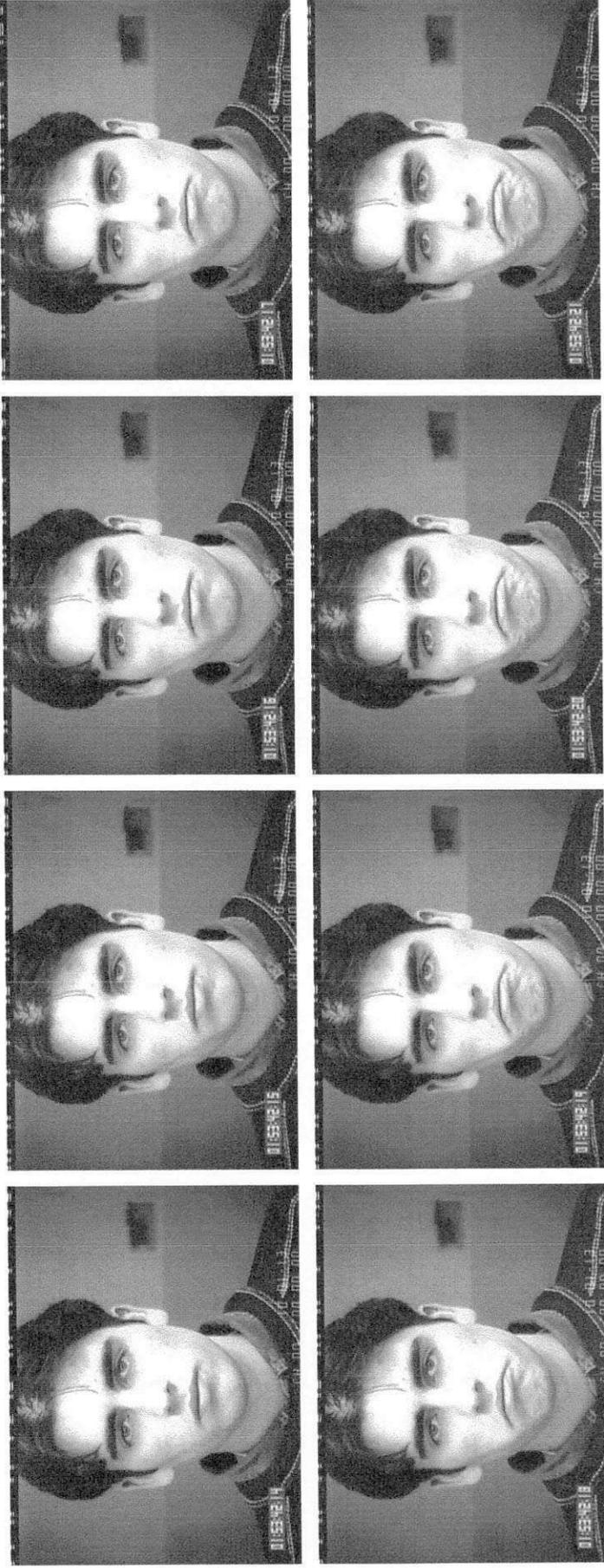




S106 ©Jeffrey Cohn







S124 ©Jeffrey Cohn



Apêndice A Imagens Anotadas para Treinamento

Dentre as 8789 imagens que compõem o banco de imagens Cohn-Kanade, foram escolhidas, aleatoriamente, 2000 imagens que foram anotadas manualmente. Neste apêndice está exibido um subconjunto das imagens do banco de imagens Cohn-Kanade cujos pontos fiduciais foram anotados para o treinamento da técnica AAM que foi desenvolvida nesta pesquisa.

