

CENTRO UNIVERSITÁRIO DA FEI

IGOR RADAMES RODRIGUES XAVIER

**ANÁLISE MULTIVARIADA DE IMAGENS FRONTAIS DE FACE NORMALIZADAS
ESPACIALMENTE**

São Bernardo do Campo

2015

IGOR RADAMES RODRIGUES XAVIER

**ANÁLISE MULTIVARIADA DE IMAGENS FRONTAIS DE FACE NORMALIZADAS
ESPACIALMENTE**

Dissertação de Mestrado apresentada ao Centro
Universitário da FEI para obtenção do título de Mes-
tre em Engenharia Elétrica, orientada pelo Prof. Dr.
Carlos Eduardo Thomaz.

São Bernardo do Campo

2015

Xavier, Igor Radames Rodrigues.

Análise multivariada de imagens frontais de face normalizadas espacialmente / Igor Radames Rodrigues Xavier. São Bernardo do Campo, 2015.

83 f. : il.

Dissertação - Centro Universitário da FEI.

Orientador: Prof. Dr. Carlos Eduardo Thomaz.

1. Deformações livres de forma. 2. Registro de imagens de face.
3. Análise estatística multivariada. 4. Composição de faces. I. Thomaz, Carlos Eduardo, orient. II. Título.

CDU 519.2

Aluno: Igor Radames Rodrigues Xavier

Matrícula: 113201-8

Título do Trabalho: Análise multivariada de imagens frontais de face normalizadas espacialmente.

Área de Concentração: Processamento de Sinais

Orientador: Prof. Dr. Carlos Eduardo Thomaz

Data da realização da defesa: 23/06/2015

ORIGINAL ASSINADA

A Banca Examinadora abaixo-assinada atribuiu ao aluno o seguinte:

APROVADO

REPROVADO

São Bernardo do Campo, 23 de Junho de 2015.

MEMBROS DA BANCA EXAMINADORA

Prof. Dr. Carlos Eduardo Thomaz

Ass.: _____

Prof. Dr. Nelson Delfino D'Ávila Mascarenhas

Ass.: _____

Prof. Dr. Paulo Sérgio Silva Rodrigues

Ass.: _____

VERSÃO FINAL DA DISSERTAÇÃO

**ENDOSSO DO ORIENTADOR APÓS A INCLUSÃO DAS
RECOMENDAÇÕES DA BANCA EXAMINADORA**

Aprovação do Coordenador do Programa de Pós-graduação

Prof. Dr. Carlos Eduardo Thomaz

Dedico este trabalho a minha amada noiva Simone, a minha querida mãe Elzia e ao meu saudoso pai Manoel.

AGRADECIMENTOS

Ao meu orientador Prof o Dr. Carlos Eduardo Thomaz pela paciência, generosidade, dedicação e orientação na elaboração de artigos, experimentos e desta dissertação.

A todos os meus familiares, em especial minha mãe Elzia, meus irmãos Anderson, Eloane, Eloize e ao meu falecido pai Manoel que não pôde me acompanhar neste momento.

Ao Dr. Dirk-Jan Kroon da Universidade de Twente disponibilizou publicamente uma versão de código em Matlab do algoritmo de registro de imagens utilizado neste trabalho.

À Elemidia, empresa onde trabalho, por permitir que eu dedicasse tempo a esta pesquisa, em especial ao meu chefe Rodrigo Cadena.

A todos os meus colegas e amigos que sempre me incentivaram e torceram pelo meu sucesso.

E ao programa PROSUP/CAPES pelo suporte financeiro.

“There’s no such thing as a free lunch. ”

Milton Friedman

RESUMO

No contexto de reconhecimento automático ou percepção cognitiva de faces, uma etapa importante é o alinhamento das imagens de face de entrada a um atlas de referência para posterior extração multivariada de características. Para alinhar duas imagens de face, pode-se utilizar transformações rígidas e afins que corrigem variações de rotação e escala, porém deformações devido a expressões faciais, por exemplo, podem ser tanto globais quanto locais e de natureza não-linear ou não-rígida. Portanto, em situações desse tipo precisa-se de um método de alinhamento que possa lidar com tais variações. Nas últimas décadas, desenvolveu-se uma técnica de deformação de objetos sólidos, denominada em inglês de *Free-Form Deformation*, que foi adaptada primeiramente para normalização espacial de imagens médicas e mais recentemente para imagens frontais de face. Paralelamente, técnicas estatísticas de análise multivariada têm sido aplicadas ao contexto de imagens de face, tais como Análise de Componentes Principais (PCA, *Principal Components Analysis*) e Análise de Discriminantes Lineares (LDA, *Linear Discriminant Analysis*), para modelagem, reconstrução e interpretação desses padrões. Este estudo teve como objetivo realizar uma análise multivariada de imagens frontais de face normalizadas espacialmente com relação a um atlas de referência. Os resultados mostram que esta abordagem produziu imagens artificiais de faces humanas com realismo promissor, tanto nas direções de maior variação obtidas pelo PCA quanto nas direções mais discriminantes obtidas pelo LDA.

Palavras-chave: deformações livres de forma. Registro de imagens de face. Análise estatística multivariada. Composição de faces.

ABSTRACT

In the context of face recognition and perception, an important step is the alignment of the input face images to a reference atlas for further multivariate feature extraction. To align two face images, we can use rigid transformations to minimize scale and rotation variations, but deformation due to facial expressions can be both global and local, or may be non-linear or non-rigid. Therefore, in such situations we need an alignment method that can deal with these variations. In the past decades a technique of deformation of solid objects known as Free-Form Deformation has been developed and first adapted for spatial normalization of medical images and more recently to frontal face images. At the same time, methods of multivariate statistics have been applied to the context of face images, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), for modeling and reconstructing these patterns. The main goal of this work has been to perform a multivariate statistical analysis of frontal face images using non-rigid deformation fields previously aligned to a reference atlas. The experimental results carried out have showed that our approach has been able to generate artificial face images with a realistic human-like appearance, in both PCA and LDA most expressive and discriminant directions.

Keywords: free-form deformation. Face image registration. Multivariate statistical analysis. Facial composite.

LISTA DE FIGURAS

Figura 1	Exemplo de FFD aplicada a um modelo de sólido. Fonte: Hahmann et al. adaptado de Hahmann et al. (2012).	18
Figura 2	Exemplo de face marcada com os 46 pontos fiduciais do trabalho de Tenorio e Thomaz (2011).	26
Figura 3	Máscara com 61 pontos posicionada sobre uma imagem de face usando uma estimativa inicial baseada na posição do centros dos olhos e ponta do queixo. Pontos marcados com circulo vermelho são os pontos principais. Pontos azuis são pontos de controle das curvas interpoladoras e linhas azuis são as curvas interpoladoras baseadas em <i>B-Splines</i>	28
Figura 4	Posicionamento final dos 61 pontos fiduciais requeridos para interpolação usando <i>B-Splines</i>	29
Figura 5	Conjunto de 284 pontos fiduciais delineando a maior parte das características faciais.	30
Figura 6	Transformação rígida aplicada a uma imagem do banco de faces da FEI. . .	33
Figura 7	Transformação afim aplicada a uma imagem do banco de faces da FEI. . .	34
Figura 8	Malha de pontos de controle sobreposta a uma imagem de face. As posições dos pontos de controle são as intersecções das linhas da malha.	36
Figura 9	Resultado da aplicação do Algoritmo 1 em uma imagem de face: (a) Imagem de face original sobreposta por uma malha de pontos de controle; (b) Face de referência usada para registrar a face de entrada; (c) A mesma imagem de face deformada após a aplicação da FFD sobreposta pela malha de pontos otimizada.	39
Figura 10	Imagem média neutra calculada a partir do banco de faces da FEI.	46
Figura 11	Procedimento iterativo de criação de um atlas para registro não-rígido de imagens frontais de face.	47
Figura 12	Um subconjunto de imagens de face retiradas do banco de faces da FEI. . .	55
Figura 13	Anotações feitas nas imagens do banco de faces da FEI. À esquerda: face média neutra do banco de dados anotada com 46 pontos fiduciais. À direita: exemplo de imagem do banco de dados anotada com os mesmos 46 pontos da face média.	56

Figura 14 Exemplos de resultados da marcação semi-automática de 284 pontos fiduciais no conjunto original de faces originais do banco de faces da FEI.	57
Figura 15 Região de interesse demarcada pela envoltória convexa dos pontos fiduciais anotados nas imagens de face.	59
Figura 16 Exemplos de faces deformadas usando o algoritmo de registro FFD. A linha superior apresenta imagens de face originais, a linha do meio apresenta as mesmas imagens deformadas do conjunto 46 e a linha inferior apresenta, analogamente, as imagens do conjunto 284.	61
Figura 17 Navegação nas oito primeiras componentes principais descrevendo as características mais expressivas do conjunto de faces original.	63
Figura 18 Navegação nas oito primeiras componentes principais descrevendo as características mais expressivas do conjunto não-rigidamente registrado com 46 pontos fiduciais.	64
Figura 19 Navegação nas oito primeiras componentes principais descrevendo as características mais expressivas do conjunto não-rigidamente registrado com 284 pontos fiduciais.	65
Figura 20 Características mais discriminantes do conjunto de faces originais, da esquerda para a direita: masculino contra feminino (acima); neutra contra sorrindo (abaixo).	66
Figura 21 Características mais discriminantes do conjunto de faces registradas com 46 pontos fiduciais, da esquerda para a direita: masculino contra feminino (acima); neutra contra sorrindo (abaixo).	67
Figura 22 Características mais discriminantes do conjunto de faces com 284 pontos fiduciais, da esquerda para a direita: masculino contra feminino (acima); neutra contra sorrindo (abaixo).	67
Figura 23 Taxas de classificação da reta de separação das amostras, gênero da esquerda para a direita: 94%, 96.75% e 96.25%; Expressão Facial; 95%, 96.25% e 94.75%.	68
Figura 24 Projeção das amostras no vetor discriminante de gênero: (CM) - amostras masculinas usadas no cálculo da reta discriminante; (CF) - amostras femininas usadas no cálculo da reta discriminante; (MM) - média das amostras masculinas; (MF) - média das amostras femininas; (DI) - faces geradas artificialmente usando o método supervisionado; (EX) - faces geradas artificialmente pelo método não-supervisionado.	70

Figura 25	Projeção das amostras no vetor discriminante de expressão faciais: (CN) - amostras neutras usadas no cálculo da reta discriminante; (CS) - amostras sorrindo usadas no cálculo da reta discriminante; (MN) - média das amostras neutras; (MS) - média das amostras sorrindo; (DI) - faces geradas artificialmente usando o método supervisionado; (EX) - faces geradas artificialmente pelo método não-supervisionado.	71
Figura 26	Gráficos de probabilidade normal e histograma dos grupos de amostras de treinamento de gênero: (a) - gráficos de probabilidade normal das amostras femininas; (b) - histograma das amostras femininas; (c) - gráficos de probabilidade normal das amostras masculinas; (d) - histograma das amostras masculinas.	72
Figura 27	Gráficos de probabilidade normal e histograma dos grupos de amostras de treinamento de expressão facial: (a) - gráficos de probabilidade normal das amostras sorrindo; (b) - histograma das amostras sorrindo; (c) - gráficos de probabilidade normal das amostras neutras; (d) - histograma das amostras neutras.	73

LISTA DE TABELAS

Tabela 1	Imagens de referência usadas para registrar as 400 imagens da base original e as respectivas médias de correlação cruzadas obtidas no processo de registro.	60
----------	---	----

LISTA DE ALGORITMOS

Algoritmo 1	Algoritmo de registro de imagens. Fonte: Rueckert et al. adaptado de (RUECKERT et al., 1999)	38
-------------	---	----

SUMÁRIO

1 INTRODUÇÃO	15
1.1 Objetivo	16
1.2 Organização	16
2 REVISÃO BIBLIOGRÁFICA	17
2.1 Considerações Complementares	22
3 METODOLOGIA	24
3.1 Marcação de Pontos Fiduciais	24
3.1.1 Marcação com 46 Pontos Fiduciais	25
3.1.2 Marcação com 284 Pontos Fiduciais	27
3.2 Registro de Imagens de Face	31
3.2.1 Transformações Rígidas	33
3.2.2 Transformações Afins	34
3.2.3 Transformações Não-Rígidas baseadas em FFD	35
3.3 Métricas de Similaridade	40
3.3.1 Métricas baseadas na Intensidade dos Pixels	40
3.3.1.1 <i>Correlação Cruzada (CC)</i>	40
3.3.1.2 <i>Distância Euclidiana (DE)</i>	41
3.3.2 Métricas Baseadas em Entropia	41
3.3.2.1 <i>Entropia Conjunta (EC)</i>	41
3.3.2.2 <i>Informação Mútua (IM)</i>	42
3.3.2.3 <i>Informação Mútua Normalizada (IMN)</i>	43
3.3.3 Métrica de Distância entre Pontos Fiduciais	43
3.4 Construção de um Atlas	44
3.5 Características Mais Expressivas	48
3.6 Características Mais Discriminantes	50
4 EXPERIMENTOS E RESULTADOS	54
4.1 Banco de Faces	54
4.2 Marcação de 46 Pontos Fiduciais	55
4.3 Marcação de 284 Pontos Fiduciais	56

4.4	Construção do Atlas de Referência	57
4.5	Geração de Novos Conjuntos de Imagens de Face	61
4.6	Resultados Mais Expressivos	62
4.7	Resultados Mais Discriminantes	66
5	CONCLUSÃO	74
	REFERÊNCIAS	76
	APÊNDICE A – PCA COM NÚMERO PEQUENO DE AMOSTRAS	81

1 INTRODUÇÃO

Imagens de face têm sido aplicadas em problemas de Engenharia e Computação principalmente como sinais biométricos para o reconhecimento automático de pessoas. Entretanto, tais sinais são também úteis para investigações científicas relacionadas à compreensão da percepção facial humana.

Em ambos contextos, reconhecimento automático e percepção humana, uma etapa importante é o alinhamento das imagens de face de entrada a um atlas de referência para posterior extração de características, de acordo com uma base de conhecimento pré-existente (AHONEN; HADID; PIETIKÄINEN, 2004; KIRBY; SIROVICH, 1990). Em outras palavras, antes de comparar duas amostras de faces distintas, deve-se normalizar espacialmente a localização das características mais relevantes de uma amostra em relação às mesmas características da outra.

Em aplicações na área da Psicologia, por exemplo, alguns trabalhos visam o estudo de julgamentos e inferências sociais baseados em estímulos visuais por imagens de face (TODOROV et al., 2008; TODOROV; OOSTERHOF, 2011). Estes estudos assumem que dimensões de um espaço multidimensional, comumente chamado de espaço de faces, definem características fisionômicas relevantes para codificação de informações faciais interpretadas em nosso sistema cognitivo como desvios a partir de uma face padrão, ou protótipo, localizada na origem ou centro desse espaço de faces.

De fato, mesmo duas imagens de face da mesma pessoa podem apresentar variações de pose, escala e expressão, tornando a tarefa de codificação dessa informação muito mais problemática do que seria sem o alinhamento correto de suas características. Em alguns problemas específicos, esta etapa de pré-processamento pode ser ainda mais determinante. Por exemplo, na tentativa de interpretação de expressões faciais, o alinhamento precisa ser realizado entre imagens de face da mesma pessoa, pois, geralmente, o sinal a ser analisado é uma sequência de imagens, onde a expressão é uma variação espacial e temporal da mesma face (KOELSTRA; PANTIC, 2008).

No caso de imagens de face da mesma pessoa, um alinhamento rígido pode corrigir a variação de rotação e um alinhamento afim pode lidar com a variação de escala. Porém, quando trabalham-se com imagens de pessoas distintas ou quando lidam-se com expressões faciais, têm-se variações de natureza não-rígida. Portanto, precisa-se de um método de alinhamento que dê suporte às variações de naturezas global e local deste padrão.