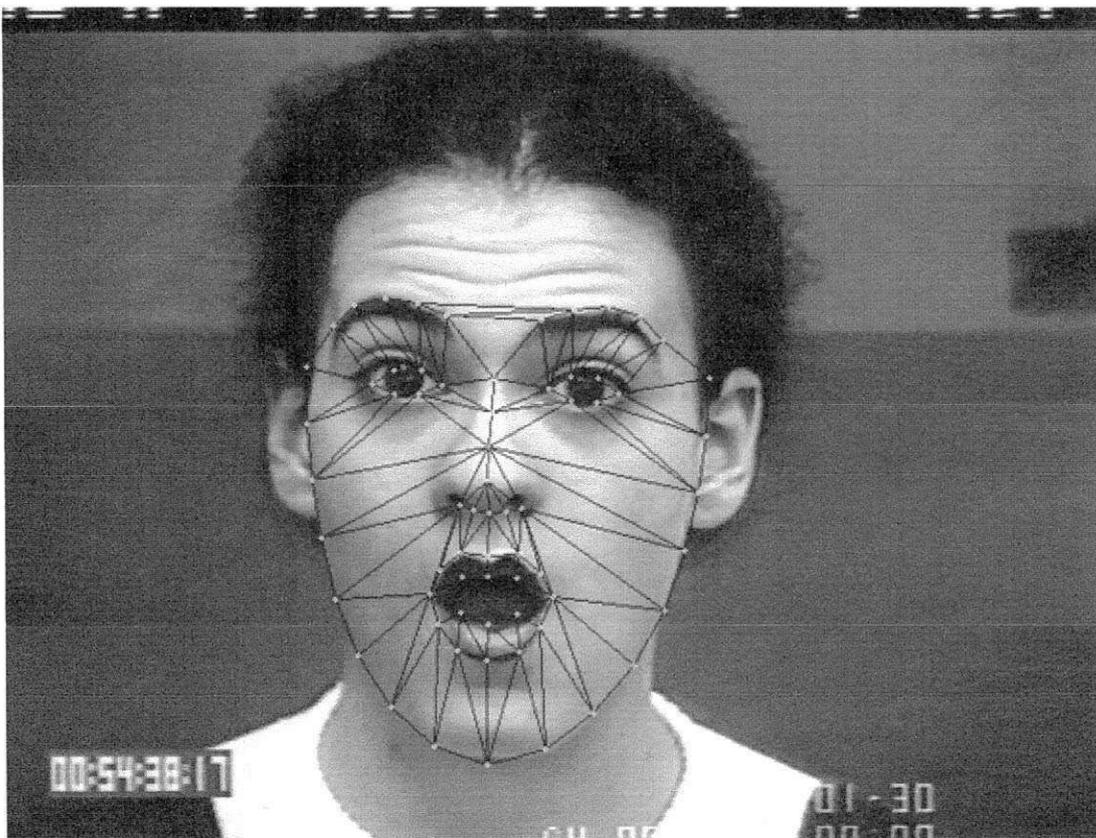
S055 ©Jeffrey Cohn



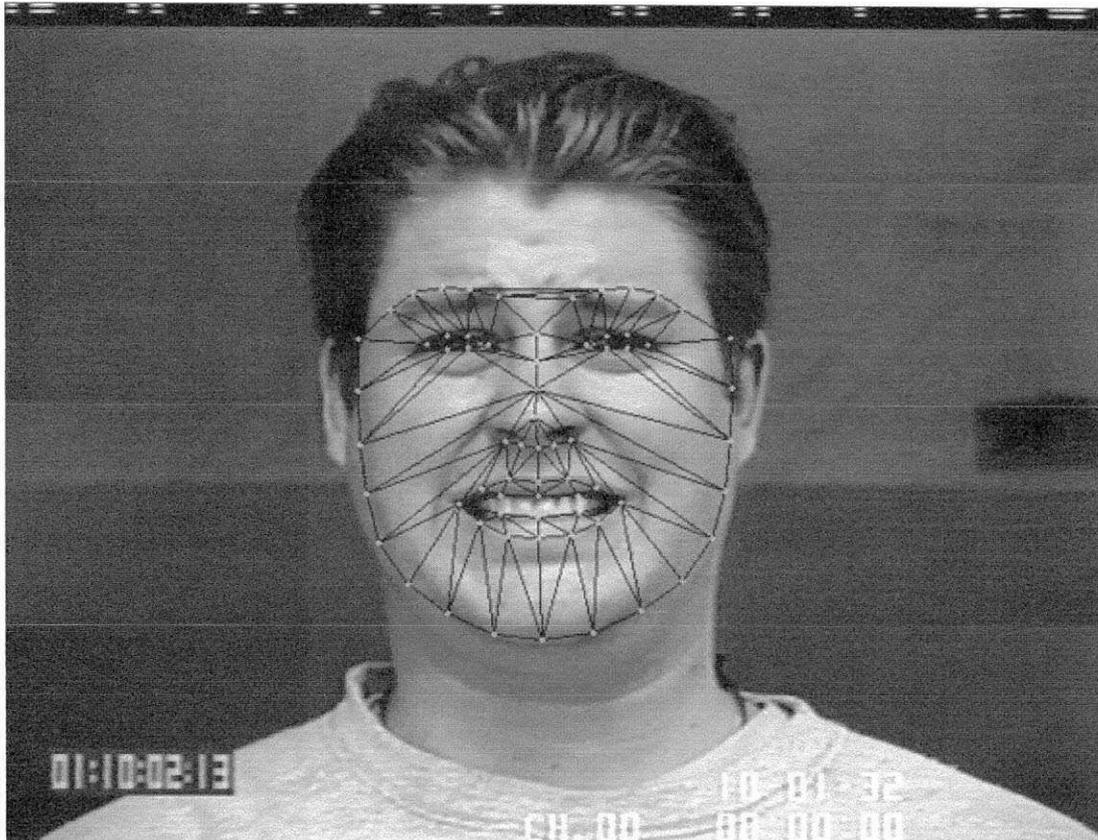
S124 ©Jeffrey Cohn



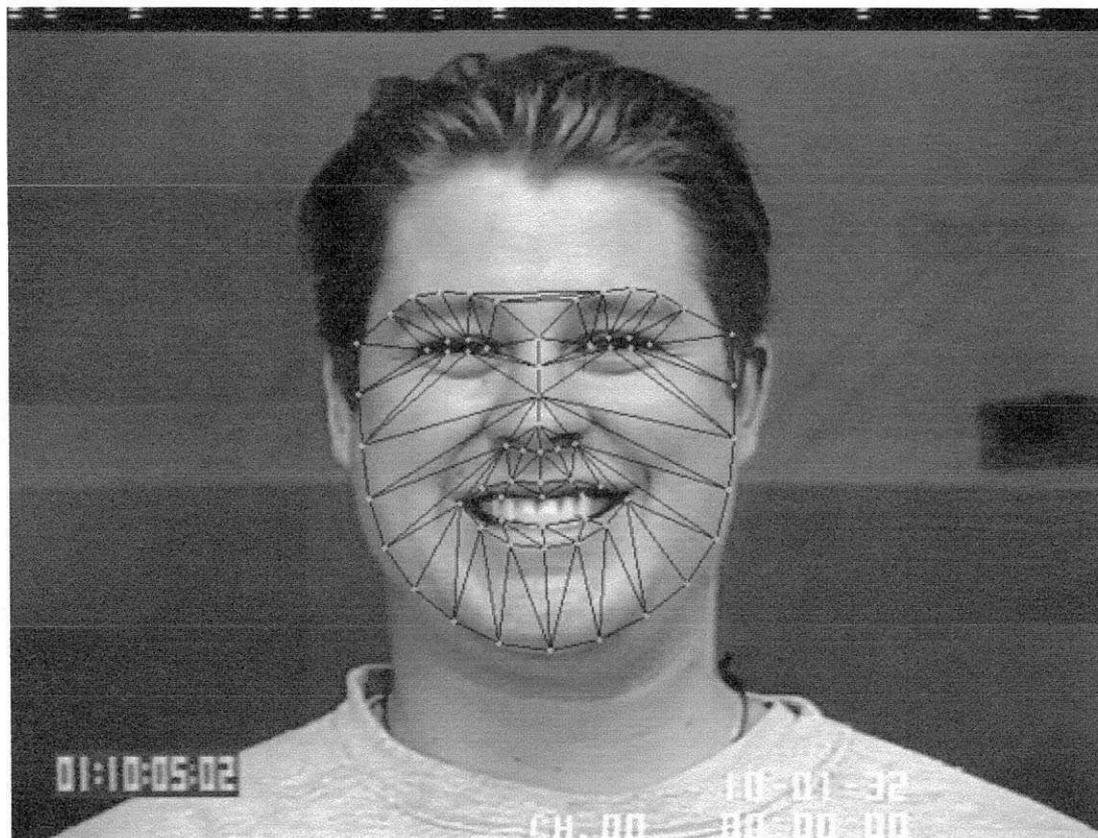
S130 ©Jeffrey Cohn



S132 ©Jeffrey Cohn



S132 ©Jeffrey Cohn



Apêndice B Amostras de Imagens Ajustadas após o Treinamento

Neste Apêndice estão exibidos alguns resultados do ajuste automático realizado pela ferramenta desenvolvida.

Imagens com Ajuste Adequado

As imagens exibidas nesta seção do apêndice exibem exemplos de imagens cujo ajuste automático, do modelo AAM desenvolvido nesta pesquisa, foi realizado corretamente.

S111 ©Jeffrey Cohn



S124 ©Jeffrey Cohn



S130 ©Jeffrey Cohn



Imagem com Ajuste Inadequado

Nesta seção do apêndice está exibida uma das imagens que exibe um ajuste inadequado.

S052 ©Jeffrey Cohn



Apêndice C Formulário para o Cadastro de Participação

Este apêndice contém o formulário para o cadastro de participação nas sessões de teste que foram realizadas nesta pesquisa. Os dados contidos neste formulário tem a finalidade de facilitar a convocação para a realização de cada sessão de teste

Nome

Endereço Residencial

Logradouro	Número
Complemento	
Bairro	CEP
Cidade	Estado
Telefone	Celular
E-mail	
Meio Preferencial para contato	Horário Preferencial para contato

Observações

<hr/> <hr/> <hr/> <hr/> <hr/> <hr/>

Apêndice D Documento de Aceitação das Condições de Teste

Este apêndice contém a autorização para a utilização das imagens e sons gravados na realização de cada sessão de teste.

Autorização

Autorizo a utilização das imagens e sons registrados durante a sessão de teste, realizada em _____ de _____ de _____.

Saliento que tais imagens e sons poderão ser utilizados para fins de análise de dados e geração de relatórios.

Campina Grande, _____ de _____ de _____.

Nome:

CPF:

RG:

Apêndice E Termo de Confidencialidade

O termo de confidencialidade, contido neste apêndice, tem a finalidade de reduzir a chance de que informações sobre cada teste seja propagada e possam afetar a forma com que usuários venham a reagir durante a realização de seus respectivos testes.

Termo de CONFIDENCIALIDADE

Comprometo-me a manter completo e absoluto sigilo, em relação a quais quer dados, materiais, informações transmitidas, documentos, de que venha a ter conhecimento, ou acesso de forma verbal e/ou escrita; ou que a mim venha a ser confidenciado em razão deste teste, realizado em _____ de _____ de _____. Não podendo, sob qualquer pretexto, reproduzir, divulgar, ceder, doar, explorar, comercializar, revelar, utilizar ou de qualquer modo dar conhecimento a terceiros.

Declaro estar ciente de que, na forma de lei, sou responsável civilmente pela divulgação indevida, descuidada ou incorreta utilização das informações de natureza confidencial de me tenham sido reveladas.