1.1 Objetivo

O objetivo principal deste trabalho é realizar uma análise multivariada de imagens frontais de face alinhadas não-rigidamente a um atlas de referência. Pretende-se comparar esta análise com aquela realizada em imagens de face previamente normalizadas de maneira rígida, gerando amostras artificiais com realismo comparável às imagens já existentes a partir de técnicas de estatística multivariada. Como resultados complementares deste estudo, tem-se o objetivo de investigar os efeitos do uso de diferentes configurações de pontos fiduciais utilizados para o alinhamento não-rígido das faces e disponibilizar publicamente uma base de imagens de faces frontais normalizadas espacialmente com potencial aplicação em problemas de percepção cognitiva deste padrão

1.2 Organização

Esta dissertação está dividida em cinco capítulos. No capítulo 2, apresenta-se uma breve revisão bibliográfica da literatura e das técnicas e ferramentas computacionais utilizadas. No capítulo 3, descreve-se o método de marcação de pontos fiduciais em faces, além dos métodos de normalização espacial de imagens, construção de um atlas de referência e estatística multivariada empregados. No capítulo 4, os experimentos e resultados deste estudo são detalhados e, por fim, no capítulo 5, apresenta-se a conclusão e os possíveis trabalhos futuros.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo descreve sucintamente os principais trabalhos relacionados aos assuntos estudados nesta dissertação de mestrado.

Em meados da década de 80, o trabalho de Sederberg e Parry (1986) introduziu uma técnica computacional de deformação de objetos sólidos. Esta técnica de deformação livre de forma foi denominada em inglês de *Free-Form Deformation* (FFD) e teve como motivação principal criar uma ferramenta que desse a mesma flexibilidade de um artista escultor ao modelador computacional para deformação de objetos. Já neste estudo de 1986, os autores sugeriram que a técnica poderia ser aplicada não somente a modelos de sólidos como também de superfícies ou polígonos.

Uma boa analogia física para a FFD é considerar um paralelepípedo de plástico flexível e transparente no qual podemos embutir um ou mais objetos. Imagina-se que os objetos embutidos são tão flexíveis quanto o paralelepípedo que os envolve. Logo, ao deformarmos o paralelepípedo estaremos deformando os objetos embutidos nele. Um exemplo pode ser visto na Figura 1 adaptada de Hahmann et al. (2012).

Matematicamente, a FFD pode ser definida em termos do produto tensorial trivariado do Polinômio de Bernstein:

$$\mathbf{X}_{ffd} = \sum_{i=0}^l \binom{l}{i} (1-s)^{l-i} s^i \left[\sum_{j=0}^m \binom{m}{j} (1-t)^{m-j} t^j \left[\sum_{k=0}^n \binom{n}{k} (1-u)^{n-k} u^k \mathbf{P}_{ijk} \right] \right]. \quad (2.1)$$

Esta equação implica que os objetos embutidos no paralelepípedo imaginário têm um sistema de coordenadas local onde cada ponto \mathbf{X} tem coordenadas (s, t, u) e que as novas coordenadas dos pontos são afetadas pelos pontos de controle \mathbf{P}_{ijk} , os quais formam a grade que compõe o paralelepípedo imaginário. No artigo supracitado os autores indicaram que a FFD pode ser formulada em termos de outras bases polinomiais, tais como produtos tensoriais de B-Splines e produtos não-tensoriais de Polinômios de Bernstein.

Aproximadamente 10 anos mais tarde, em Lee, Chwa e Shin (1995) e Lee et al. (1996) introduziu-se a deformação livre de forma em múltiplos níveis, comumente denominada em inglês de *Multilevel Free-Form Deformation* (MFFD) como uma extensão para FFD (SEDERBERG; PARRY, 1986). Com o intuito de derivar funções de deformação a partir de restrições posicionais impostas por características da imagem (pontos fiduciais), os autores usaram o pro-

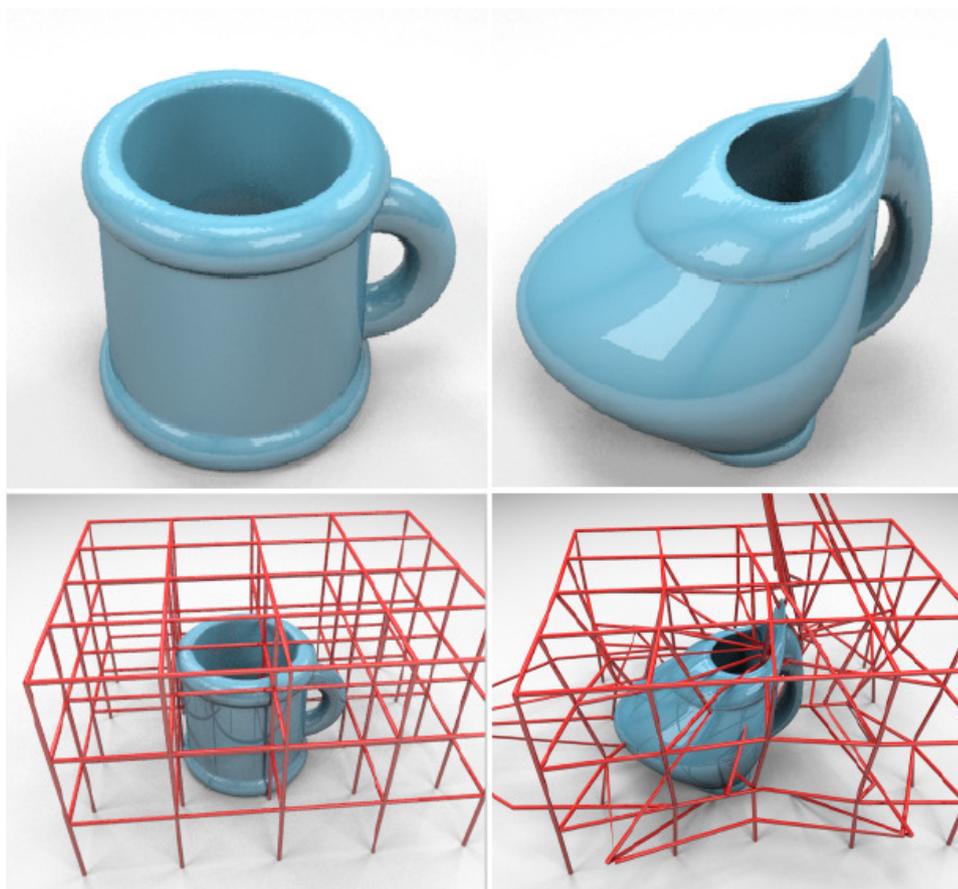


Figura 1 – Exemplo de FFD aplicada a um modelo de sólido. Fonte: Hahmann et al. adaptado de Hahmann et al. (2012).

duto tensorial bivariado de B-Splines cúbicas como bases polinomiais para a FFD. Também foi proposta uma técnica para manipulação direta desta FFD. Uma hierarquia de malhas de pontos de controle é aplicada a cada iteração da MFFD, onde cada malha tem o espaçamento entre os pontos de controle menor que a malha anterior. A função de deformação final é uma composição das funções de deformação estimadas em cada iteração, objetivando compor deformações que não poderiam ser derivadas por FFD devido aos limites impostos pelas condições de continuidade desta última.

Paralelamente, Cootes et al. (1995) desenvolveram modelos de distribuição de pontos, do inglês *Point Distribution Models*, a partir de pontos fiduciais manualmente anotados nas bordas de conjuntos de objetos pertencentes a uma mesma classe, tais como, imagens de resistores, exames de ecocardiograma e mãos. Em tais modelos, visou-se capturar a forma e variação típicas que os objetos modelados poderiam apresentar. Os autores aplicaram a análise de componentes principais nas coordenadas dos pontos anotados do conjunto de imagens de treinamento, obtendo-se, assim, um conjunto de parâmetros estatísticos para controlar a variação do formato da classe de objetos e também garantir a especificidade destes. Mostrou-se também que é possível gerar artificialmente novos exemplos de objetos somando-se ao formato

médio as k primeiras componentes de variação ponderadas pelos seus desvios correspondentes. Adicionalmente, mostrou-se uma técnica iterativa de busca de imagens a partir dos modelos de distribuição de pontos, denominada *Active Shape Models*, ajustando os parâmetros do modelo em cada iteração a fim de obter uma correspondência ótima entre a forma gerada pelo modelo e a forma da imagem de exemplo. Os autores argumentaram a importância de anotar corretamente o conjunto de imagens de treinamento para que o modelo apenas capturasse a variação no formato dos objetos e não incorporasse a variação de ruídos.

Ainda no contexto de funções de deformação, Lee, Wolberg e Shin (1997) generalizam o algoritmo MFFD, proposto em Lee, Chwa e Shin (1995), Lee et al. (1996), em termos da interpolação de dados n -dimensionais. Para demonstrar a eficiência desta técnica, foram apresentadas três aplicações: (1) reconstrução de imagens de face a partir de amostras de intensidades de uma imagem como um problema de interpolação de uma dimensão; (2) deformação de uma imagem de face a partir da movimentação de características (pontos fiduciais) como um problema em duas dimensões; e (3) reconstrução de volumes em três dimensões a partir de uma amostra de dados de intensidade de uma imagem de face em duas dimensões como um problema em três dimensões.

Baseando-se nas técnicas de interpolação FFD desenvolvidas anteriormente, Rueckert et al. (1999) desenvolveram uma técnica de registro não-rígido de imagens utilizando, além de alinhamento afim, funções de deformação derivadas da FFD baseada em B-splines. Neste trabalho, a técnica foi aplicada para registrar imagens de exames de ressonância magnética em mamas. O objetivo era melhorar o alinhamento das várias imagens do exame, antes dificultado pelo movimento dos pacientes. Os autores mostraram que o alinhamento não-rígido por FFD endereçava o problema de alinhar deformações locais sem atrapalhar o alinhamento global das imagens. Os autores também propuseram um algoritmo de otimização dos parâmetros da FFD usando o método da descida do gradiente e mostraram a eficiência desta abordagem em relação a outras como *Thin-plate Splines* e *Elastic Body Registration*. Estas últimas têm seu desempenho degradado quando necessitam-se otimizar parâmetros de locais afetando a otimização de todos os outros.

No início dos anos 2000, Cootes, Edwards e Taylor (2001) estenderam o modelo estatístico do formato de objetos proposto em Cootes et al. (1995), adicionando um modelo semelhante de textura, onde procurou-se não apenas modelar a variação e a especificidade do formato como também o padrão de intensidade dos pixels no interior do objeto modelado. Neste trabalho, mostrou-se que os modelos separados de formato e textura estão estatisticamente correlacionados e, portanto, os autores fundiram os parâmetros dos modelos em um

único parâmetro. Os autores também mostraram que, variando-se os parâmetros do modelo, conforme mostrado anteriormente em Cootes et al. (1995), foi possível obter exemplos não somente de formatos, mas, também de texturas artificiais que capturavam a variação de uma base de treinamento com 400 imagens de face. Os autores aplicaram também uma técnica de otimização dos parâmetros a fim de que o modelo estatístico do formato e aparência convergisse para uma imagem de teste. O modelo foi testado tanto para imagens de face quanto de ressonância magnética do joelho. Tal método foi denominado de *Active Appearance Models* e os testes quantitativos mostraram que em 13% das imagens o modelo não convergiu satisfatoriamente, de acordo com os critérios definidos pelos autores.

Paralelamente em Tiddeman e Perrett (2001) e Tiddeman, Burt e Perrett (2001), desenvolveu-se um método para criar sequências de imagens de face visualmente realísticas a partir de modelos baseados em pontos fiduciais que demarcam características faciais. Nestes trabalhos, 179 pontos foram usados para descrever um modelo para faces femininas e outro modelo para faces masculinas. A textura é derivada calculando-se a média das intensidades de cada canal RGB provenientes de uma função de interpolação linear aplicada em cada imagem de treinamento para evitar descontinuidades. Os autores mostraram ser possível gerar artificialmente faces com certas características como envelhecimento simplesmente computando-se a diferença entre uma face fonte, usualmente da mesma classe da face a ser deformada, e uma face alvo, aquela com os atributos que desejam-se incorporar. O método proposto no artigo gerou novas representações das imagens reais através das diferenças entre modelos previamente calculados, ou seja, pode-se navegar entre modelos de faces dada uma face qualquer de um modelo origem para um modelo destino. Os autores demonstraram diversos exemplos, desde modelos baseados em imagens humanas até modelos baseados em imagens de desenhos animados e animais. Outra contribuição importante destes trabalhos foi o melhoramento no realismo da textura das faces a partir de características adicionadas por wavelets, criando faces com características humanas extremamente realísticas.

Em 2003, Gibson, Solomon e Bejarano (2003) estenderam o *Active Appearance Models* para gerar imagens de faces realísticas de forma iterativa navegando no espaço de faces usando três diferentes estratégias evolucionárias. Em cada caso, é requerido a um usuário interessado em obter uma composição de face do seu interesse escolher uma entre várias faces geradas pelo algoritmo dinamicamente. Dependendo da escolha, o algoritmo atualiza a busca no espaço de faces a fim de alcançar a correspondência correta. Esta por sua vez é uma proposta de desenvolver um sistema de retrato falado interativo, sem a necessidade de um desenhista treinado

e composto pela própria testemunha. Testes com testemunhas virtuais foram elaborados para avaliar a viabilidade da busca iterativa no espaço de faces.

Como resultado prático desses trabalhos, Gibson desenvolveu uma ferramenta completa de composição de faces a partir de um aprendizado estatístico de faces (GIBSON, 2006). O EigenFIT incorpora mais características no aprendizado tais como: cicatrizes, barbas e bigodes, defeitos faciais, entre outros para compor retratos falados de forma iterativa através de escolhas sucessivas por parte das testemunhas. Infelizmente, no entanto, as faces produzidas não têm um aspecto realístico e possuem aspecto de modelos artificiais, porém, cumprem o compromisso assumido de produzir retratos falados com as características descritas pela testemunha sem a intervenção de desenhista treinado.

Mais recentemente, alguns trabalhos exploraram as características estatísticas de baixo nível em imagens de face para gerar interpretações e reconstruções de alto nível da percepção humana, tais como gênero e expressão facial. Em Thomaz et al. (2009), aplicou-se a análise de componentes principais (PCA) para analisar de forma não supervisionada as direções de maior variância calculadas a partir do processamento de bases de dados de imagens de face 2D. Os autores navegaram nas componentes principais e reconstruíram imagens de face na direção dessas componentes a partir da face média ou protótipo. Como resultado, perceberam-se que estas componentes retêm várias características faciais não necessariamente exclusivas como variação de gênero e expressão presentes nos dados de treinamento. No mesmo trabalho, os autores também aplicaram uma análise supervisionada unindo PCA com a análise de discriminantes lineares regularizada (PCA+MLDA) para derivar hiperplanos de separação que discriminaram gênero, expressão facial e idade. Ao navegar nestes hiperplanos a partir da face média, os autores puderam mostrar que é possível partir de um protótipo e reconstruir imagens de face com as características desejadas.

Ainda neste contexto, o trabalho Thomaz e Giraldi (2010) estendeu a abordagem de interpretação e navegação no espaço de faces usando uma nova abordagem para o PCA, que reordena as componentes principais pelas mais discriminantes e não pelas que retêm maior variância. Esta reordenação possibilitou um aumento nas taxas de classificação ao usar menos componentes como características para três tipos de classificadores. Os autores também reportaram resultados superiores na reconstrução de faces artificiais usando a nova ordenação do PCA.

Na mesma época, Koelstra e Pantic (2008) aplicaram o alinhamento não-rígido de imagens baseado em uma versão adaptada das FFD para derivar vetores de deformação que capturam o movimento facial em vídeos de expressões faciais. A transformação necessária para

alinhar os quadros de um vídeo em relação ao primeiro quadro neutro gerou um conjunto de vetores de deformação, a partir dos quais os autores construíram histogramas do movimento facial. Tal informação foi utilizada como característica da expressão facial para treinamento de um sistema de classificação de expressões faciais, o qual obteve uma média de 94.3% de classificação contra a base de expressões MMI.

Em 2011, Saragih, Lucey e Cohn (2011) estenderam o conceito apresentado em Cootes et al. (1995) de ajustar um modelo deformável a um objeto, no caso, composto por uma imagem de face. Porém, ao contrário do modelo de distribuição de pontos, formulou-se um arcabouço probabilístico de possíveis localizações dos pontos fiduciais do modelo. Tal método é denominado em inglês de *Constrained Local Models*. Os autores mostraram que esta metodologia apresenta bons resultados também na presença de oclusões parciais.