Campina grande, _____ de _____ de _____.

Nome:

CPF:

RG:

Apêndice F Questionário para o Delineamento do Perfil do Usuário

Este questionário tem a finalidade de delinear o perfil de cada usuário de teste. O perfil dos usuários de teste é avaliado na seção 4.3.4 (Análise do Perfil do Usuário) desta dissertação.

1. Qual o seu grau de instrução?
 Ensino médio incompleto Ensino superior completo
 Ensino médio completo Pós-graduação incompleto
 Ensino superior incompleto Pós-graduação completo
2. Qual o seu sexo?
 Masculino Feminino
3. Você usa óculos ou lentes de contato?
 Sim Não
4. Em que faixa etária você se encaixa?
 Menor de idade 30-35
 18-23 acima de 35
 24-29
5. Você tem experiência prévia no uso de computadores? (Caso sua resposta seja 'Não', marque a opção 'Não se aplica' nas questões de seis até quatorze e pule para a última questão.)
 Sim Não
6. Há quanto tempo você usa computadores?
 Menos de três meses Mais de um ano
 De três meses a um ano Não se aplica
7. Com que frequência você usa computadores?
 Diariamente Ocasionalmente
 A cada semana
 A cada mês Não se aplica
8. Que sistema operacional você utiliza com mais frequência?
 Windows Não sabe responder
 Linux
 Outro: Não se aplica
9. Qual o seu nível de conhecimento de informática?

- Básico Programação
 Intermediário
 Avançado Não se aplica

10. Você já usou algum aplicativo para composição de apresentações multimídia (PowerPoint, Keynote, IMPRESS)? (Caso sua resposta seja 'Não', marque a opção 'Não se aplica' na questão onze.)

- Sim Não

11. Com que frequência você usa o seu aplicativo de composição de apresentações multimídia?

- Diariamente Ocasionalmente
 A cada semana
 A cada mês Não se aplica

12. Você já usou a suíte de aplicativos do *OpenOffice.org*? (Caso sua resposta seja 'Não', marque a opção 'Não se aplica' nas questões treze e quatorze.)

- Sim Não

13. Você já usou o aplicativo *IMPRESS*?

- Sim Não
 Não se aplica

14. Com que frequência você utiliza o aplicativo *IMPRESS*?

- Diariamente Ocasionalmente
 A cada semana
 A cada mês Não se aplica

15. Qual o seu nível de habilidade com a língua inglesa?

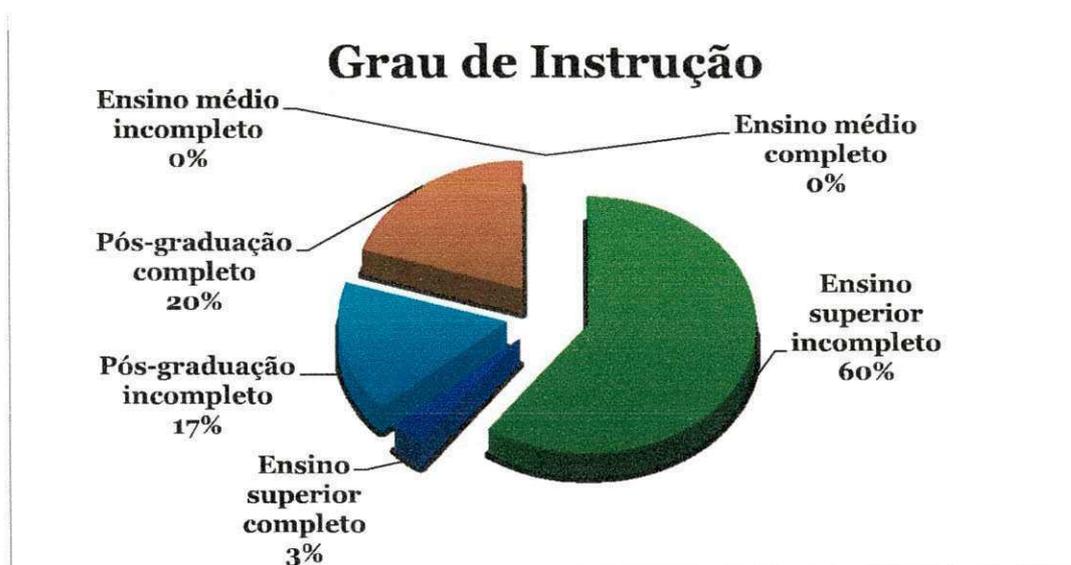
	Nenhum	Básico	Intermediário	Avançado
Lê		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Escreve	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fala		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Compreende		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Apêndice G Análise dos Dados dos Questionários para o Delineamento do Perfil do Usuário

Neste apêndice são avaliados os dados obtidos dos questionários para o delineamento do perfil do usuário, exibido no Apêndice F. Os dados dos questionários para o delineamento do perfil do usuário está exibido no Apêndice H onde há uma síntese dos dados coletados.

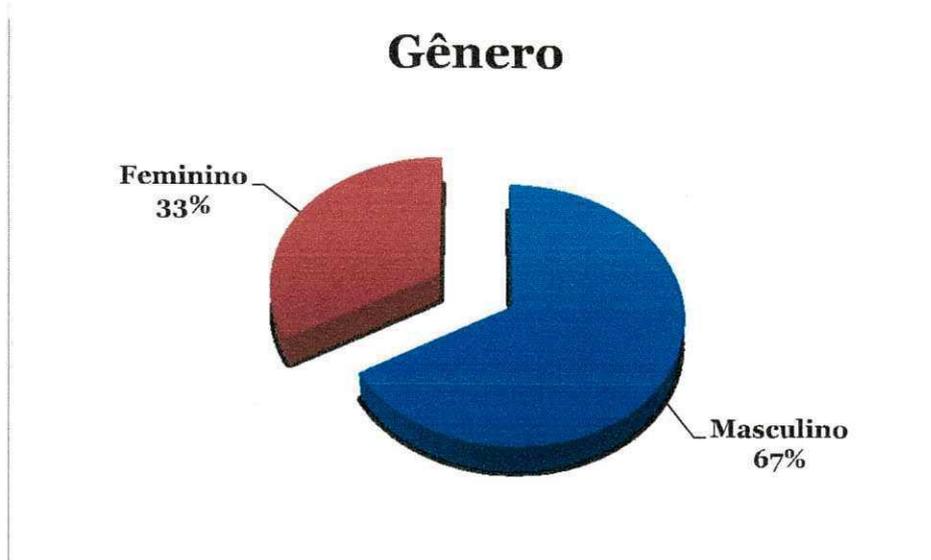
A Figura 28 corresponde ao primeiro item do questionário de sondagem do perfil do usuário. De acordo com a figura, os usuários de teste deste experimento são, em sua maioria, estudantes de graduação (18 indivíduos representando 60% do espaço amostral), seguido por estudantes de pós-graduação (mestrado e doutorado) (11 indivíduos representando 37% dos usuários).

Figura 28 – Distribuição numérica do Grau de Instrução dos usuários de teste



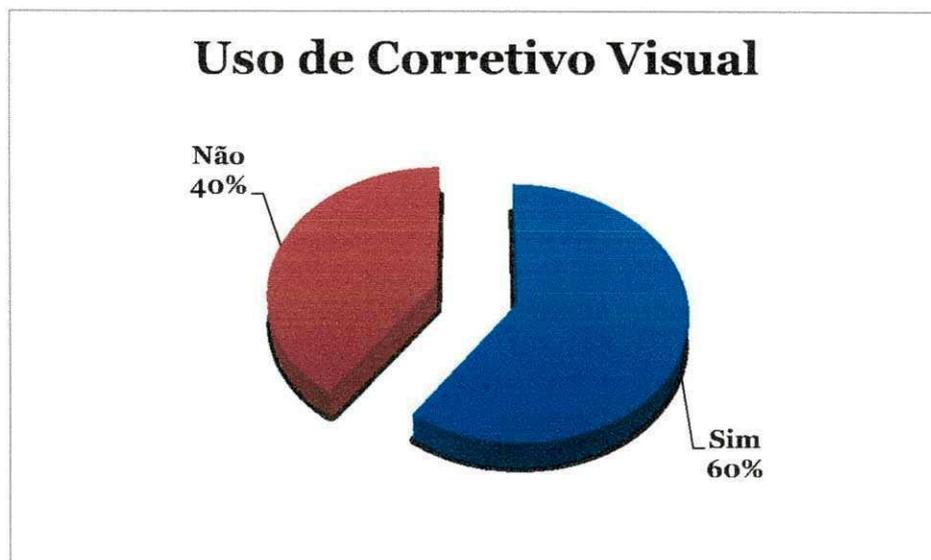
Na Figura 29 ilustra-se a distribuição dos usuários de teste em função dos gêneros. De acordo com a figura, o sexo masculino foi predominante entre os usuários deste experimento (20 representando dois terços do espaço amostral).

Figura 29 – Distribuição numérica do Gênero dos usuários de teste



Taxa semelhante dos usuários de teste deste experimento declarou usar algum tipo de corretivo visual (óculos ou lentes de contato) (18 indivíduos representando 60% do espaço amostral) (Figura 30).

Figura 30 – Distribuição numérica dos usuários de teste que usam algum tipo de corretivo visual



Com relação à faixa etária dos indivíduos que participaram do experimento, graficamente representada na Figura 31, pode ser observada a predominância de indivíduos na faixa etária entre 18 e 23 anos de idade (20, representando dois terços do espaço amostral), seguidos de indivíduos na faixa de 24 a 29 anos (8 indivíduos representando 27% do espaço amostral). A conjunção destas duas faixas etárias se faz corresponder a 95% do espaço

amostral.

Figura 31 – Distribuição numérica da Faixa Etária dos usuários

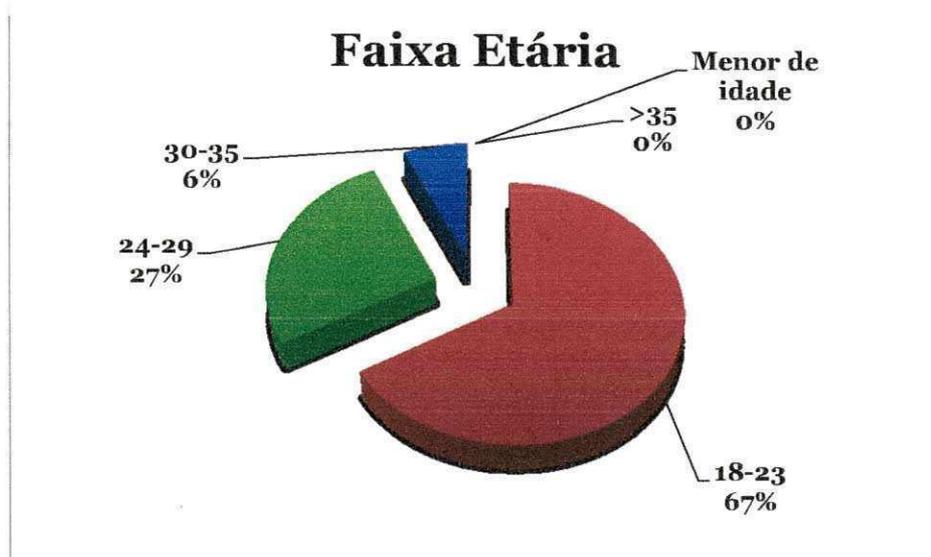


Figura 32 – Distribuição numérica dos Ambiente Operacionais mais utilizados pelos usuários de teste



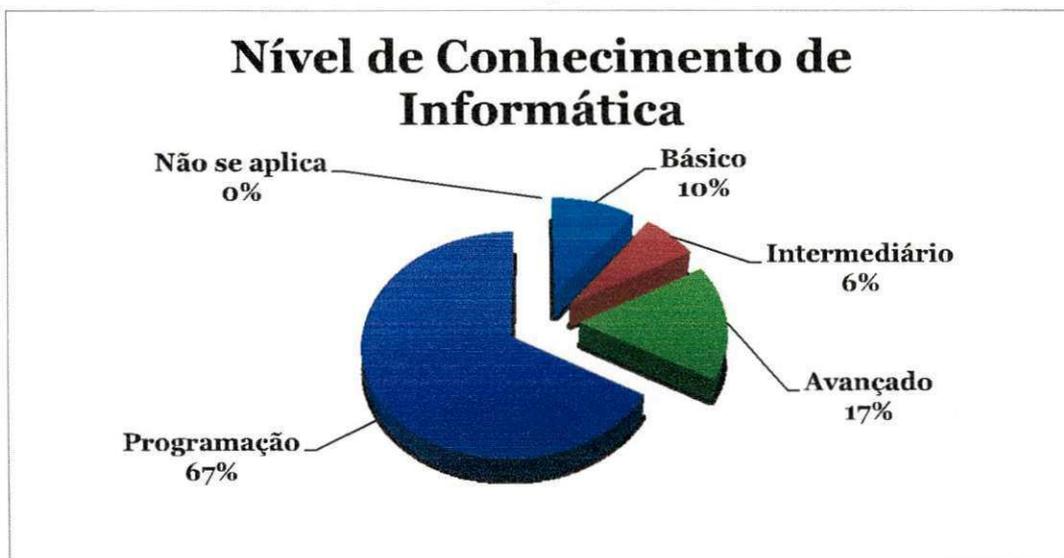
Todos os usuários de teste que participaram deste experimento possuem experiência prévia no uso de computadores, sendo este um pré-requisito para a participação do candidato no experimento descrito nesta seção. Coincidentemente, todos os candidatos a participar deste experimento possuíam experiência de mais de um ano com sistemas computacionais e atualmente utilizam os referidos sistemas diariamente.

Na Figura 32, estão representados os ambientes operacionais mais utilizados pelos usuários no espaço amostral do experimento. Conforme se

pode observar, a maior parte dos usuários utiliza o ambiente operacional *Windows* (22 indivíduos representando 69% dos usuários), seguido do ambiente *Linux* (8 indivíduos representando 25% dos usuários), dois usuários selecionaram ambos os ambiente *Windows* e *Linux*, então há uma pequena sobreposição entre estes ambiente. Outros dois usuários escolheram outro ambiente, e declararam que o ambiente operacional mais utilizado por eles é o *OSX*.

Na Figura 33, representa-se o nível de conhecimento de informática que cada candidato declarou. Neste item do questionário, os usuários de teste selecionaram uma dentre quatro opções de níveis de conhecimento fornecidas. Além das opções normais de níveis de conhecimento básico, intermediário e avançado, os usuários deste experimento tiveram a opção de escolher o nível de programação. A opção extra foi colocada considerando que uma parte significativa dos usuários de teste seria de estudantes do curso de Ciência da Computação.