Nos últimos anos, Sandbach et al. (2012) estenderam o trabalho de Koelstra e Pantic (2008) ao capturarem características do movimento facial em uma sequência de imagens de face 3D utilizando FFD. Assim como em Koelstra e Pantic (2008), as deformações encontradas foram utilizadas como características de expressões faciais. Os autores propuseram o uso de imagens 3D porque diminuem os efeitos de luminosidade encontrados em imagens 2D.

2.1 Considerações Complementares

Durante as últimas décadas, o registro de imagens têm tido um papel fundamental nas mais diversas áreas de processamento de imagens, desde sensoriamento remoto, passando por cartografia, visão computacional e processamento de imagens médicas (ZITOVA; FLUSSER, 2003). No contexto de imagens de face, o registro de imagens têm sido aplicado como etapa de pré-processamento para extração de características em reconhecimento de expressões faciais principalmente (KOELSTRA; PANTIC, 2008; SANDBACH et al., 2011; SANDBACH et al., 2012). Em conjunto com essas abordagens e levando-se em conta resultados da área de percepção facial que demonstram o fato que seres humanos têm habilidades naturais de processamento holístico de imagens de face (TODOROV et al., 2008; TODOROV; OOSTERHOF, 2011; SINHA et al., 2006), aplicou-se o registro de imagens não-rígido com o intuito de compensar artificialmente imagens de face eliminando artefatos provocados pelo desalinhamento de características internas da face.

No mesmo período, foram desenvolvidos métodos estatísticos de análise multivariada dos pontos fiduciais anotados em imagens com o intuito de derivar modelos de formato de objetos de interesse (COOTES et al., 1995) e, também, modelos de formato e textura de imagens de face (COOTES; EDWARDS; TAYLOR, 2001). Nesta dissertação, utiliza-se a abordagem estatística adotada nestes trabalhos, porém, diferentemente de Cootes, Edwards e Taylor (2001), deriva-se um único modelo de face que leva em consideração tanto a forma quanto a textura simultaneamente. Assim como em Cootes et al. (1995), Cootes, Edwards e Taylor (2001), desenvolve-se um modelo de face que pode gerar exemplos artificiais variando seus parâmetros.

Outras abordagens, como em Tiddeman e Perrett (2001), Tiddeman, Stirrat e Perrett (2005), Tiddeman, Burt e Perrett (2001), também fizeram a composição de faces e geraram novos exemplos a partir de modelos previamente computados, contudo, estes trabalhos também desenvolveram modelos separados de textura e formato além de inserir características específicas de pares de imagens somente. Nesta dissertação, também utilizam-se pontos fiduciais para gerar o modelo de face, no entanto, procura-se partir de um único modelo que engloba tanto formato quando textura para descrever as características mais expressivas e discriminantes de um conjunto de amostras de faces.

3 METODOLOGIA

A metodologia proposta neste trabalho seguiu os seguintes passos principais: marcação das imagens de face com pontos fiduciais, registro não-rígido das imagens anotadas a um atlas de referência, criando novos conjuntos de imagens deformadas, e análises estatísticas multivariadas baseadas em técnicas supervisionadas e não-supervisionadas dos conjuntos de faces gerados. Nenhuma dessas técnicas é nova e todas já foram aplicadas com sucesso para estudar em separado composições faciais (COOTES; EDWARDS; TAYLOR, 2001; COOTES; TAYLOR et al., 2004; COOTES et al., 1995; GIBSON, 2006), registro de imagens médicas (RUECKERT et al., 1999) e reconstrução e modelagem rígida de imagens de face (THOMAZ et al., 2009). Acredita-se que a aplicação conjunta destas técnicas seja uma contribuição importante desta pesquisa.

As seções seguintes descrevem os fundamentos teóricos das técnicas e ferramentas computacionais utilizadas nesta dissertação.

3.1 Marcação de Pontos Fiduciais

Para normalizar espacialmente todas as imagens de faces de interesse, necessitam-se, preliminarmente, marcar um conjunto de pontos fiduciais em cada uma das faces. Com tal conjunto de pontos, pretendem-se descrever certas regiões faciais que representam características importantes a serem normalizadas, tais como: olhos, nariz, perfil do rosto e boca.

Alguns trabalhos propuseram detectar automaticamente um conjunto de pontos fiduciais em imagens de face 2D. Os trabalhos de Vukadinovic e Pantic (2005) e Valstar et al. (2010) detectaram aproximadamente 20 pontos fiduciais em imagens do banco de dados de faces BioID (JESORSKY; KIRCHBERG; FRISCHHOLZ, 2001), onde obtiveram acurácia de detecção acima de 90%. Mais tarde Belhumeur et al. (2011) e Cao et al. (2014) superaram o desempenho dos métodos anteriores e estenderam o protocolo de comparação em bancos de dados de faces mais desafiadores, como no banco de dados LFPW (BELHUMEUR et al., 2011). Este apresenta imagens de face com maior variação de pose, escala e iluminação. Recentemente, em Sun, Wang e Tang (2013), utilizou-se um método com redes de convolução que supera os trabalhos anteriores mesmos nas condições extremas apresentadas pelas recentes bases de da-

dos. Entretanto, os autores realizaram a comparação entre métodos para detecção de apenas 5 pontos fiduciais.

Nesta dissertação, não utilizaram-se nenhum dos métodos anteriores para marcar as faces por algumas razões pontuais: nem todos os autores disponibilizam o código fonte ou executável de seus métodos de forma pública; a maioria dos métodos detecta poucos pontos fiduciais; e, principalmente, mesmo o estado da arte não supera ainda um usuário treinado na tarefa de marcar corretamente as imagens de face. Por tais motivos, utilizou-se o banco de faces da FEI, pois, este encontra-se previamente marcado com 46 pontos fiduciais (TENORIO; THOMAZ, 2011). Realizou-se também, para efeito de comparação, uma nova marcação com 284 pontos fiduciais com a ajuda de uma ferramenta de marcação semi-automática de pontos fiduciais em imagens de face (GIBSON, 2006). Nas seguintes sub-seções, são descritos os dois modelos de marcação de pontos fiduciais usados nesta dissertação.

3.1.1 Marcação com 46 Pontos Fiduciais

Primeiramente, aproveitou-se do trabalho realizado em Tenorio e Thomaz (2011), onde foram marcados 46 pontos fiduciais nas faces da base da FEI. Neste trabalho, realizou-se uma análise multilinear de imagens frontais de face baseando-se nos formatos derivados dos 46 pontos fiduciais. Do total de pontos, 19 são coincidentes com os definidos no trabalho de Vukadinovic e Pantic (2005) e 27 são oriundos do trabalho de Lanitis, Taylor e Cootes (1995). Um exemplo de face marcada com os 46 pontos fiduciais pode ser observado na Figura 2.



Figura 2 – Exemplo de face marcada com os 46 pontos fiduciais do trabalho de Tenorio e Thomaz (2011).

3.1.2 Marcação com 284 Pontos Fiduciais

Além dos pontos previamente mencionados e com o intuito de comparar diferentes configurações e quantidades de pontos fiduciais, utilizou-se a ferramenta *Face Landmarking Tool* (FLT). Esta é uma versão estendida da ferramenta *Shape Delineation Tool* (SDT) (GIBSON, 2006). A diferença entre ambas é que com a FLT marcam-se mais características do que com a SDT. Marcando-se as imagens de face com esta ferramenta, dispõe-se de um conjunto mais detalhado de pontos fiduciais e por sua vez delineiam-se mais características das faces do que com a marcação de 46 pontos.

Usando-se a ferramenta FLT não há necessidade de marcar ponto-a-ponto os 284 pontos fiduciais. Ao invés disso, para cada imagem de face, necessita-se posicionar manualmente um conjunto de 61 pontos fiduciais pré-definidos e, no final, chegam-se aos 284 pontos por meio de interpolações não-lineares de curvas polinomiais definidas para cada região da face.

Primeiramente, a ferramenta solicita que sejam marcados os seguintes pontos: centro do olho direito, centro do olho esquerdo e ponta do queixo. Estes pontos são usados pela ferramenta para realizar uma estimativa inicial da localização de uma máscara com 61 pontos fiduciais, conforme Figura 3. Essa estimativa inicial é feita com base em uma transformação a partir de uma configuração média desses pontos (representando o formato de uma face média) e usando a posição dos três pontos informados inicialmente.

Após a estimativa inicial e utilizando o recurso de clicar e arrastar, posicionam-se corretamente os pontos fiduciais principais (marcados com círculos vermelhos na Figura 3) de acordo com um modelo mostrado pela ferramenta. Estes representam pontos salientes das características faciais, por exemplo: cantos dos olhos e boca. Em seguida, posicionam-se os pontos de controle (desenhados em azul) de modo que as características faciais entre os pontos principais estejam corretamente delineadas. Em outras palavras, estes pontos controlam a forma das curvas que delineiam as características faciais. Ao moverem-se estes pontos altera-se o formato dessas curvas (desenhadas em azul) localmente. As curvas, neste caso, são interpolações entre os pontos principais usando *B-Splines*. Podem-se observar as posições finais dos 61 pontos fiduciais na Figura 4.

Após posicionar manualmente os 61 pontos descritos acima, interpolam-se automaticamente estes pontos ao longo das curvas em azul para chegar ao resultado final de 284 pontos fiduciais. Um exemplo de resultado dessa interpolação pode ser observado na Figura 5.

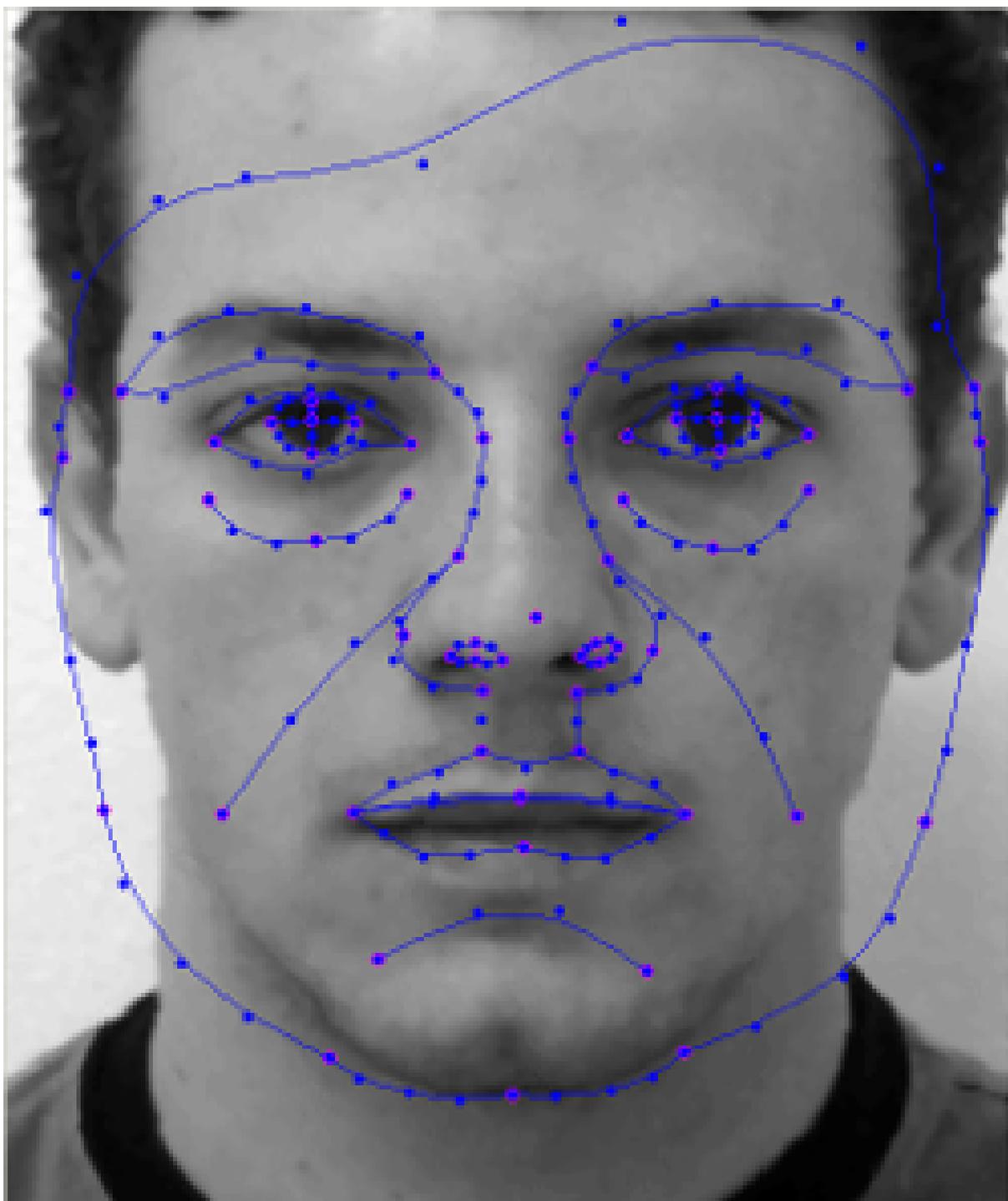


Figura 3 – Máscara com 61 pontos posicionada sobre uma imagem de face usando uma estimativa inicial baseada na posição do centros dos olhos e ponta do queixo. Pontos marcados com círculo vermelho são os pontos principais. Pontos azuis são pontos de controle das curvas interpoladoras e linhas azuis são as curvas interpoladoras baseadas em *B-Splines*.

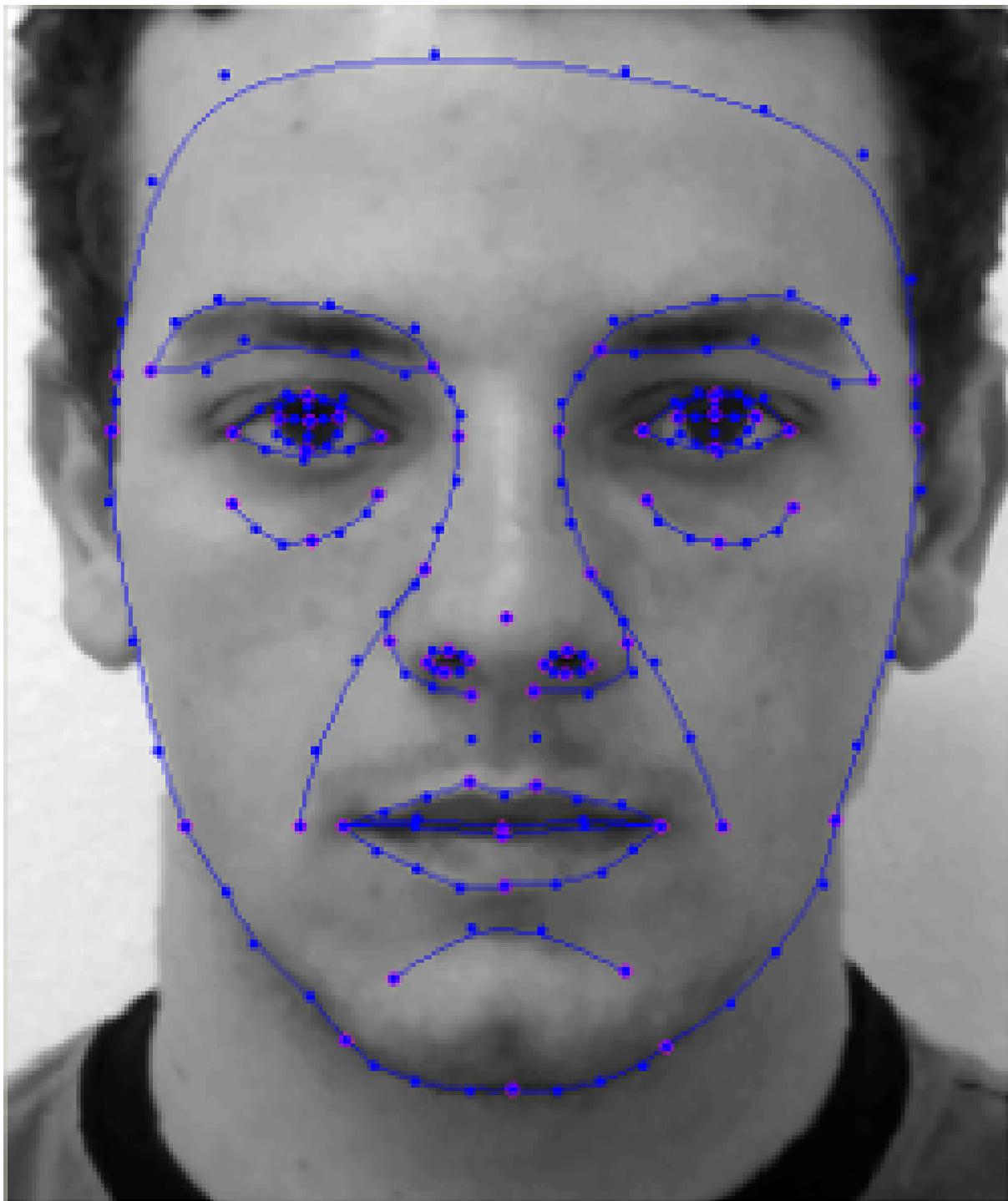


Figura 4 – Posicionamento final dos 61 pontos fiduciais requeridos para interpolação usando *B-Splines*.