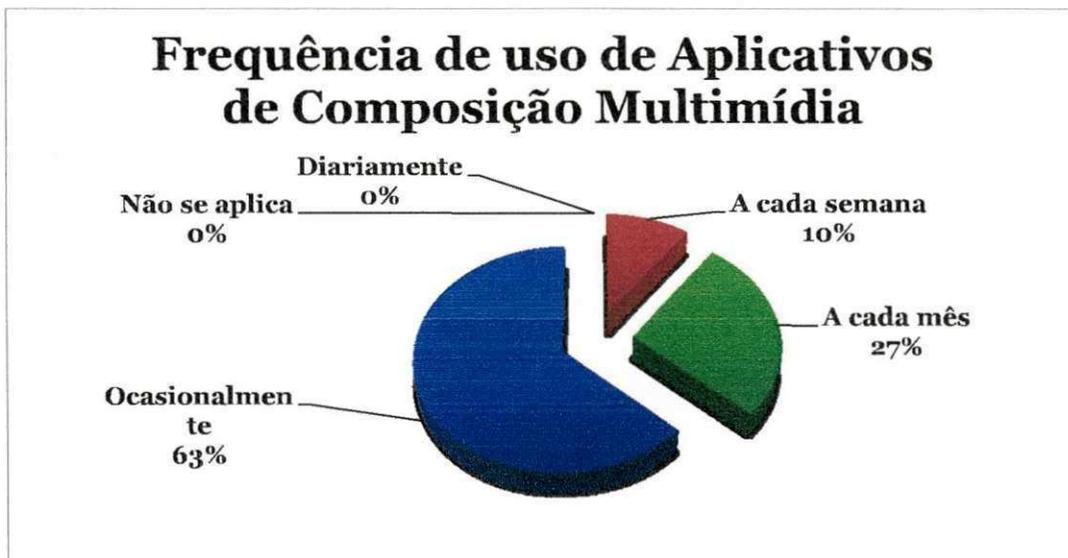
Figura 33 – Distribuição numérica do nível de conhecimento em informática declarado pelos usuários



O resultado da sondagem do nível de conhecimento de informática dos usuários, na Figura 33, mostra que há uma predominância dos que se classificaram no nível de programação (20 indivíduos representando dois terços do espaço amostral), seguidos de usuários que se declararam avançados (5 indivíduos representando 17% dos usuários).

A sondagem mostrou que todos os usuários de teste já utilizaram algum aplicativo de composição de apresentações multimídia. A frequência de uso de tais aplicativos está descrita na Figura 34. Verifica-se que nenhum dos usuários deste experimento utiliza aplicativos de composição de apresentações multimídia diariamente. A maior parte dos usuários utiliza os aplicativos descritos com uma frequência inferior a uma vez a cada mês (19 indivíduos representando 63% dos usuários). A parcela seguinte dos usuários utiliza aplicativos de composição de apresentações multimídia ao menos uma vez a cada mês (8 indivíduos representando 27% do espaço amostral).

Figura 34 – Distribuição numérica da Frequência de Uso de Aplicativos de Composição multimídia



A suíte de aplicativos do LibreOffice.org utiliza uma interface semelhante para todos os produtos que fazem parte da suíte. Desta forma, a familiaridade que os usuários de teste possam ter com a interface do aplicativo que será usado é um fator a ser considerado. Neste sentido, representa-se na Figura 35 a proporção dos usuários que já utilizaram aplicativos da suíte de aplicativos para escritório do LibreOffice.org. A maior parte dos usuários declarou que já utilizaram algum aplicativo da suíte (25 indivíduos representando 83% do espaço amostral).

Figura 35 – Distribuição numérica da experiência com a suíte de aplicativos do *LibreOffice.org*



Além da experiência com aplicativos da suíte de aplicativos do *LibreOffice.org*, outro fator importante foi o conhecimento dos usuários do próprio aplicativo sob avaliação. Na Figura 36 representa-se a proporção dos usuários que participaram do teste que já utilizaram o aplicativo *IMPRESS*. Apenas uma pequena parcela dos usuários de teste já utilizou o aplicativo sob avaliação (7 indivíduos representando 23% do espaço amostral).

Figura 36 – Distribuição numérica da Experiência com o Aplicativo *IMPRESS*



Ainda referente à experiência do usuário, é sumariada na Figura 37. O item 14 do questionário para o delineamento do perfil do usuário diz respeito à frequência de uso do aplicativo *IMPRESS*. De acordo com a Figura 37, dos usuários que, alguma vez, já utilizaram o aplicativo *IMPRESS*, apenas um utilizava o aplicativo ao menos uma vez ao mês (3% do espaço amostral). O

restante dos usuários declarou utilizar o aplicativo com frequência inferior a uma vez ao mês (6 usuários, representando 20% do espaço amostral).

O último item do questionário para o delineamento do perfil do usuário faz referência à habilidade dos usuários em relação à língua inglesa. De acordo com as figuras 38, 39, 40 e 41, as quais sumariam a habilidade dos usuários com respeito à leitura, escrita, fala e compreensão da língua inglesa, o nível de habilidade dos usuários se concentra no intermediário. A quantidade de usuários no nível de conhecimento intermediário está entre 14 (na habilidade de leitura) (representando 46% do espaço amostral) e 17 usuários (representando 57% do espaço amostral) (na capacidade de compreensão). No geral, o nível básico de cada habilidade é o segundo nível predominante, com a quantidade de usuários variando desde 7 (23% do espaço amostral) (na capacidade de compreensão) até 10 (34% do espaço amostral) (na habilidade de escrita). A exceção ocorreu na habilidade de escrita dos usuários, para a qual o segundo nível de habilidade predominante é o avançado, conforme declarado por 11 dos usuários (representando 37% do espaço amostral).

Figura 37 – Distribuição numérica da frequência de Uso do Aplicativo *IMPRESS*

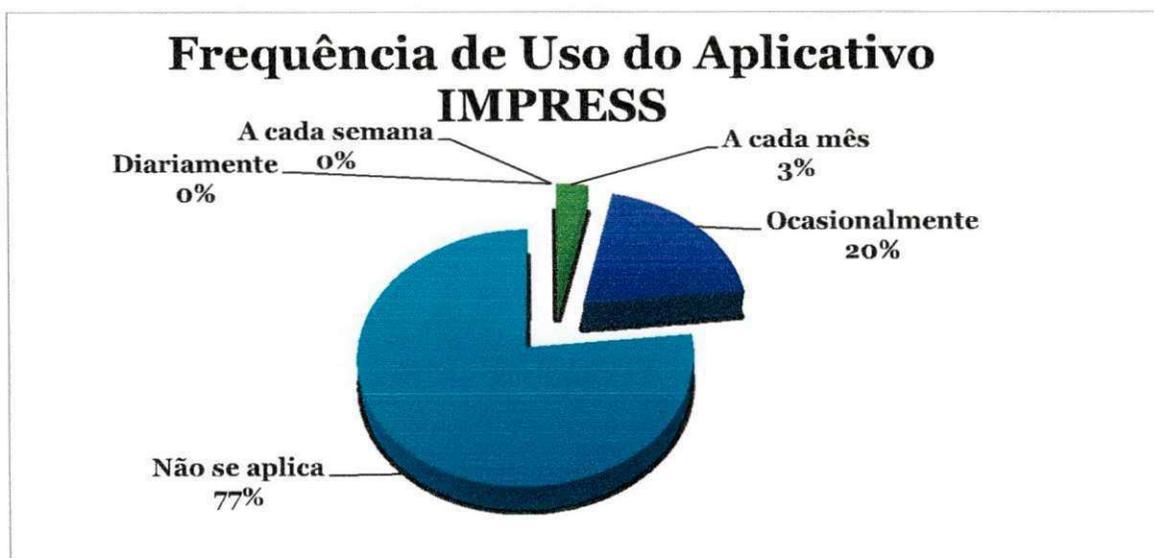


Figura 41 – Distribuição numérica da Habilidade de Leitura da Língua Inglesa

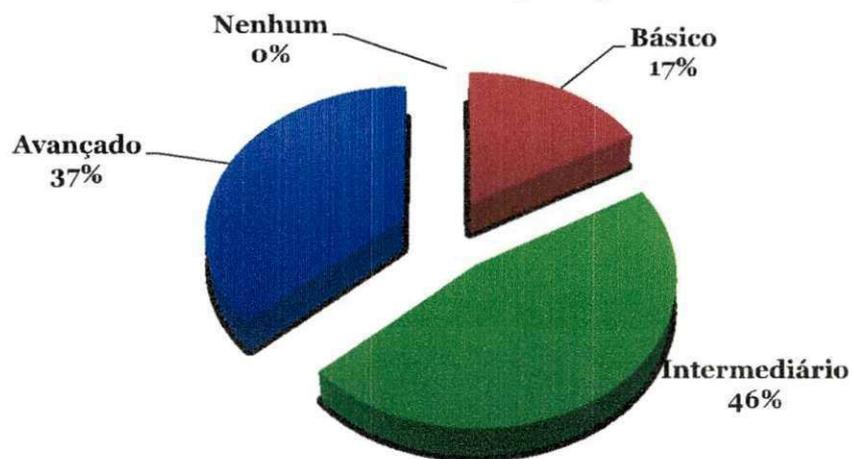


Figura 39 – Distribuição numérica da Habilidade de Escrita da Língua Inglesa

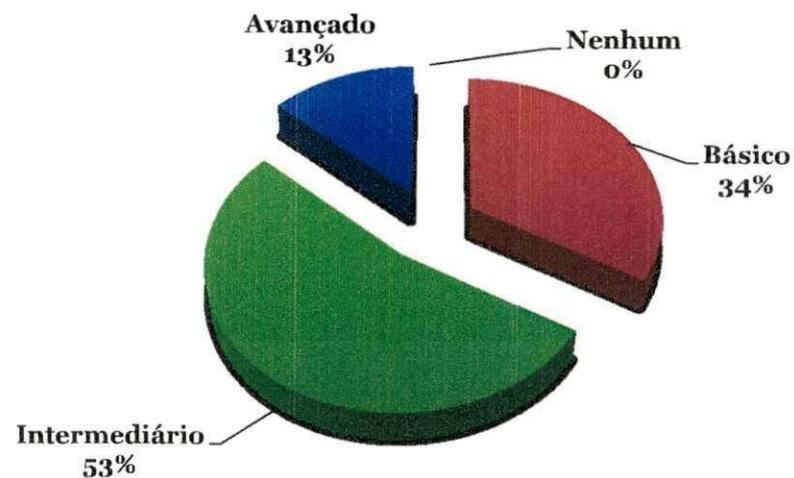


Figura 38 – Distribuição numérica da Habilidade de Fala da Língua Inglesa

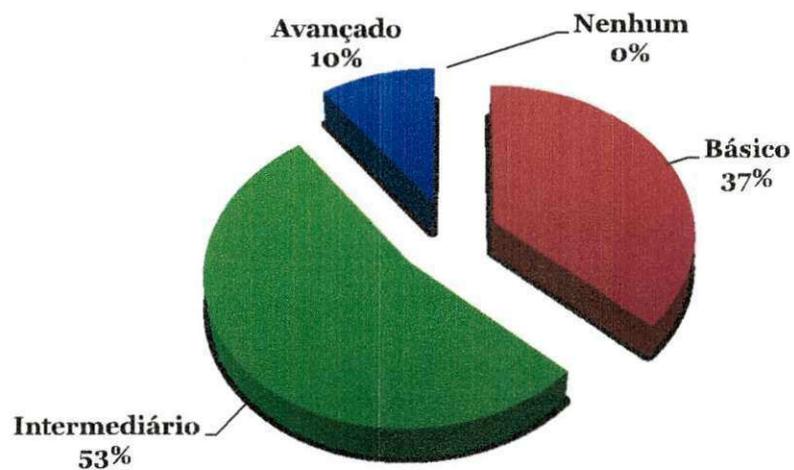
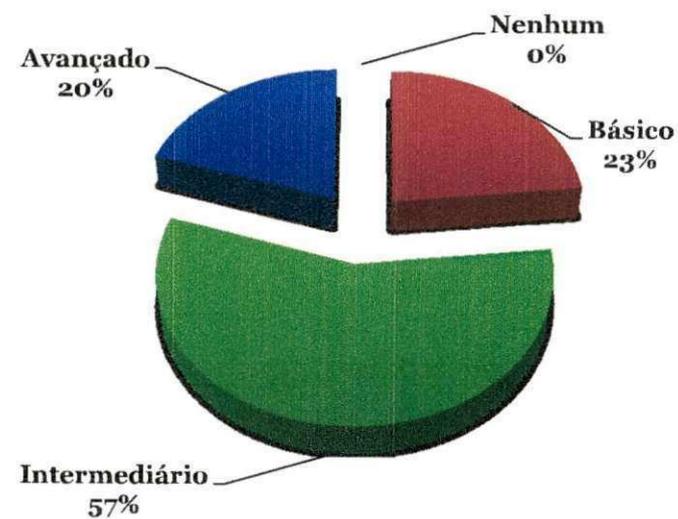


Figura 40 – Distribuição numérica da habilidade de compreensão da Língua Inglesa



Apêndice H Síntese dos Resultados da Sondagem do Perfil do Usuário

Neste apêndice estão sintetizadas no Quadro 9 as respostas dos questionário para o delineamento do perfil de cada um dos voluntários que participaram das sessões de teste de usabilidade conduzidas nesta pesquisa.

Quadro 9 – Síntese dos resultados da sondagem do perfil do usuário

Questão	Resposta	Total
Qual o seu grau de Instrução?	Ensino médio incompleto	0
	Ensino médio completo	0
	Ensino superior incompleto	18
	Ensino superior completo	1
	Pós-graduação incompleto	5
	Pós-graduação completo	6
Qual o seu Sexo?	Masculino	20
	Feminino	10
Você usa óculos ou lentes de contato?	Sim	18
	Não	12
Em que faixa etária você se encaixa?	Menor de idade	0
	18-23	20
	24-29	8
	30-35	2
	>35	0
Você tem experiência prévia no uso de computadores?	Sim	30
	Não	0
Há quanto tempo você usa computadores?	Menos de três meses	0
	De três meses a um ano	0
	Mais de um ano	30
	Não se aplica	0
Com que frequência você usa computadores?	Diariamente	30
	A cada semana	0
	A cada mês	0
	Ocasionalmente	0
	Não se aplica	0
Que sistema operacional você utiliza com mais frequência?	Windows	22
	Linux	8
	Outro: OSX	2
	Não sabe responder	0

	Não se aplica	0
Qual o seu nível de conhecimento de informática?	Básico	3
	Intermediário	2
	Avançado	5
	Programação	20
	Não se aplica	0
Você já usou algum aplicativo para a composição de apresentações multimídia (PowerPoint, Keynote, <i>IMPRESS</i>)?	Sim	30
	Não	0
	Não se aplica	0
Com que frequência você utiliza o seu aplicativo de composição de apresentações multimídia?	Diariamente	0
	A cada semana	3
	A cada mês	8
	Ocasionalmente	19
	Não se aplica	0
Você já usou a suíte de aplicativos do <i>OpenOffice.org</i> ?	Sim	25
	Não	5
	Não se aplica	0
Você já usou o aplicativo <i>IMPRESS</i> ?	Sim	7
	Não	18
	Não se aplica	5
Com que frequência você utiliza o aplicativo <i>IMPRESS</i> ?	Diariamente	0
	A cada semana	0
	A cada mês	1
	Ocasionalmente	6
	Não se aplica	23
Qual o seu nível de habilidade com a língua inglesa?		
Lê	Nenhum	0
	Básico	5
	Intermediário	14
	Avançado	11
Escreve	Nenhum	0
	Básico	10
	Intermediário	16
	Avançado	4
Fala	Nenhum	0
	Básico	11
	Intermediário	16
	Avançado	3
Compreende	Nenhum	0
	Básico	7
	Intermediário	17
	Avançado	6

Apêndice I Dados Coletados para a Mensuração de Desempenho dos Usuários

Neste apêndice estão coletados os dados obtidos da condução das sessões de teste. Os dados obtidos consistem na contagem de eventos ocorridos durante a execução de cada tarefa por parte de cada usuário. Foram observadas a quantidade de acessos a opções incorretas, a quantidade de execuções de ações incorretas, a quantidade de erros que foram repetidos e a quantidade de acessos ao menu de ajuda.