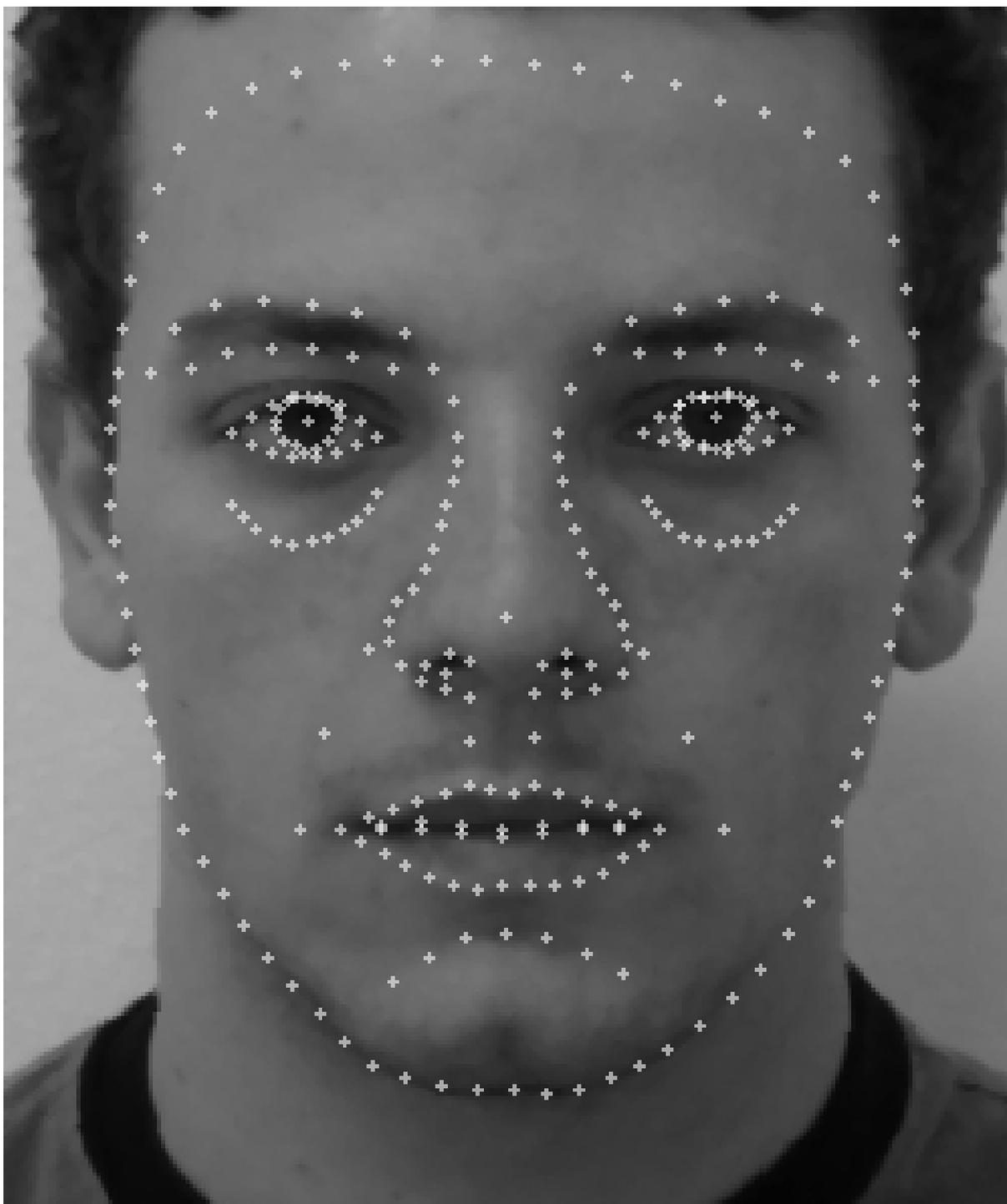


Figura 5 – Conjunto de 284 pontos fiduciais delineando a maior parte das características faciais.

3.2 Registro de Imagens de Face

Para comparar e interpretar a informação entre duas imagens de face, em muitas aplicações, se faz necessário normalizá-las espacialmente, tal que, as características faciais de uma imagem estejam sobrepostas às características correspondentes da outra imagem. Em alguns métodos de reconhecimento de faces, por exemplo, alinhar a região de interesse é de suma importância, tanto que as imagens de face precisam passar por uma etapa de registro antes de comparar suas características (KIRBY; SIROVICH, 1990; AHONEN; HADID; PIETIKÄINEN, 2004). Isto se faz necessário porque mesmo uma imagem de face da mesma pessoa sob duas poses diferentes pode variar de tal maneira que compará-las pode produzir resultados totalmente inesperados. Neste caso, um alinhamento rígido da região facial pode corrigir a variação causada por rotação da pose. Este tipo de variação pode ser classificada como variação global ou rígida. Por outro lado, a face de uma mesma pessoa também pode variar localmente, ou seja, de maneira não-rígida. Por exemplo, quando a variação se dá por causa de uma expressão facial. Isto pode ser compensado aplicando-se um alinhamento não-rígido para as regiões da face correspondentes.

Em linhas gerais, o processo de registro de imagens consiste em encontrar uma transformação T que irá mapear os pontos a de uma imagem de origem para os pontos correspondentes b de uma imagem de referência, tal que:

$$b = T(a), \quad (3.1)$$

onde $a = (\mathbf{a}_x, \mathbf{a}_y)$ e $b = (\mathbf{b}_x, \mathbf{b}_y)$ são pontos nas duas imagens e que correspondem espacialmente um ao outro baseando-se em alguma medida de similaridade. Existem muitos tipos de técnicas de registro de imagens e a escolha da técnica apropriada depende do propósito do registro (ZITOVA; FLUSSER, 2003). Devido ao número elevado de alternativas, a escolha do método apropriado depende muito da deformação geométrica assumida entre as imagens que precisam ser registradas (ZITOVA; FLUSSER, 2003). Outros fatores que influenciam nesta escolha são a métrica de similaridade usada no processo de otimização dos parâmetros da função de transformação e o algoritmo de interpolação usado para deformar a imagem a ser registrada.

No presente trabalho, as distorções geométricas assumidas são de natureza global e local, rígidas e não-rígidas, pois, tem-se o propósito de alinhar imagens de face de pessoas distintas, apresentando variações de escala, leves variações de rotação e, principalmente, variações de expressão facial.

Como as imagens de face utilizadas neste trabalho foram obtidas de uma única modalidade, isto é, apenas um tipo de sensor foi utilizado para capturar as imagens (OLIVEIRA JR; THOMAZ, 2006), pode-se utilizar uma métrica de similaridade que compara diretamente a intensidade dos pixels. Por outro lado, mesmo em ambientes controlados, podem existir efeitos de diferenças de luminosidade nas imagens. Para diminuir a influência destes efeitos no processo de registro das faces, pode-se acrescentar a métrica de similaridade uma medida do deslocamento espacial entre os pontos fiduciais das duas imagens de face de interesse, ou seja, a imagem de face a ser registrada e a imagem de referência.

Portanto, necessita-se de um método de registro de imagens que dê suporte as características observadas acima. Para lidar com as distorções rígidas, optou-se por uma transformação afim que aplica de uma vez transformações de translação, rotação e escala. No caso das distorções não-rígidas, pode-se recorrer a um dos métodos clássicos de registro como aqueles que usam funções de base radial. O mais frequentemente usado é o *Thin-Plate Splines - TPS* (ZITOVA; FLUSSER, 2003) que apresenta bons resultados no registro de imagens, porém, a otimização dos seus parâmetros se torna ineficiente a medida que o número de pontos de controle aumenta. Ou seja, quando se necessita de um ajuste fino para registrar as imagens, demandam-se mais pontos de controle e a otimização de um ponto de controle afeta todos os parâmetros (ZITOVA; FLUSSER, 2003; RUECKERT et al., 1999). Outra abordagem muito utilizada é a *Elastic Body Registration*. Esta técnica também sofre quando as deformações das imagens são muito localizadas (ZITOVA; FLUSSER, 2003), ou seja, a computação de todos os parâmetros é afetada quando precisa-se ajustar apenas um deles. Outra técnica de registro proposta por Rueckert et al. (1999) é a *Free-Form Deformation - FFD*. Ao contrário das técnicas anteriores, na FFD a otimização de um ponto de controle afeta apenas seus vizinhos mais próximos. Isso torna a otimização mais eficiente do que os anteriores (LEE et al., 1996; RUECKERT et al., 1999). A FFD já inclui uma etapa preliminar de otimização dos parâmetros para o registro afim das imagens e a partir destes parâmetros otimizam-se aqueles responsáveis pelas deformações não-rígidas. Nesse contexto, otimizar significa posicionar o ponto de controle de forma que a função de transformação que recebe este ponto como parâmetro possa registrar a imagem de entrada em relação a imagem de referência.

Nas sub-seções seguintes descrevem-se matematicamente as técnicas utilizadas nesta dissertação para registrar as imagens de face em relação a uma imagem de referência.

3.2.1 Transformações Rígidas

As transformações rígidas são definidas como transformações geométricas que preservam todas as distâncias entre os pontos do objeto transformado. Essas transformações também preservam a proporção entre as linhas e os ângulos internos do objeto transformado (SONKA; FITZPATRICK, 2000). Formalmente, este tipo de transformação pode ser descrita como uma combinação linear de rotação R e translação t , tal que:

$$\mathbf{b} = R\mathbf{a} + \mathbf{t}, \quad (3.2)$$

onde $\mathbf{a} = (a_x, a_y)$ é o ponto de entrada, $\mathbf{b} = (b_x, b_y)$ é o ponto transformado, R é a matriz 2×2 de rotação, tal que:

$$R = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \quad (3.3)$$

e $\mathbf{t} = [t_x, t_y]$ é o vetor que descreve a translação do ponto transformado. Um exemplo de transformação rígida aplicada a uma imagem de face pode ser observado na Figura 6.

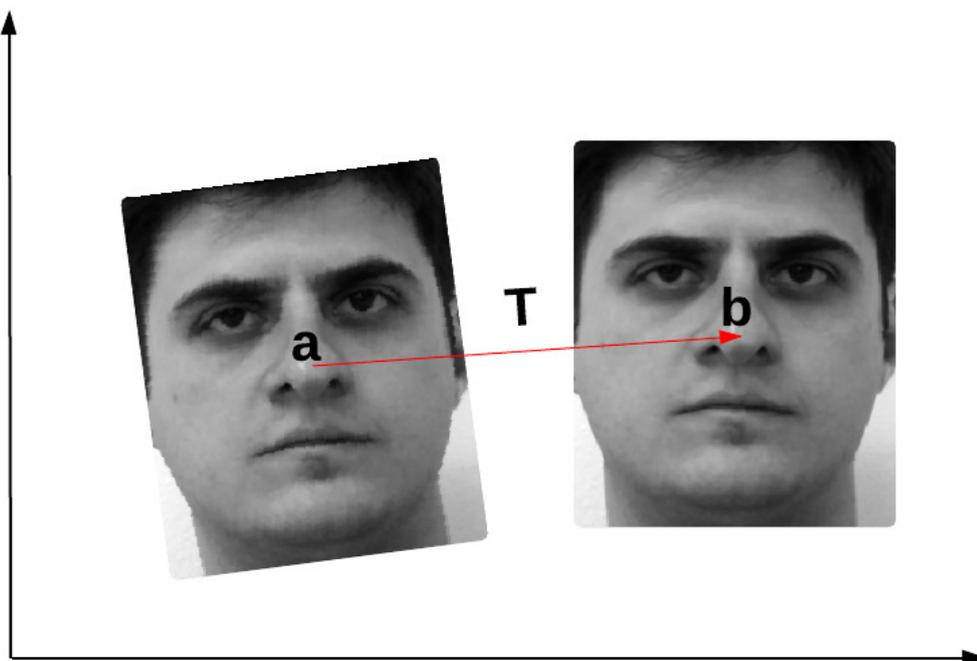


Figura 6 – Transformação rígida aplicada a uma imagem do banco de faces da FEI.

3.2.2 Transformações Afins

Uma classe mais geral de transformações lineares são as transformações afins. Estas diferenciam-se das transformações rígidas somente por um fator de escala. Tais transformações podem ser descritas matematicamente por

$$\mathbf{b} = RE\mathbf{a} + \mathbf{t} \quad (3.4)$$

e

$$\mathbf{b} = ER\mathbf{a} + \mathbf{t}, \quad (3.5)$$

onde $E = \text{diag}(e_x, e_y)$. Como RE geralmente é diferente de ER , as Equações 3.4 e 3.5 representam classes diferentes de transformações afins (SONKA; FITZPATRICK, 2000). Quando $e_x = e_y$, pode-se reduzir a matriz E para um escalar "e". Um exemplo de transformação afim pode ser visto na Figura 7.

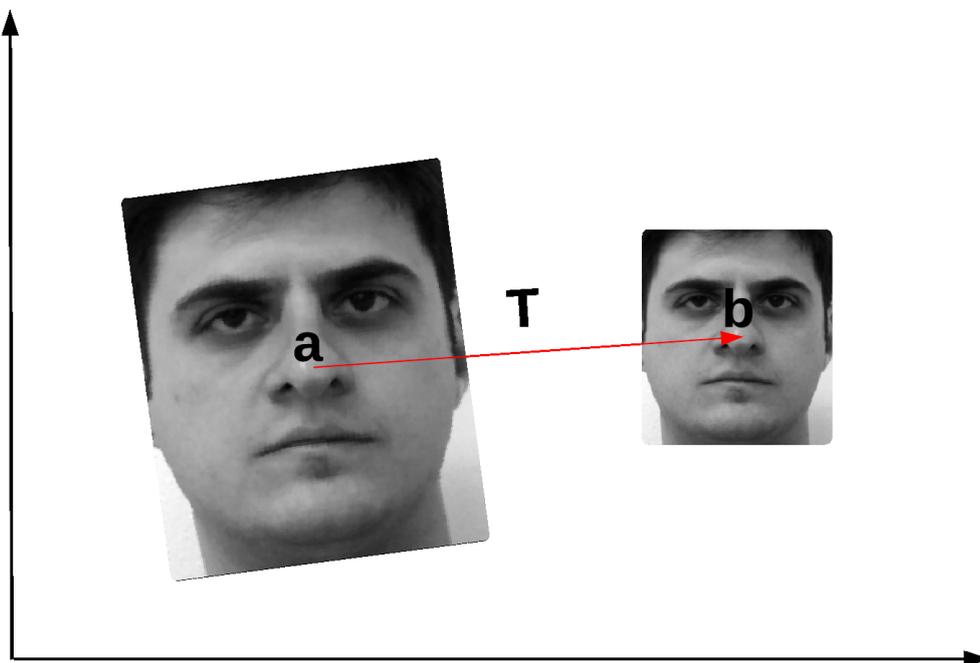


Figura 7 – Transformação afim aplicada a uma imagem do banco de faces da FEI.