Tarefa	Indicador	1	2	3	4	5	6	7	8	9	10
Tarefa 00	Tempo de Execução	0:01:14	0:02:03	0:00:51	0:00:52	0:00:40	0:01:03	0:01:13	0:00:45	0:01:02	0:00:49
	Nº de Opções Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	1	0	0	0	0	0	0	0	0
Tarefa 01	Tempo de Execução	0:04:56	0:08:13	0:03:23	0:03:28	0:02:42	0:04:11	0:04:52	0:02:59	0:04:09	0:03:18
	Nº de Opções Incorretas	2	4	1	2	0	2	2	0	2	0
	Nº de Ações Incorretas	1	0	0	1	0	0	0	1	0	0
	Nº de Erros Repetidos	0	1	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Tarefa 02	Tempo de Execução	0:04:27	0:07:24	0:03:02	0:03:07	0:02:26	0:03:46	0:04:23	0:02:41	0:03:44	0:02:58
	Nº de Opções Incorretas	1	2	0	0	0	0	2	0	1	0
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Tarefa 03	Tempo de Execução	0:04:05	0:06:47	0:02:47	0:02:51	0:02:13	0:03:27	0:04:01	0:02:28	0:03:26	0:02:43
	Nº de Opções Incorretas	0	1	0	0	0	1	1	0	1	0
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0

Tarefa 04	Tempo de Execução	0:05:26	0:09:03	0:03:43	0:03:48	0:02:58	0:04:36	0:05:21	0:03:17	0:04:34	0:03:37
	Nº de Opções Incorretas	0	2	0	1	0	1	1	0	1	0
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Tarefa 05	Tempo de Execução	0:17:17	0:28:47	0:11:49	0:12:07	0:09:26	0:14:37	0:17:02	0:10:28	0:14:33	0:11:32
	Nº de Opções Incorretas	4	8	3	3	2	3	4	2	4	3
	Nº de Ações Incorretas	1	3	0	0	0	1	1	0	0	0
	Nº de Erros Repetidos	0	2	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	3	0	0	0	0	1	0	0	0
Tarefa 06	Tempo de Execução	0:05:04	0:08:26	0:03:28	0:03:33	0:02:46	0:04:17	0:04:59	0:03:04	0:04:16	0:03:23
	Nº de Opções Incorretas	0	1	0	0	0	1	1	0	1	0
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Tarefa 07	Tempo de Execução	0:03:42	0:06:10	0:02:32	0:02:36	0:02:01	0:03:08	0:03:39	0:02:15	0:03:07	0:02:28
	Nº de Opções Incorretas	0	1	0	0	0	0	0	0	0	0
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Tarefa 08	Tempo de Execução	0:03:13	0:05:21	0:02:12	0:02:15	0:01:45	0:02:43	0:03:10	0:01:57	0:02:42	0:02:08
	Nº de Opções Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Total	Tempo de Execução	0:49:24	1:22:14	0:33:45	0:34:36	0:26:57	0:41:47	0:48:40	0:29:54	0:41:34	0:32:56
	Nº de Opções Incorretas	7	19	4	6	2	8	11	2	10	3
	Nº de Ações Incorretas	1	3	0	0	0	1	1	0	0	0
	Nº de Erros Repetidos	0	3	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	4	0	0	0	0	1	0	0	0

Tarefa	Indicador	11	12	13	14	15	16	17	18	19	20
Tarefa 00	Tempo de Execução	0:01:00	0:00:57	0:01:12	0:01:12	0:01:29	0:01:25	0:01:09	0:00:47	0:01:09	0:01:07
	Nº de Opções Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Tarefa 01	Tempo de Execução	0:04:01	0:03:48	0:04:48	0:04:49	0:05:56	0:05:39	0:04:36	0:03:08	0:05:00	0:04:55
	Nº de Opções Incorretas	2	0	2	1	3	2	1	0	2	2
	Nº de Ações Incorretas	0	0	1	1	0	1	0	0	0	1
	Nº de Erros Repetidos	0	0	0	0	1	0	0	0	0	0
	Nº de Consultas à Ajuda	1	1	0	0	0	0	0	0	0	0
Tarefa 02	Tempo de Execução	0:03:36	0:03:25	0:04:19	0:04:20	0:05:20	0:05:05	0:04:08	0:02:49	0:04:33	0:04:11
	Nº de Opções Incorretas	0	0	1	2	2	0	0	1	2	1
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Tarefa 03	Tempo de Execução	0:03:18	0:03:08	0:03:57	0:03:59	0:04:54	0:04:40	0:03:47	0:02:35	0:04:01	0:03:47
	Nº de Opções Incorretas	1	0	1	1	1	1	0	0	1	1
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Tarefa 04	Tempo de Execução	0:04:25	0:04:11	0:05:16	0:05:18	0:06:32	0:06:13	0:05:03	0:03:26	0:05:31	0:05:23
	Nº de Opções Incorretas	1	1	1	1	1	1	1	0	1	1
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0

Tarefa 05	Tempo de Execução	0:14:02	0:13:19	0:16:47	0:16:52	0:20:46	0:19:46	0:16:05	0:10:56	0:17:05	0:16:41
	Nº de Opções Incorretas	4	3	4	3	7	6	5	2	4	4
	Nº de Ações Incorretas	0	0	1	1	2	1	0	0	1	1
	Nº de Erros Repetidos	0	0	0	0	1	1	0	0	1	0
	Nº de Consultas à Ajuda	0	0	1	0	1	0	0	0	1	1
Tarefa 06	Tempo de Execução	0:04:07	0:03:54	0:04:55	0:04:56	0:06:05	0:05:47	0:04:42	0:03:12	0:01:09	0:04:57
	Nº de Opções Incorretas	1	0	1	1	1	0	0	0	1	1
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Tarefa 07	Tempo de Execução	0:03:00	0:02:51	0:03:36	0:03:37	0:04:27	0:04:14	0:03:27	0:02:21	0:03:49	0:03:30
	Nº de Opções Incorretas	0	0	0	0	1	1	0	0	0	0
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Tarefa 08	Tempo de Execução	0:02:36	0:02:28	0:03:07	0:03:08	0:03:51	0:03:40	0:02:59	0:02:02	0:03:20	0:03:00
	Nº de Opções Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Total	Tempo de Execução	0:40:05	0:38:02	0:47:57	0:48:11	0:59:20	0:56:29	0:45:56	0:31:15	0:49:53	0:47:44
	Nº de Opções Incorretas	9	4	10	9	16	11	7	3	11	10
	Nº de Ações Incorretas	0	0	1	1	2	1	0	0	1	1
	Nº de Erros Repetidos	0	0	0	0	2	1	0	0	1	0
	Nº de Consultas à Ajuda	1	1	1	0	1	0	0	0	1	1

Tarefa	Indicador	21	22	23	24	25	26	27	28	29	30
Tarefa 00	Tempo de Execução	0:01:02	0:01:00	0:01:03	0:00:49	0:01:29	0:01:09	0:00:40	0:01:09	0:01:14	0:01:09
	Nº de Opções Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Tarefa 01	Tempo de Execução	0:03:08	0:04:01	0:04:11	0:03:18	0:05:56	0:04:36	0:02:42	0:05:00	0:04:56	0:04:36
	Nº de Opções Incorretas	2	2	2	0	3	1	0	2	2	1
	Nº de Ações Incorretas	0	0	0	0	1	0	1	1	0	0
	Nº de Erros Repetidos	0	0	0	0	1	0	0	0	0	0
	Nº de Consultas à Ajuda	0	1	0	0	0	0	0	0	0	0
Tarefa 02	Tempo de Execução	0:03:17	0:03:36	0:03:46	0:02:58	0:05:20	0:04:08	0:02:26	0:04:33	0:04:27	0:04:08
	Nº de Opções Incorretas	0	0	0	0	2	0	0	2	1	0
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Tarefa 03	Tempo de Execução	0:02:41	0:03:18	0:03:27	0:02:43	0:04:54	0:03:47	0:02:13	0:04:01	0:04:05	0:03:47
	Nº de Opções Incorretas	0	1	1	0	1	0	0	1	0	0
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Tarefa 04	Tempo de Execução	0:03:58	0:04:25	0:04:36	0:03:37	0:06:32	0:05:03	0:02:58	0:05:31	0:05:26	0:05:03
	Nº de Opções Incorretas	1	1	1	0	1	1	0	1	0	1
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0

Tarefa 05	Tempo de Execução	0:12:02	0:17:52	0:15:14	0:09:44	0:19:17	0:19:07	0:06:46	0:17:21	0:19:46	0:18:55
	Nº de Opções Incorretas	3	3	3	1	4	3	1	3	3	2
	Nº de Ações Incorretas	0	0	1	0	1	1	0	0	1	2
	Nº de Erros Repetidos	0	0	0	0	1	0	0	0	0	1
	Nº de Consultas à Ajuda	0	0	0	0	1	0	0	0	0	1
Tarefa 06	Tempo de Execução	0:03:38	0:03:54	0:04:16	0:03:04	0:04:59	0:04:42	0:02:46	0:04:57	0:05:04	0:04:55
	Nº de Opções Incorretas	0	0	1	0	1	0	0	1	0	1
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Tarefa 07	Tempo de Execução	0:02:31	0:02:51	0:03:07	0:02:15	0:03:39	0:03:27	0:02:01	0:03:30	0:03:42	0:03:36
	Nº de Opções Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Tarefa 08	Tempo de Execução	0:02:21	0:02:28	0:02:42	0:01:57	0:03:10	0:02:59	0:01:45	0:03:00	0:03:13	0:03:07
	Nº de Opções Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Ações Incorretas	0	0	0	0	0	0	0	0	0	0
	Nº de Erros Repetidos	0	0	0	0	0	0	0	0	0	0
	Nº de Consultas à Ajuda	0	0	0	0	0	0	0	0	0	0
Total	Tempo de Execução	0:34:16	0:40:34	0:39:13	0:28:10	0:51:37	0:45:32	0:22:16	0:45:32	0:48:10	0:45:40
	Nº de Opções Incorretas	6	7	8	1	12	5	1	10	6	5
	Nº de Ações Incorretas	0	0	1	0	1	1	0	0	1	2
	Nº de Erros Repetidos	0	0	0	0	2	0	0	0	0	1
	Nº de Consultas à Ajuda	0	1	0	0	1	0	0	0	0	1

Apêndice J Roteiro das Tarefas de Teste

Este apêndice exibe os slides que foram usados no aplicativo IMPRESS como roteiro de tarefas de teste seguidos pelos usuários que participaram das sessões de avaliação da usabilidade conduzidas nesta pesquisa.

Roteiro de Tarefas – Versão do *Usuário*

Roteiro:

Você deverá fazer uma apresentação de slides sobre o seu curso de graduação em um evento no qual se impõe dez minutos como limite de tempo para cada apresentação. Deste modo, você deverá elaborar uma apresentação bastante sucinta, utilizando o programa *IMPRESS*, da suíte de programas de escritório *LibreOffice 3*.

Tempo total Estimado das tarefas:

45 minutos

Observações:

- Sinta-se à vontade para consultar a ajuda do programa (menu **Help**) quando julgar necessário;
- Caso encontre alguma dificuldade que não comprometa a realização da tarefa, não se preocupe e prossiga;

Tarefa 1 – Criar slide inicial I

Roteiro:

Nesta tarefa, você irá criar o slide inicial de sua apresentação, o qual deverá conter, pelo menos, a sigla da Universidade Federal de Campina Grande, o nome do seu curso e seu nome completo.

Tempo Estimado:

5 minutos

Instruções:

1. Crie um novo slide;
2. Clique com o botão direito do mouse (no canto inferior direito do mousepad, com dois dedos sobre qualquer lugar do mousepad ou pressione o botão command ⌘) sobre o slide atualmente selecionado, do lado esquerdo da tela;
3. Selecione **New Slide**;
4. Clique do lado esquerdo da tela para navegar entre os slides desta apresentação;

2

Tarefa 1 – Criar slide inicial II

Instruções:

5. Adicione um título a sua apresentação;
6. Insira mais detalhes sobre o tema de sua apresentação;
7. Insira as informações que você achar conveniente;
8. Insira suas informações profissionais e de contato.

3

Tarefa 2 – Finalizar slide inicial

Roteiro:

Nesta tarefa, você irá finalizar o primeiro slide de sua apresentação, incluindo as logomarcas da UFCG e de seu curso.

Tempo Estimado:

5 minutos

Instruções:

1. Adicione ao slide inicial as logomarcas da UFCG e de seu curso;
2. Na barra de menus principal da aplicação, selecione **Insert > Picture > From File...**;
3. Encontre as imagens no *desktop* do sistema;
4. Ajuste o tamanho das imagens; e
5. Ajuste a posição das imagens.

4

Tarefa 3 – Criar slide com histórico

Roteiro:

Você decide que é pertinente iniciar a apresentação com um breve histórico de seu curso, introduzido no Brasil na década de 60 e criado na UFCG (na época, Campus II da UFPB) em 13 de maio de 1976.

Tempo Estimado:

5 minutos

Instruções:

1. Crie um novo slide;
2. Modifique o título do slide para o nome de seu curso seguido de um hífen e do nome **Histórico**;
3. Insira os dados históricos relativos ao seu curso dos quais você tem conhecimento.

5

Tarefa 6 – Criar slide opinativo

Roteiro:

Nesta tarefa, você deve criar um slide contendo a sua opinião acerca de seu curso.

Tempo Estimado:

5 minutos

Instruções:

1. Crie um novo slide;
2. Altere o título do slide;
3. Adicione à sua apresentação a imagem representativa de seu curso;
4. Escreva características que sejam apropriadas ao seu curso;
5. Escreva o que ainda pode melhorar em seu curso.

8

Tarefa 7 – Criar slide de agradecimento

Roteiro:

Nesta tarefa, você irá agradecer pela oportunidade de participar deste evento.

Tempo Estimado:

5 minutos

Instruções:

1. Crie um novo slide;
2. Modifique o corpo do slide para agradecer a sua oportunidade;
3. Modifique o tamanho da fonte para que fique claramente legível.

9

Tarefa 8 – Simular a apresentação

Roteiro:

Nesta tarefa, você irá visualizar e revisar a sua apresentação.

Tempo Estimado:

10 minutos

Instruções:

1. Inicie a apresentação de slides;
2. Revise os passos de sua apresentação;
3. Simule, em sua mente, a apresentação de seu trabalho.