3.2.3 Transformações Não-Rígidas baseadas em FFD

Quando deseja-se registrar uma imagem de face em relação a uma face de referência, as distorções geométricas assumidas entre as duas imagens são de natureza não-rígida, ou seja, existem deformações localizadas que não podem ser resolvidas por transformações globais tais como transformações afins. Nestes casos, opta-se por utilizar um método de registro de imagens que dê suporte a transformações globais e locais. Nesta dissertação, implementou-se uma versão 2D da técnica de registro de imagens FFD (RUECKERT et al., 1999).

Esta técnica consiste primeiramente em otimizar os parâmetros de uma função de transformação que mapeia as características da imagem a ser registrada para as características da imagem de referência, para depois aplicar essa função otimizada na deformação da imagem a ser registrada. Esta função de transformação é uma combinação de uma transformação global e uma transformação local, tal que, no caso 2D, pode ser escrita como:

$$\mathbf{T}(x, y)_{ffd} = \mathbf{T}_{global}(x, y) + \mathbf{T}_{local}(x, y). \quad (3.6)$$

A componente global da equação (3.6) é modelada como uma transformação afim de acordo com a subseção 3.2.2. A componente local é definida como a FFD baseada em *B-Splines* (LEE et al., 1996; LEE; WOLBERG; SHIN, 1997). A ideia básica desta transformação é deformar um objeto que está sob uma malha de pontos de controle. Ao movimentar estes pontos, controla-se a forma do objeto subjacente à malha.

Inicia-se a definição da FFD baseada em *B-Splines* definindo-se o domínio da área da imagem a ser registrada como um conjunto $\Omega = \{(x, y) | 1 \leq x < h_1, 1 \leq y < h_2\}$ de pixels e um conjunto $\Phi_{i,j}$ de $n_x \times n_y$ pontos de controle, com espaçamento uniforme δ entre eles, formando uma malha sobreposta à imagem Ω , conforme pode-se observar na Figura 8.

Dadas as definições acima, a FFD pode ser matematicamente descrita como um produto tensorial 2D das já bem estabelecidas *splines* cúbicas (BARTELS; BEATTY; BARSKY, 1987), tal que:

$$\mathbf{T}(u, v)_{local} = \sum_{k=0}^3 \sum_{l=0}^3 B_k(u) B_l(v) \phi_{(i+k)(j+l)}, \quad (3.7)$$

onde $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, $u = x/n_x - \lfloor x/n_x \rfloor$ e $v = y/n_y - \lfloor y/n_y \rfloor$. $B_k(u)$ e $B_l(v)$ são *B-splines* cúbicas usadas como funções de base para a FFD, avaliadas em u e v e são definidas por:

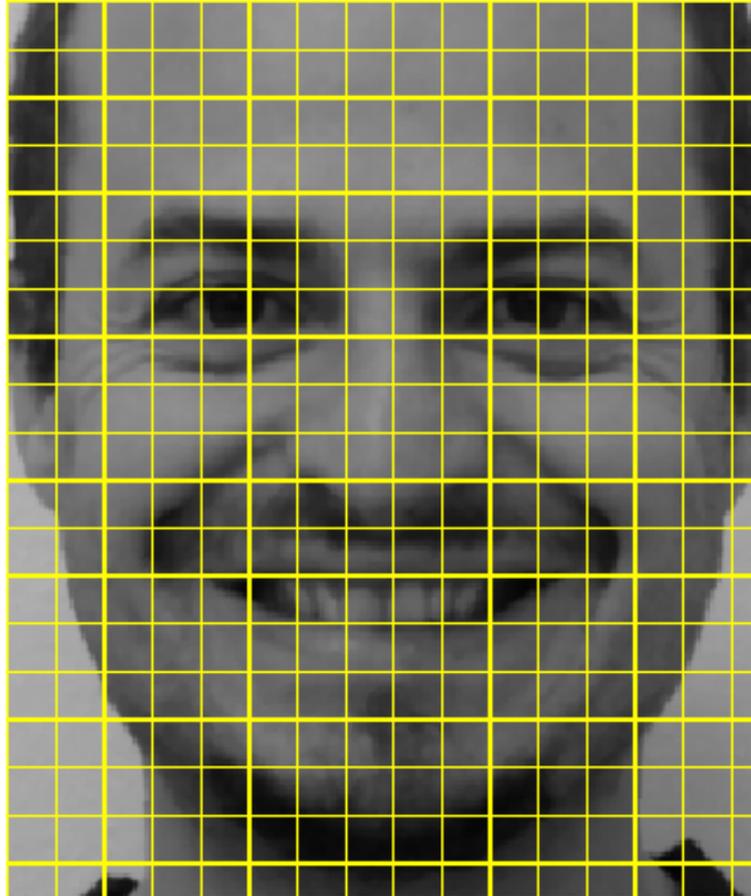


Figura 8 – Malha de pontos de controle sobreposta a uma imagem de face. As posições dos pontos de controle são as intersecções das linhas da malha.

$$B_0(t) = (-t^3 + 3t^2 - 3t + 1)/6$$

$$B_1(t) = (3t^3 - 6t^2 + 4)/6$$

$$B_2(t) = (-t^3 + 3t^2 + 3t + 1)/6$$

$$B_3(t) = t^3/6,$$

onde o parâmetro t , que no caso acima é substituído por u e v , varia de acordo com $0 \leq t < 1$.

Percebe-se que ao movimentar os pontos de controle $\Phi_{i,j}$ de suas posições originais i, j para posições i', j' , pode-se deformar a imagem subjacente aplicando-se a função 3.6 em todo o domínio Ω . Pretende-se então determinar quais são os movimentos necessários aos pontos de controle para que a imagem a ser registrada, depois de deformada, alcance o objetivo do registro, ou seja, minimizar o distância entre os pontos fiduciais da face de entrada e da face de referência e maximizar a similaridade entre elas. Tal objetivo é definido como a minimização de uma função custo C que pode ser descrito matematicamente como:

$$\operatorname{argmin}[C], \quad (3.8)$$

tal que:

$$C = -C_{\text{similaridade}} + \lambda C_{\text{suavidade}}, \quad (3.9)$$

onde $C_{\text{similaridade}}$ é a métrica de similaridade entre as duas imagens. $C_{\text{suavidade}}$ é o termo que regulariza a suavidade da função 3.7 (RUECKERT et al., 1999) e pode ser descrito para a versão 2D da FFD como:

$$C_{\text{suavidade}} = \frac{1}{A} \int_0^{h_1} \int_0^{h_2} \left[\left(\frac{\partial T}{\partial x^2} \right)^2 + \left(\frac{\partial T}{\partial y^2} \right)^2 + \left(\frac{\partial T}{\partial xy} \right)^2 \right] dx dy, \quad (3.10)$$

onde o parâmetro λ controla a importância do termo de suavidade na função custo C e deve-se ajustá-lo para prover um bom compromisso entre a exatidão do alinhamento das imagens e a suavidade da função de deformação, A é a área da imagem a ser transformada e T é a função de transformação usada para registrar a imagem, neste caso $\mathbf{T}(x, y)_{ffd}$.

Para obter uma função de transformação ótima, procura-se um método para minimizar a função custo C que pode ser baseado no seu gradiente, pois este representa a direção da máxima taxa de variação da função C em direção a um mínimo ou máximo local. Métodos de otimização baseados na descida do gradiente calculam o gradiente da função em cada iteração e a busca pelo mínimo é ajustada na direção do gradiente calculado. Porém, este método de otimização pode muitas vezes levar a busca na direção de um mínimo local, o que não necessariamente representa a direção de um mínimo global. A fim de evitar que a busca fique presa em um mínimo local, pode-se otimizar a função custo hierarquicamente, refinando o espaço de busca, ou seja, o número de graus de liberdade cada vez que um valor mínimo é encontrado.

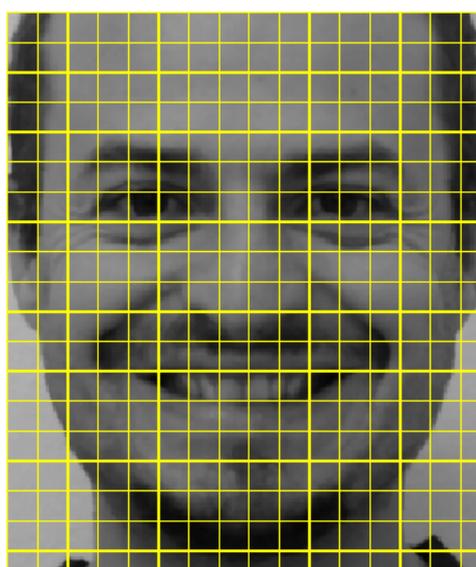
Levando-se em conta as considerações acima, aplica-se o Algoritmo 1, onde primeiro inicializa-se o conjunto de pontos de controle $\Phi_{i,j}$ aplicando-se a transformação afim obtida ao tentar alinhar os pontos fiduciais de ambas imagens. Depois, inicia-se um procedimento de busca iterativo pelo mínimo da função custo C baseado na descida do gradiente, onde, em cada iteração, calcula-se o gradiente da função custo em relação ao conjunto de pontos de controle. Enquanto a norma do vetor gradiente for maior que um determinado limiar ϵ , movimentam-se os pontos de controle proporcionalmente à direção do versor do gradiente de C , regula-se o tamanho destes movimentos pelo parâmetro de passo μ . Ao atingir o valor ótimo

da função custo em uma determinada resolução dos pontos de controle, o algoritmo dobra esta resolução (FORSEY; BARTELS, 1988) até a maior resolução permitida; ou seja, o algoritmo busca novamente o valor ótimo da função custo em cada nível de resolução.

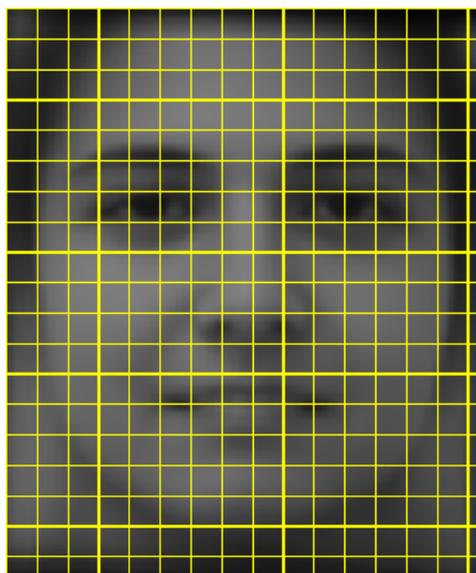
initialize the control points $\Phi_{i,j}$ with the global affine alignment between the two sets of landmarks.
repeat
 calculate the gradient vector of the cost function $\nabla C = \frac{\delta C(\Phi_{i,j})}{\delta \Phi_{i,j}}$
 while $\|\nabla C\| > \epsilon$ **do**
 recalculate the control points $\Phi_{i,j} = \Phi_{i,j} + \mu \frac{\nabla C}{\|\nabla C\|}$
 recalculate the gradient vector ∇C
 end while
 increase the control points resolution by calculating the new lattice of control points $\Phi_{i,j}^{l+1}$ from $\Phi_{i,j}^l$
until Finest level of FFD resolution is reached

Algoritmo 1– Algoritmo de registro de imagens. Fonte: Rueckert et al. adaptado de (RUECKERT et al., 1999)

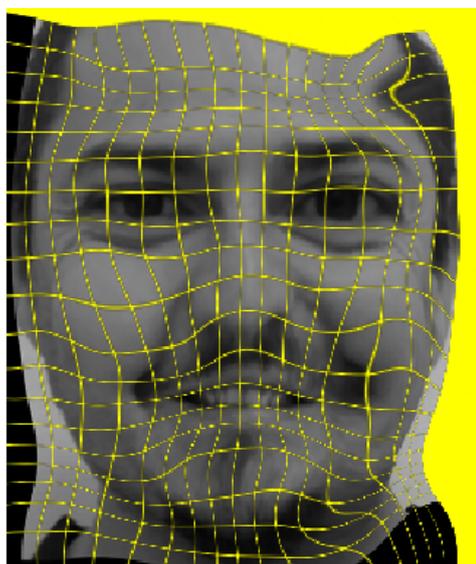
Após a otimização da malha $\Phi_{i,j}$, aplica-se a função 3.6 em todo o domínio da imagem para deformá-la, um exemplo de aplicação pode ser visualizado na Figura 9, onde uma face foi registrada em relação a uma face de referência e as malhas foram sobrepostas nas imagens para ilustrar a movimentação dos pontos de controle da FFD.



(a)



(b)



(c)

Figura 9 – Resultado da aplicação do Algoritmo 1 em uma imagem de face: (a) Imagem de face original sobreposta por uma malha de pontos de controle; (b) Face de referência usada para registrar a face de entrada; (c) A mesma imagem de face deformada após a aplicação da FFD sobreposta pela malha de pontos otimizada.

3.3 Métricas de Similaridade

O objetivo do registro de imagens usando 2D FFD é maximizar a semelhança entre uma imagem de entrada e uma imagem de referência. Para isso, aplica-se o Algoritmo 1 que procura minimizar a função custo C composta por um termo de suavidade e uma métrica de similaridade. Uma revisão de algumas métricas de similaridade é feita nas sub-seções abaixo.

3.3.1 Métricas baseadas na Intensidade dos Pixels

Estas métricas são baseadas na diferença entre as intensidades dos pixels correspondentes em ambas as imagens. Dadas duas imagens I_1 e I_2 , a similaridade entre as duas imagens é a somatória das distâncias de cada par de pixels correspondentes.

3.3.1.1 Correlação Cruzada (CC)

A CC entre duas imagens pode ser descrita como:

$$S_{CC} = \sum_{x=1}^n (I_1(x) \cdot I_2(x)). \quad (3.11)$$

Esta métrica de similaridade assume uma correspondência linear entre as intensidades dos pixels das imagens, sendo sensível à diferenças de luminosidade e contraste entre elas. Para reduzir o impacto destes ruídos, uma versão normalizada desta métrica pode ser utilizada, tal que:

$$S_{CC} = \sum_{x=1}^n \frac{(I_1(x) - \bar{I}_1)(I_2(x) - \bar{I}_2)}{\sqrt{(I_1(x) - \bar{I}_1)^2 (I_2(x) - \bar{I}_2)^2}}, \quad (3.12)$$

onde \bar{I}_1 e \bar{I}_2 são as intensidades médias das duas imagens comparadas.

3.3.1.2 Distância Euclidiana (DE)

A distância Euclidiana entre duas imagens baseando-se nas intensidades dos pixels, pode ser definida como:

$$S_E = \sum_{x=1}^n \sqrt{(I_1(x) - I_2(x))^2}. \quad (3.13)$$

Quando os pixels de ambas as imagens têm a mesma intensidade, o valor desta métrica atinge o mínimo zero. Assim como na correlação cruzada, esta métrica pode ser seriamente afetada pelas diferenças de luminosidade das imagens. Uma versão menos suscetível a estes ruídos é a distância de Manhattan, matematicamente definida como:

$$S_E = \sum_{x=1}^n |(I_1(x) - I_2(x))|. \quad (3.14)$$

3.3.2 Métricas Baseadas em Entropia

Na tentativa de suprimir os efeitos negativos dos ruídos provocados pelas diferenças de luminosidade, métricas de similaridade baseadas em entropia, as quais usam a informação da imagem inteira, podem ser empregadas.

3.3.2.1 Entropia Conjunta (EC)

A entropia conjunta entre duas imagens procura medir quão bem alinhadas elas estão usando o conceito de entropia de *Shannon*. Como mencionado anteriormente, podemos representar as intensidades de duas imagens pelos vetores I_1 e I_2 e sua entropia conjunta pode ser descrita como:

$$S_{EC} = H(I_1, I_2) = - \sum_{x=1}^n p(I_1(x), I_2(x)) \log p(I_1(x), I_2(x)), \quad (3.15)$$

onde $p(I_1(x), I_2(x))$ representa a função de densidade de probabilidade conjunta das imagens I_1 e I_2 . É esperado que quanto mais alinhadas as imagens estiverem, menor será entropia conjunta entre elas. Porém, como mostrado em Collignon et al. (1995), valores baixos de entropia conjunta podem ser alcançados com um alinhamento pobre entre as imagens, por exemplo, alinhando apenas o plano de fundo das imagens.

3.3.2.2 Informação Mútua (IM)

Uma métrica alternativa a EC é a informação mútua (IM) entre as imagens que, além da entropia conjunta, leva em consideração as entropias marginais das duas imagens comparadas. A IM mede quanta informação uma variável tem sobre a outra (neste caso as variáveis são as intensidades das imagens), em lugar de comparar as intensidades diretamente.

As entropias marginais das imagens I_1 e I_2 podem ser descritas como:

$$H(I_1) = - \sum_{x=1}^n p(I_1(x)) \log p(I_1(x)), \quad (3.16)$$

$$H(I_2) = - \sum_{x=1}^n p(I_2(x)) \log p(I_2(x)), \quad (3.17)$$

onde $p(I_1(x))$ e $p(I_2(x))$ são as probabilidades dos pixels com intensidades $I_1(x)$ e $I_2(x)$ ocorrerem nas imagens correspondentes. A informação mútua entre as imagens é dada por:

$$S_{IM}(I_1, I_2) = H(I_1) + H(I_2) - H(I_1, I_2) = \sum_{x=1}^n p(I_1(x), I_2(x)) \log \frac{p(I_1(x), I_2(x))}{p(I_1(x)) \cdot p(I_2(x))}. \quad (3.18)$$

Pode-se observar pela Equação (3.18) que minimizar a entropia conjunta aumenta a informação mútua. Porém, a introdução das entropias marginais penaliza uma redução do montante de informação em cada imagem. Portanto, a IM é menos sensível ao alinhamento pobre das imagens do que a entropia conjunta.

A equação (3.18) é equivalente a divergência de Kullback-Leibler (KULLBACK; LEIBLER, 1951) entre as distribuições de probabilidades $p(I_1(x), I_2(x))$ e $p(I_1(x)) \cdot p(I_2(x))$. Quando I_1 e I_2 são completamente independentes, $p(I_1(x), I_2(x)) = p(I_1(x)) \cdot p(I_2(x))$ e

$S_{IM} = 0$. A IM também pode ser vista como uma medida de dependência entre as imagens: quanto mais dependente elas forem, maior será o valor de S_{IM} .

Adicionalmente, a informação mútua entre duas imagens pode ser equivalentemente expressa por:

$$\begin{aligned} S_{IM}(I_1, I_2) &= H(I_1) + H(I_2) - H(I_1, I_2) \\ &= H(I_1) - H(I_1|I_2) \\ &= H(I_2) - H(I_2|I_1), \end{aligned} \quad (3.19)$$

onde $H(I_1|I_2)$ é a entropia condicional definida como:

$$H(I_1|I_2) = - \sum_{x=1}^n p(I_1(x), I_2(x)) \log p(I_1(x)|I_2(x)). \quad (3.20)$$

De acordo com a formulação acima, pode-se interpretar a informação mútua como a redução da incerteza de I_1 (ou I_2) devido ao conhecimento sobre I_2 (ou I_1).

3.3.2.3 Informação Mútua Normalizada (IMN)

Embora a métrica IM seja menos sensível ao alinhamento pobre das imagens do que a EC, ela ainda não apresenta a propriedade da invariância. Outra métrica que procura resolver as deficiências da IM é a informação mútua normalizada (IMN) proposta por Studholme, Hill e Hawkes (1999) e matematicamente descrita por:

$$S_{IMN}(I_1, I_2) = \frac{H(I_1) + H(I_2)}{H(I_1, I_2)}. \quad (3.21)$$

3.3.3 Métrica de Distância entre Pontos Fiduciais

Além das métricas de similaridade baseadas na informação dada pelos pixels das imagens, pode-se optar por métricas baseadas em características extrínsecas, tais como pontos fiduciais manual ou automaticamente marcados. Pode-se então, comparar as localizações dos

pontos marcados nas duas imagens usando uma medida de deslocamento espacial entre eles. Dadas duas matrizes de pontos fiduciais $C_{q \times 2}$ e $D_{q \times 2}$, pode-se definir a distância normalizada entre os pontos como:

$$DNP = \frac{\sum_1^q [(C_{q,1} - D_{q,1})^2 + (C_{q,2} - D_{q,2})^2]}{n_1^2 + n_2^2}, \quad (3.22)$$

onde cada linha das matrizes C e D representa um ponto fiducial de cada imagem comparada e a primeira coluna de cada matriz é a coordenada x de sua localização na imagem e a segunda coluna sua coordenada y . As variáveis n_1 e n_2 são, respectivamente, os números de pixels das imagens 1 e 2.

3.4 Construção de um Atlas

Para registrar as imagens usando um método afim ou não-rígido, faz-se necessário a construção de um atlas, o qual servirá como imagem de referência para alinhamento de todas as outras imagens. Essa imagem de referência é geralmente construída com uma ou mais amostras da população estudada e deve representar a estrutura típica desta população (BHATIA, 2007).

Dado que a estrutura de uma única amostra de imagem de face nunca poderia representar a face típica de uma população, há um interesse em desenvolver faces de referência que representem a população estudada usando a informação de múltiplos exemplos da amostra. Um método para combinar múltiplos indivíduos na construção de uma imagem de referência baseia-se nas intensidades dos pixels.

Imagens de referência baseadas em intensidade envolvem gerar uma representação média das faces calculando a média pixel a pixel a partir dos múltiplos exemplos utilizados na construção, de acordo com a equação

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i, \quad (3.23)$$

onde N é o número de exemplos utilizados e \mathbf{x}_i é o vetor contendo as intensidades dos pixels da i -ésima face de exemplo da população estudada.

Porém, antes de calcular a face de referência utilizando a média dos pixels dos exemplos disponíveis, é necessário normalizar os exemplos de acordo com algum sistema de coordenadas

padrão, ou seja, para que a face de referência apresente uma estrutura típica algumas características das amostras precisam estar alinhadas.

Uma maneira de alinhar as amostras de face é estabelecer dois pontos que representem características particulares que, no caso de imagens frontais de face, podem ser as coordenadas dos centros das pupilas e a distância entre elas. Um exemplo de atlas construído desta forma, utilizando um alinhamento afim (AMARAL; THOMAZ, 2009), pode ser visto na Figura 10. Este atlas foi construído utilizando como amostras as 200 imagens neutras e frontais do banco da FEI. O resultado é uma imagem suavizada na maioria das regiões da face, exceto na região dos olhos: característica utilizada com referência para normalização de todas as faces.

Na tentativa de criar uma imagem de referência que represente a estrutura típica da amostra de faces utilizada neste trabalho, porém sem o aspecto suavizado da Figura 10, aplicou-se um procedimento iterativo de registro não-rígido das faces em relação a uma face de referência, conforme Figura 11. Este procedimento aplicado diversas vezes tem por objetivo minimizar o efeito de suavização visto na média produzido pelas amostras rigidamente normalizadas e também de criar uma face de referência mais andrógina e sem expressão possível.



Figura 10 – Imagem média neutra calculada a partir do banco de faces da FEI.

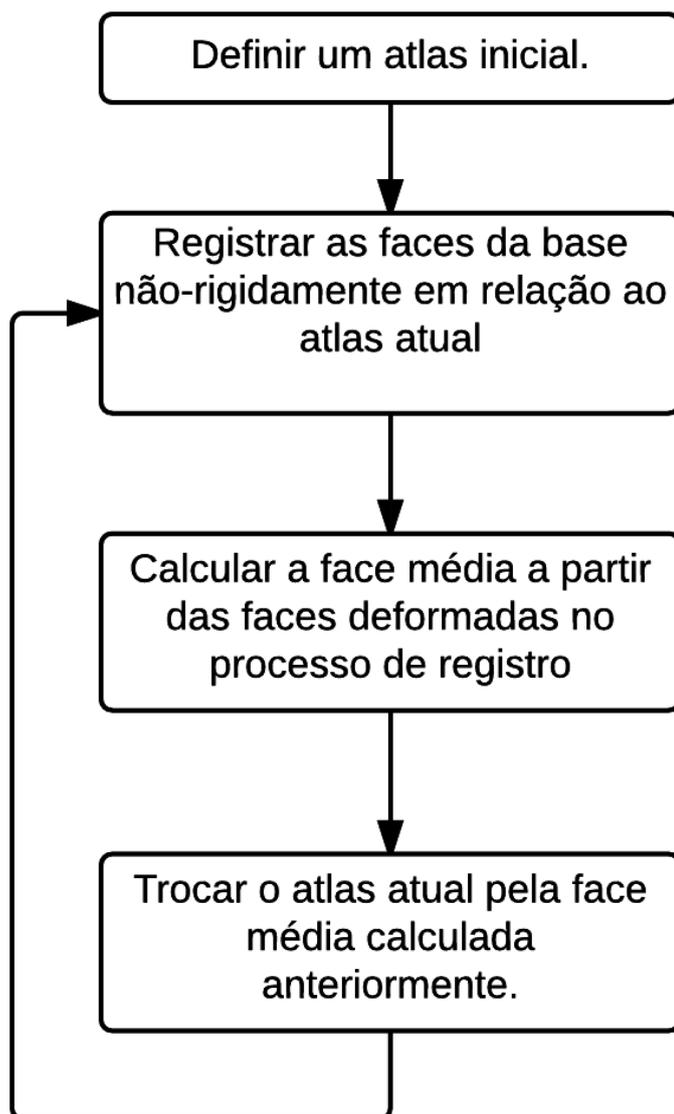


Figura 11 – Procedimento iterativo de criação de um atlas para registro não-rígido de imagens frontais de face.

3.5 Características Mais Expressivas

Assim como em outros estudos (KIRBY; SIROVICH, 1990; SWETS; WENG, 1996; COOTES; EDWARDS; TAYLOR, 2001; COOTES; TAYLOR et al., 2004; TODOROV et al., 2008; TODOROV; OOSTERHOF, 2011), aplicou-se a Análise de Componentes Principais ou, em inglês, *Principal Components Analysis* (PCA), para descrever de maneira não-supervisionada as características mais expressivas de um conjunto de imagens de face em termos da variância das amostras. Embora o PCA seja uma técnica já bem estabelecida (FUKUNAGA, 1990; JOHNSON; WICHERN, 2007) e tenha sido utilizada em muitos problemas de reconhecimento de imagem, extração de características e redução de dimensionalidade, aborda-se a seguir sua teoria de forma contextualizada para sustentar o uso desta técnica.

O PCA é uma técnica de estatística multivariada que visa explicar a estrutura de variância-covariância de uma amostra de dados através de combinações lineares de menor dimensionalidade que os dados originais. No contexto de percepção facial e análise de padrões, uma imagem com n pixels pode ser interpretada como um ponto em um espaço n -dimensional, também conhecido como espaço de faces (KIRBY; SIROVICH, 1990; TURK; PENTLAND, 1991), ou seja, cada pixel da imagem representa o valor de uma das coordenadas que formam o vetor $\mathbf{x}^T = \{x_1, x_2, \dots, x_n\}$, obtido ao concatenar as linhas (ou colunas) da matriz de pixels da imagem. Sabe-se que imagens de face normalizadas espacialmente em relação a um sistema de coordenadas, tal como em Amaral e Thomaz (2009), ou em relação a um atlas de referência são altamente redundantes. Como consequência, uma imagem com n pixels pode ser projetada em um espaço com menor dimensionalidade sem perda significativa de informações.

Seja a matriz $X_{n \times N} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}^T$ um conjunto de treinamento, composto de N amostras de entrada (ou padrões de interesse, tais como imagens de face) com n variáveis (ou atributos, tais como pixels). Cada linha da matriz $X_{n \times N}$ representa valores particulares de uma variável observada em N amostras. Seja $S_{n \times n}$ a matriz de covariância das amostras em $X_{n \times N}$, tal que

$$S = \frac{1}{(N-1)} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T, \quad (3.24)$$

onde $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T$ e $\bar{\mathbf{x}}$ é o vetor médio de $X_{n \times N}$ dado por

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i = \frac{1}{N} \sum_{i=1}^N [x_{i1}, x_{i2}, \dots, x_{in}]^T. \quad (3.25)$$

Problemas que envolvem padrões de interesse cuja dimensionalidade n é muito maior que a número de amostras N , tais como imagens de faces, são chamados de problemas com número pequeno de amostras (FUKUNAGA, 1990). Em tais casos, sabe-se que o posto da matriz de covariância $S_{n \times n}$ é $N - 1$ ou menor. Para este tipo de problema, pode-se optar por calcular a matriz de covariância das amostras da seguinte forma:

$$S_{N \times N} = \frac{1}{(N - 1)} (U^T U)_{N \times N}, \quad (3.26)$$

onde $U = [\mathbf{x}_1 - \bar{\mathbf{x}} \dots \mathbf{x}_N - \bar{\mathbf{x}}]_{n \times N}$. Em Fukunaga (1990), demonstrou-se que é possível obter os autovalores e autovetores da matriz de covariância $S_{N \times N}$ sem perder informação, dado que os outros $(n - N)$ autovalores são iguais a zero e os correspondentes autovetores são indefinidos. Os detalhes desse truque algébrico podem ser vistos no Apêndice A.

Sejam P e Λ as matrizes de autovetores e autovalores de $S_{n \times n}$, tal que:

$$P_{k \times n}^T S_{n \times n} P_{n \times k} = \Lambda_{k \times k}. \quad (3.27)$$

É um resultado provado na literatura que o conjunto de k ($k \leq n$) autovetores de S , os quais correspondem aos k maiores autovalores, minimiza o erro médio quadrático de reconstrução sobre todas as escolhas possíveis de base com k vetores ortonormais (FUKUNAGA, 1990), neste caso $k = N - 1$. Dado que o conjunto de autovetores $P = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k]$ define um novo sistema de coordenadas ortogonal da matriz de treinamento $X_{n \times N}$, os k componentes principais podem substituir as n variáveis iniciais e o conjunto de dados original, consistindo de N amostras com n variáveis, é reduzido para um conjunto de dados de N amostras com k componentes principais.

Levando-se em conta que a face média $\bar{\mathbf{x}}$ é um ponto n -dimensional que retém todas as características comuns das amostras de face, pode-se usar este ponto como um protótipo ou referência para se mover ao longo dos componentes principais e gerar artificialmente as características mais expressivas no espaço de faces. Analogamente a muitos trabalhos (COOTES; EDWARDS; TAYLOR, 2001; COOTES; TAYLOR et al., 2004; COOTES et al., 1995; GIBSON, 2006) podem-se criar novas imagens de face navegando-se pelos componentes principais usando a seguinte expressão:

$$\mathbf{y}_{(n \times 1)} = \bar{\mathbf{x}}_{(n \times 1)} + j \sqrt{\lambda_i} \cdot \mathbf{p}_{i(n \times 1)}, \quad (3.28)$$

onde $j \in \{-3, -2, -1, 0, 1, 2, 3\}$ e \mathbf{p}_i são os k primeiros componentes principais ordenados de forma decrescente de acordo com seus respectivos autovalores, tal que $\lambda_1 \geq \lambda_2 \dots \geq \lambda_k$.

Nesta dissertação, utilizou-se o PCA com o objetivo de criar alterações faciais realísticas, onde cada subconjunto de novas amostras geradas através da navegação nas correspondentes componentes descrevem faces inexistentes. Ao usar tal estratégia não-supervisionada, no entanto, não garante-se que uma única componente principal capturará exclusivamente a variação de uma característica facial específica, por exemplo, gênero ou expressão facial. Em tais situações, seria mais apropriado navegar ao longo de dimensões do espaço de faces estimadas por métodos supervisionados.

3.6 Características Mais Discriminantes

A extração supervisionada das características mais discriminantes das imagens de face registradas não-rigidamente baseou-se na técnica de estatística multivariada chamada de Análise de Discriminantes Lineares ou, do inglês, *Linear Discriminant Analysis* (LDA) (FUKUNAGA, 1990).

O propósito principal desta técnica é separar as amostras de grupos distintos maximizando sua separabilidade inter-classes e, ao mesmo tempo, minimizando a variabilidade intra-classes. O LDA assume implicitamente que as matrizes de covariância de cada classe são iguais e por isso a mesma matriz de espalhamento intra-classes é usada para todas as classes consideradas (TENORIO; THOMAZ, 2011).

Seja a matriz de espalhamento inter-classes S_b , descrita matematicamente por

$$S_b = \sum_{i=1}^g N_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \quad (3.29)$$

e a matriz de espalhamento intra-classes S_w descrita matematicamente por

$$S_w = \sum_{i=1}^g (N_i - 1) S_i = \sum_{i=1}^g \sum_{j=1}^{N_i} (\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i)(\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i)^T, \quad (3.30)$$

onde $\mathbf{x}_{i,j}$ é amostra de imagem de face n -dimensional j da classe i , N_i é o número de padrões de treinamento da classe i e g é o número total de classes ou grupos (TENORIO; THOMAZ, 2011). O vetor $\bar{\mathbf{x}}_i$ e a matriz S_i são, respectivamente, a média e a matriz de covariância amostral da classe i (FUKUNAGA, 1990). O vetor médio global $\bar{\mathbf{x}}$ pode ser então reescrito por

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^g N_i \bar{\mathbf{x}}_i = \frac{1}{N} \sum_{i=1}^g \sum_{j=1}^{N_i} \mathbf{x}_{i,j}, \quad (3.31)$$

onde N é o número total de amostras, que também pode ser escrito como $N = N_1 + N_2 + \dots + N_g$. É importante notar que a matriz de espalhamento intra-classes definida na Equação 3.30 é essencialmente a matriz de covariância ponderada padrão S_p multiplicada por um escalar $(N - g)$, onde S_p pode ser descrita matematicamente por

$$S_p = \frac{1}{(N - g)} \sum_{i=1}^g (N_i - 1) S_i = \frac{(N_1 - 1) S_1 + (N_2 - 1) S_2 + \dots + (N_g - 1) S_g}{N - g}. \quad (3.32)$$

O principal objetivo do método LDA é encontrar uma matriz de projeção P_{lda} que maximize a razão entre o determinante da matriz de espalhamento inter-classes S_b e o determinante da matriz de espalhamento intra-classes S_w , conhecido como critério de Fisher e descrito matematicamente por

$$P_{lda} = \underset{n}{\operatorname{argmax}} \frac{\left| P^T S_b P \right|}{\left| P^T S_w P \right|}. \quad (3.33)$$

O critério de Fisher, descrito na equação 3.33, é maximizado quando a matriz de projeção P_{lda} é composta dos $(g - 1)$ autovetores de $S_w^{-1} S_b$, cujos autovalores correspondentes são não-nulos (DEVIJVER; KITTLER, 1982; FUKUNAGA, 1990). Este é o procedimento LDA padrão (KITANI; THOMAZ; GILLIES, 2006).

No entanto, o desempenho do método LDA pode ser seriamente comprometido quando o número total de amostras N não é consideravelmente maior que o número de variáveis n . Para não ocorrer instabilidades no cálculo da matriz inversa de S_w , recomenda-se que N seja de cinco a dez vezes maior que $(n + g)$ (JAIN; CHANDRASEKARAN, 1982).

Para evitar os problemas de instabilidade do procedimento padrão do método LDA acima mencionados, calculou-se a matriz de projeção P_{lda} através do método denominado MLDA (*Maximum uncertainty LDA*) que considera os problemas de instabilidade de S_w (THOMAZ; GILLIES; FEITOSA, 2004; THOMAZ; KITANI; GILLIES, 2006). Esta técnica consiste em substituir a matriz de espalhamento S_w por outra matriz regularizada S_w^* . O método MLDA pode ser descrito seguindo os seguintes passos:

Calculam-se os autovetores Φ e os autovalores Λ de S_p , onde $S_p = \frac{S_w}{N-g}$.

Calcula-se a média $\bar{\lambda}$ dos autovalores de S_p , dada por

$$\bar{\lambda} = \frac{1}{n} \sum_{i=1}^n \lambda_i = \frac{\text{Trace}(S_p)}{n}; \quad (3.34)$$

Contrói-se uma nova matriz de autovalores baseada na dispersão dos maiores valores entre cada autovalor λ_n e a média $\bar{\lambda}$, matematicamente descrita por

$$\Lambda^* = \text{diag}[\max(\lambda_1, \bar{\lambda}), \max(\lambda_2, \bar{\lambda}), \dots, \max(\lambda_n, \bar{\lambda})]; \quad (3.35)$$

Calcula-se a matriz de espalhamento intra-classes regularizada

$$S_w^* = S_p^*(N - g) = (\Phi \Lambda^* \Phi^T)(N - g). \quad (3.36)$$

O método MLDA é construído substituindo S_w por S_w^* na fórmula do critério de Fisher descrita na equação (3.33). Com esta substituição regulariza-se o critério de Fisher quando o número total de amostras é limitado com relação à dimensionalidade dos dados (KITANI; THOMAZ; GILLIES, 2006).

Assim como descrito na estratégia não-supervisionada para extrair as características mais expressivas, pode-se mover ao longo das direções discriminantes e gerar artificialmente faces compostas das respectivas características dadas pelos grupos de amostras separados. Assumindo que o espalhamento dos grupos de amostras segue uma distribuição Gaussiana, o procedimento de navegar nas dimensões mais discriminantes (TENORIO; THOMAZ, 2011) pode ser descrito matematicamente por:

$$\mathbf{y}_{(n \times 1)} = \bar{\mathbf{x}}_{(n \times 1)} + j\sigma_i \cdot P_{(n \times k)} \cdot \mathbf{w}_{(k \times 1)}, \quad (3.37)$$

onde $j \in \{-3, -2, -1, 0, 1, 2, 3\}$, σ_i é o desvio padrão de cada grupo de amostras $i \in \{1, 2\}$, cobrindo-se assim 99.73% da variação. Observa-se pela Equação 3.37 que computou-se o vetor discriminante $\mathbf{w} = P_{mlda}$ no espaço do PCA, onde primeiro computou-se o PCA a partir da matriz de covariância $S_{n \times n}$ das amostras representadas pela matriz $X_{n \times N}$, obtendo-se o conjunto de autovetores $P_{n \times k}$ e o conjunto de autovalores $\Lambda_{k \times k}$, retendo-se todos os componentes principais com os respectivos autovalores maiores que zero, sendo $k = N - 1$ componentes. Estes representando a totalidade da variância das amostras sem perda de informação. Projetaram-se as amostras originais no espaço de faces do PCA de acordo com

$$W_{N \times k} = (X^T)_{N \times n} P_{n \times k}, \quad (3.38)$$

obtendo-se um novo conjunto de amostras $W_{N \times k}$ com a dimensionalidade reduzida. Aplicou-se, então, o método MLDA nas amostras projetadas no espaço do PCA, obtendo-se o vetor discriminante w . Porém, o vetor discriminante obtido está no espaço do PCA, para navegar na sua direção no espaço original das amostras é necessário desfazer a transformação, multiplicando-o pela esquerda pela matriz de autovetores $P_{n \times k}$, conforme mostrado na Equação 3.37.

É importante enfatizar que computar o PCA como um passo intermediário torna-se necessário quando $n \gg N$, permitindo que as matrizes de espalhamento do LDA possam ser calculadas em computadores com espaço de memória regular. Em situações onde $N > n$, computar o PCA como passo intermediário para redução de dimensionalidade torna-se desnecessário.

4 EXPERIMENTOS E RESULTADOS

Neste capítulo apresentam-se os materiais utilizados nos experimentos, tais como: a base de faces processada, os pontos fiduciais pré-existent e anotados neste trabalho, o atlas de referência escolhido e os conjuntos de imagens deformados não-rigidamente. Descrevem-se também os resultados dos estudos das características mais expressivas e mais discriminantes dos conjuntos de imagens de face processados, fazendo-se uma comparação dessas reconstruções usando diferentes conjuntos de imagens de face.

4.1 Banco de Faces

Para este trabalho utilizou-se o banco de faces da FEI disponível publicamente para pesquisa ¹ (THOMAZ; GIRALDI, 2010). Esta base contém 400 imagens de faces 2D, sendo 200 neutras ou sem expressão e 200 com expressão facial sorrindo e é formado por 200 indivíduos, sendo 100 homens e 100 mulheres, onde cada indivíduo contribuiu com uma imagem de face neutra e outra sorrindo. A maioria das imagens representa indivíduos entre 19 e 40 anos. Todas as imagens foram tomadas frontalmente contra um plano de fundo branco e homogêneo, onde variou-se a escala das faces aproximadamente em 10%. O tamanho original de cada imagem é 640×480 pixels. As imagens estão em formato RGB e os indivíduos representados apresentam diferenças em aparência, estilo do cabelo e adornos.

Para o propósito de comparações, utilizou-se uma versão normalizada desta base de dados, onde todas as 400 amostras foram previamente alinhadas usando a posição das pupilas e a direção determinadas entre elas como medidas de referências, tal que as características baseadas nos pixels das imagens correspondam grosseiramente a mesma localização entre todas elas. Após a normalização, recortaram-se todas as imagens de modo que ficassem com 250 pixels de largura e 300 pixels de altura e também convertidas em escala de cinza de 8 *bits*, correspondendo a 256 níveis. A Figura 12 ilustra algumas amostras rigidamente registradas, nomeadas aqui de conjunto de faces originais.

Para executar uma análise multivariada não-supervisionada, todas as 400 amostras do conjunto original de faces foram utilizadas como entrada para o estudo das características mais

¹<http://www.fei.edu.br/cet/facedatabase.html>



Figura 12 – Um subconjunto de imagens de face retiradas do banco de faces da FEI.

expressivas. Para a análise multivariada supervisionada, os seguintes grupos de amostras foram organizados: masculino contra feminino e neutro contra sorrindo, permitindo a análise de características faciais orientadas pela variação de gênero e expressão facial. Nos experimentos de gênero, os conjuntos de treinamento foram compostos de 200 amostras masculinas, contendo faces neutras e sorrindo e 200 amostras femininas análogas. Para os experimentos com expressões, utilizaram-se 200 imagens neutras, com amostras femininas e masculinas e 200 sorrindo, analogamente com mistura de gênero.

4.2 Marcação de 46 Pontos Fiduciais

Além do banco de faces da FEI, utilizaram-se os pontos fiduciais marcados nas 400 faces rigidamente registradas, de acordo com o trabalho de Tenorio e Thomaz (TENORIO; THOMAZ, 2011). Como mencionado anteriormente, no trabalho citado marcou-se um conjunto de 46 pontos fiduciais em cada imagem de face com o intuito de realizar uma análise multilinear do formato das faces, ou seja, derivou-se o formato das faces delineando-se semi-retas entre alguns desses pontos. Dos 46 pontos, 19 são semanticamente coincidentes as existentes no trabalho de Vukadinovic e Pantic (VUKADINOVIC; PANTIC, 2005) e 27 deles oriundos do trabalho de Lanitis *et al.* (LANITIS; TAYLOR; COOTES, 1995). Um exemplo de face marcada com os 46 pontos fiduciais do conjunto original de faces pode ser observado na Figura 13.

A face média compostas pelas 200 amostras neutras do conjunto original de faces também foi anotada manualmente com os respectivos 46 pontos fiduciais das outras imagens, tal

como em (XAVIER; THOMAZ, 2014). Exemplos dessas marcações também podem ser vistos na Figura 13.

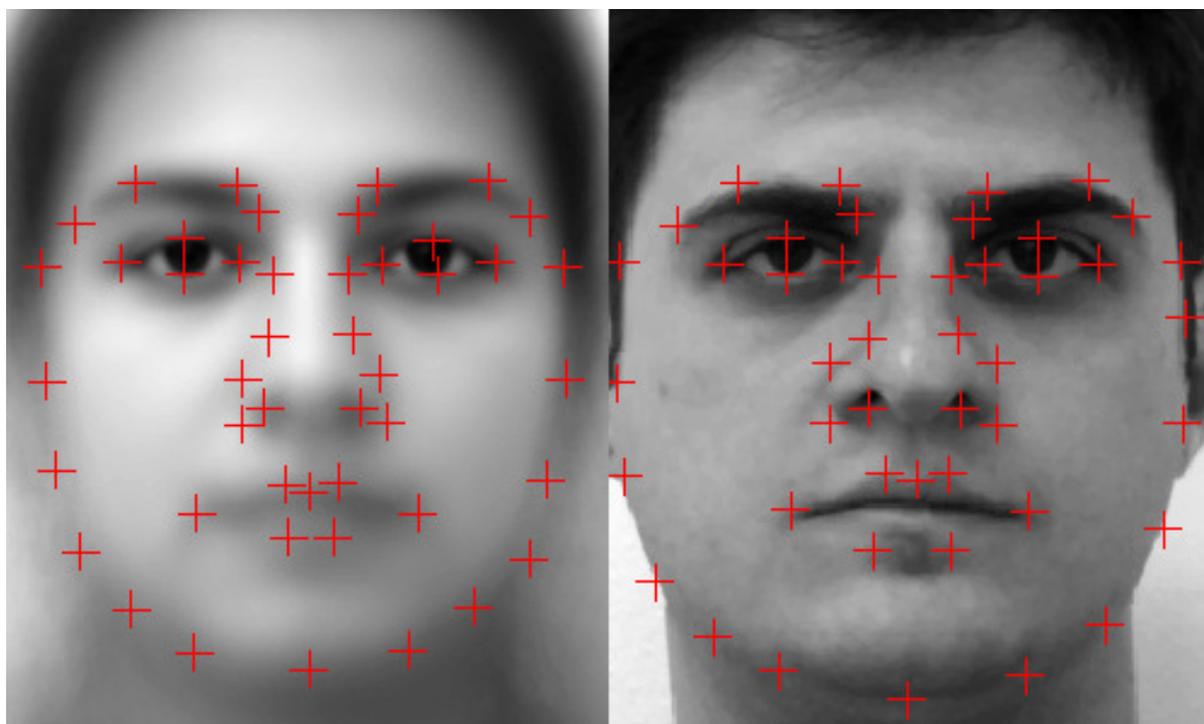


Figura 13 – Anotações feitas nas imagens do banco de faces da FEI. À esquerda: face média neutra do banco de dados anotada com 46 pontos fiduciais. À direita: exemplo de imagem do banco de dados anotada com os mesmos 46 pontos da face média.

As marcações foram feitas visando destacar as principais características das faces, tais como olhos, nariz, boca, perfil do rosto e sobrancelhas. Essas marcações foram aproveitadas para os experimentos de registro de faces através do método FFD não-rígido, gerando um conjunto novo de faces deformadas e utilizadas como entrada nos experimentos de análise multivariada não-supervisionadas e supervisionadas.

4.3 Marcação de 284 Pontos Fiduciais

Com o propósito de registrar um número maior de características faciais do que aquele definido na Seção 4.2, aplicou-se o procedimento descrito na Sub-seção 3.1.2 para marcar as 400 imagens do conjunto original de faces com 284 pontos fiduciais. Como visto anteriormente, para cada face, posicionaram-se inicialmente 3 pontos: localização das pupilas e ponta do queixo. Após um posicionamento automático da ferramenta, ajustou-se a localização de 61 pontos como passo intermediário. Ao concluírem-se estas etapas, a ferramenta computacional

utilizada marcou automaticamente um total de 284 pontos fiduciais nas faces originais, delineando mais características faciais que o conjunto de 46 pontos. A Figura 14 mostra alguns exemplos de faces marcadas semi-automaticamente de acordo com o procedimento acima.



Figura 14 – Exemplos de resultados da marcação semi-automática de 284 pontos fiduciais no conjunto original de faces originais do banco de faces da FEI.

Assim como na Seção 4.2, as imagens de face originais marcadas com 284 pontos fiduciais foram registradas em relação a um atlas de referência, gerando um novo conjunto de imagens de face deformadas pelo processo de normalização não-rígida. Utilizaram-se também estas imagens como entrada para os experimentos de análise multivariada não-supervisionadas e supervisionadas e compararam-se os resultados com aqueles experimentos realizados tendo como entrada os outros conjuntos de faces.

4.4 Construção do Atlas de Referência

Para a execução dos experimentos de registro de imagens de faces, necessita-se primeiramente definir uma atlas de referência, o qual servirá como base para todas as outras imagens a serem registradas. O procedimento utilizado para gerar as imagens de referência é aquele definido na Seção 3.4. Utilizou-se como atlas inicial a face média neutra calculada a partir das faces originais do banco da FEI, marcada como os 46 pontos fiduciais. Executou-se uma primeira iteração pelo procedimento da Figura 11, gerando-se um conjunto de 400 imagens de face deformadas. Calculou-se então a face média deste conjunto e repetiu-se o procedimento

até a estabilização métrica da qualidade do atlas. Para registrar as faces utilizou-se o algoritmo FFD descrito anteriormente e a métrica de similaridade foi formada pela correlação cruzada (CC) e a distância normalizada dos pontos fiduciais (DNP), tal que a função custo C ficou com o seguinte formato:

$$C = DNP - CC + \lambda C_{suavidade}, \quad (4.1)$$

onde determinou-se empiricamente o parâmetro $\lambda = 0.01$ por prover um bom compromisso entre os termos competitivos da função custo.

Como medida de qualidade do atlas, utilizou-se a média das correlações cruzadas entre as imagens registradas em relação ao atlas, ou seja, calculou-se a média da métrica CC após a deformação das imagens, tal que:

$$CC_{médio} = \frac{1}{N} \sum_{i=1}^N CC_i, \quad (4.2)$$

onde N é número de imagens registradas em cada iteração. Calculou-se a correlação média em cada iteração do procedimento da Figura 11, onde também gerou-se o atlas de referência para a iteração seguinte. A Tabela 1 lista os atlas gerados e a média da correlação cruzada. Nestes experimentos utilizaram-se somente as marcações com 46 pontos fiduciais.

Em todas as iterações, calculou-se a correlação cruzada das imagens mantendo as intensidades apenas dentro da região de interesse definida pela envoltória convexa dos pontos fiduciais anotados na imagens. Os pixels internos à envoltória convexa foram mantidos e pixels externos foram definidos com intensidade 0. A Figura 15 apresenta um exemplo de face com apenas a região de interesse preservada. A medida de qualidade foi feita usando apenas a região de interesse tendo em vista que essa região é a mais afetada pela deformação do registro de imagens porque é demarcada pelos pontos fiduciais.

Como pode ser visto na Tabela 1, o $CC_{médio}$ não aumentou significativamente a partir da segunda iteração. As imagens de referência geradas a partir dessa rodada são semelhantes e todas têm o aspecto indefinido em relação ao gênero e a expressão facial. Portanto, utilizou-se nos experimentos seguintes o atlas da iteração 2.



Figura 15 – Região de interesse demarcada pela envoltória convexa dos pontos fiduciais anotados nas imagens de face.

Iteração	Atlas de Referência	$CC_{médio}$
1		0.9668
2		0.9721
3		0.9721
4		0.9723
5		0.9728

Tabela 1 – Imagens de referência usadas para registrar as 400 imagens da base original e as respectivas médias de correlação cruzadas obtidas no processo de registro.

4.5 Geração de Novos Conjuntos de Imagens de Face

A partir da escolha do atlas de referência e usando-se os mesmos parâmetros deste processo no algoritmos FFD, ou seja, $\lambda = 0.01$, combinando *DNP* e $-CC$ como métrica de similaridade, aplicou-se a FFD para registrar as 400 imagens do conjunto original da FEI marcadas com 46 pontos fiduciais, gerando 400 novas imagens deformadas que são chamadas aqui de conjunto 46. Registraram-se também as imagens do conjunto original marcadas com 284 pontos fiduciais, gerando-se um novo conjunto de imagens que são chamadas aqui de conjunto 284. Algumas imagens de face deformadas no processo de geração dos dois conjuntos são apresentadas na Figura 16. Percebe-se a diferença entre os dois processos de registros principalmente nas características internas da face, as quais foram demarcadas de maneira mais detalhada no conjunto 284. Aproveitaram-se os três conjuntos de imagens de face nos experimentos de análises multivariadas subsequentes.



Figura 16 – Exemplos de faces deformadas usando o algoritmo de registro FFD. A linha superior apresenta imagens de face originais, a linha do meio apresenta as mesmas imagens deformadas do conjunto 46 e a linha inferior apresenta, analogamente, as imagens do conjunto 284.

4.6 Resultados Mais Expressivos

As Figuras 17, 18 e 19 descrevem as características mais expressivas dos conjuntos de faces originais, 46 e 284, respectivamente. Para melhorar o contraste das imagens, aplicou-se uma equalização de histograma nos três conjuntos de dados antes de calcular as dimensões estatísticas pelo método PCA. Com exceção da equalização de histograma, todas as faces ilustradas nas figuras foram derivadas matematicamente usando a já mencionada Equação (3.28), onde $i \in \{1, 2, 3, 4, 5, 6, 7, 8\}$ e $j \in \{-3, -2, -1, 0, 1, 2, 3\}$, ou seja, nenhum retoque manual ou processamento de imagem adicional foi aplicado. Claramente, o registro de imagens não-rígido como etapa de pré-processamento eliminou os artefatos tipicamente encontrados nas faces geradas do conjunto de dados original, muito aparentes na Figura 17, mas praticamente inexistentes nas Figuras 18 e 19. Estas figuras de fato ilustram um subconjunto de faces gerados por computador que representam indivíduos inexistentes com realismo promissor.

Nas três figuras, percebe-se que a característica mais expressiva, aquela que leva em conta a maior parte da variância das amostras, corresponde a efeitos de luminosidade, porém, percebe-se também a variação das características de gênero na primeira linha de cada figura. Percebe-se também que na maioria das linhas das três figuras houve variação na expressão facial das faces, de neutra para sorrindo. Outras características menos aparentes variam dependendo da componente navegada. Em alguns casos, nota-se a variação na largura dos lábios, na espessura do nariz, abertura dos olhos e tamanho das sobrancelhas. Estas variações são mais aparentes nas faces geradas a partir dos conjuntos de dados registrados não-rigidamente.



Figura 17 – Navegação nas oito primeiras componentes principais descrevendo as características mais expressivas do conjunto de faces original.



Figura 18 – Navegação nas oito primeiras componentes principais descrevendo as características mais expressivas do conjunto não-rigidamente registrado com 46 pontos fiduciais.



Figura 19 – Navegação nas oito primeiras componentes principais descrevendo as características mais expressivas do conjunto não-rigidamente registrado com 284 pontos fiduciais.

4.7 Resultados Mais Discriminantes

As Figuras 20, 21 e 22 descrevem, respectivamente, as características mais discriminantes dos conjuntos de dados original, registrado com 46 pontos e registrado com 284 pontos fiduciais. Todos as faces das três figuras foram derivadas matematicamente usando a já mencionada equação (3.37), onde $j \in \{-3, -2, -1, 0, 1, 2, 3\}$ e $i \in \{1, 2\}$, ou seja, as amostras foram arranjadas de dois em dois grupos: masculino contra feminino e neutro contra sorrindo. Novamente, aplicou-se previamente uma equalização do histograma em todas as amostras dos três conjuntos de imagens antes de calcular as dimensões estatísticas, mas nenhum retoque ou manipulação das imagens foi aplicado posteriormente. Nestes experimentos também percebeu-se que a etapa de pré-processamento de registro de imagens, tanto com 46 quanto com 284 pontos fiduciais, removeu os artefatos da região facial. Mais importante, percebe-se nas Figuras 21 e 22, da esquerda para a direita, sem sobreposições, como a face média dos conjuntos de dados pode ser modificada para se parecer exclusivamente mais masculina ou mais feminina (linha superior), ou mais neutra para mais sorridente (linha inferior). Curiosamente, como a face média possui alguns traços de sorriso, as faces mais à esquerda das linhas inferiores das Figuras 21 e 22 são exemplos de faces andróginas e sem expressão.



Figura 20 – Características mais discriminantes do conjunto de faces originais, da esquerda para a direita: masculino contra feminino (acima); neutra contra sorrindo (abaixo).

Adicionalmente, investigou-se o efeito da etapa de registro de imagens como pré-processamento no desempenho de classificação da reta de separação das amostras. Aplicou-se o método *leave-one-out* para avaliar o desempenho desse classificador linear (THOMAZ; GIRALDI, 2010), usando os três conjuntos de faces separadamente. Para cada experimento, calculou-se a média de cada classe no sub-espço PCA+MLDA e a distância Euclidiana destas médias foi utilizada para classificar cada observação, ou seja, a amostra observada foi assi-



Figura 21 – Características mais discriminantes do conjunto de faces registradas com 46 pontos fiduciais, da esquerda para a direita: masculino contra feminino (acima); neutra contra sorrindo (abaixo).



Figura 22 – Características mais discriminantes do conjunto de faces com 284 pontos fiduciais, da esquerda para a direita: masculino contra feminino (acima); neutra contra sorrindo (abaixo).

nalada para a classe cuja distância em relação a média foi menor em cada grupo (masculino contra feminino ou neutro contra sorrindo).

A Figura 23 mostra as taxas de reconhecimento *leave-one-out* da reta de separação para as três bases de treinamento, conjunto original de faces rigidamente pré-alinhadas, conjunto não-rigidamente alinhado com 46 pontos fiduciais e conjunto não-rigidamente alinhado com 284 pontos fiduciais. Como pode-se observar, a reta de separação alcançou o desempenho de 94%, 96.75% e 96.25% na classificação de gênero nas três bases respectivamente e, analogamente, 95%, 96.25% e 94.75% na classificação de expressões faciais. Essas taxas de classificação indicam que não houve ganho ou perda significativa no desempenho deste classificador por conta da etapa de registro não-rígido das faces.

Aproveitou-se o treinamento das retas de separação para classificar as amostras geradas nos experimentos não-supervisionados, bem como as amostras geradas nos experimentos supervisionados. As Figuras 24 e 25 mostram respectivamente o resultado da projeção dessas amostras por gênero e expressão facial. Realizou-se este experimento apenas com o conjunto de

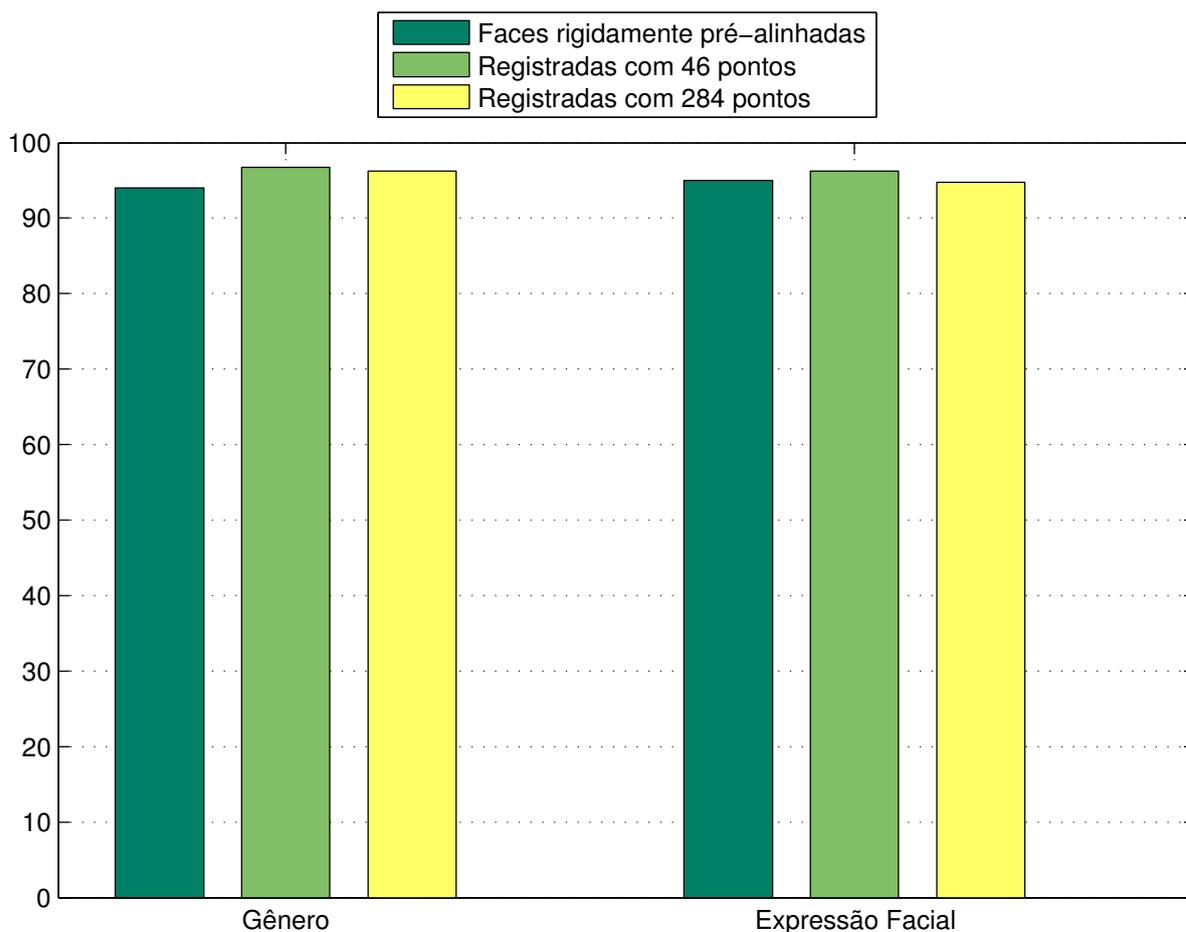


Figura 23 – Taxas de classificação da reta de separação das amostras, gênero da esquerda para a direita: 94%, 96.75% e 96.25%; Expressão Facial; 95%, 96.25% e 94.75%.

faces registradas não-rigidamente usando 284 pontos fiduciais. Em cada gráfico projetaram-se no respectivo vetor discriminante as amostras usadas para o cálculo da reta de separação usando o arcabouço PCA+MLDA, as médias de cada grupo arranjado para treinamento (masculino e feminino, neutro e sorrindo) e as faces geradas artificialmente revelando as características mais expressivas e mais discriminantes.

Conforme visto na seção 3.6, o método MLDA gera uma matriz de projeção com $g-1$ autovetores que, neste caso, corresponde a apenas 1 autovetor. Quando projetou-se cada amostra neste autovetor, obteve-se um escalar. Portanto, o resultado da projeção de todas as amostras é um conjunto de escalares. Contudo, para facilitar a visualização dessas amostras, criou-se uma dimensão arbitrária no eixo vertical dos gráficos das Figuras 24 e 25 correspondente a ordem na qual as amostras foram projetadas. Primeiramente, as amostras de treinamento receberam o valor de 1 a 200, as médias de cada grupo de treinamento receberam o valor de 100, as amostras com as características mais discriminantes receberam o valor 210 e as amostras com as características mais expressivas receberam os valores de 221 a 268.

Percebe-se no gráfico da Figura 24 que as faces geradas pelo método supervisionado foram uniformemente projetadas de acordo com a navegação realizada na direção do vetor discriminante de gênero no espaço de faces. Percebe-se também uma concentração de faces geradas pelo método não-supervisionado na fronteira que separa as amostras, isso condiz com a percepção visual da maioria destas faces, ou seja, em alguns casos não é visualmente trivial dizer se estas faces são masculinas ou femininas, pois as faces preservam traços de ambos os gêneros.

Já no gráfico da Figura 25, as faces geradas pelo método não-supervisionado foram projetadas de maneira mais espalhada, sendo a maioria projetada na região sorrindo e poucas delas aglutinadas na região de fronteira. As demais projeções no vetor discriminante de expressões faciais seguem o comportamento esperado para o experimento.

A fim de verificar se o espalhamento dos grupos de amostras segue uma distribuição Gaussiana, plotou-se cada um dos grupos de amostras em dois tipos de gráficos: Gráfico de Probabilidade Normal e Histograma. O primeiro mostra o quanto os dados da amostra estão alinhados a uma distribuição normal teórica, onde os pontos azuis devem formar uma linha reta aproximada. O segundo mostra a distribuição de frequência dos dados cuja forma deve se aproximar do formato de uma distribuição normal padrão.

Nota-se nos gráficos de probabilidade normal das Figuras 26a, 26c, 27a e 27c que a maioria dos dados está situada entre o primeiro e o quarto quartil e que nessa faixa os dados podem ser bem aproximados por uma linha reta. Percebe-se também que existem muitos pontos fora da curva nas extremidades, principalmente nos gráficos das amostras de expressões: Figuras 27a e 27c. Esses pontos fora da curva podem também ser visualizados nos histogramas das amostras. Percebe-se que os gráficos das amostras de expressões são os que apresentaram maior número desses pontos.

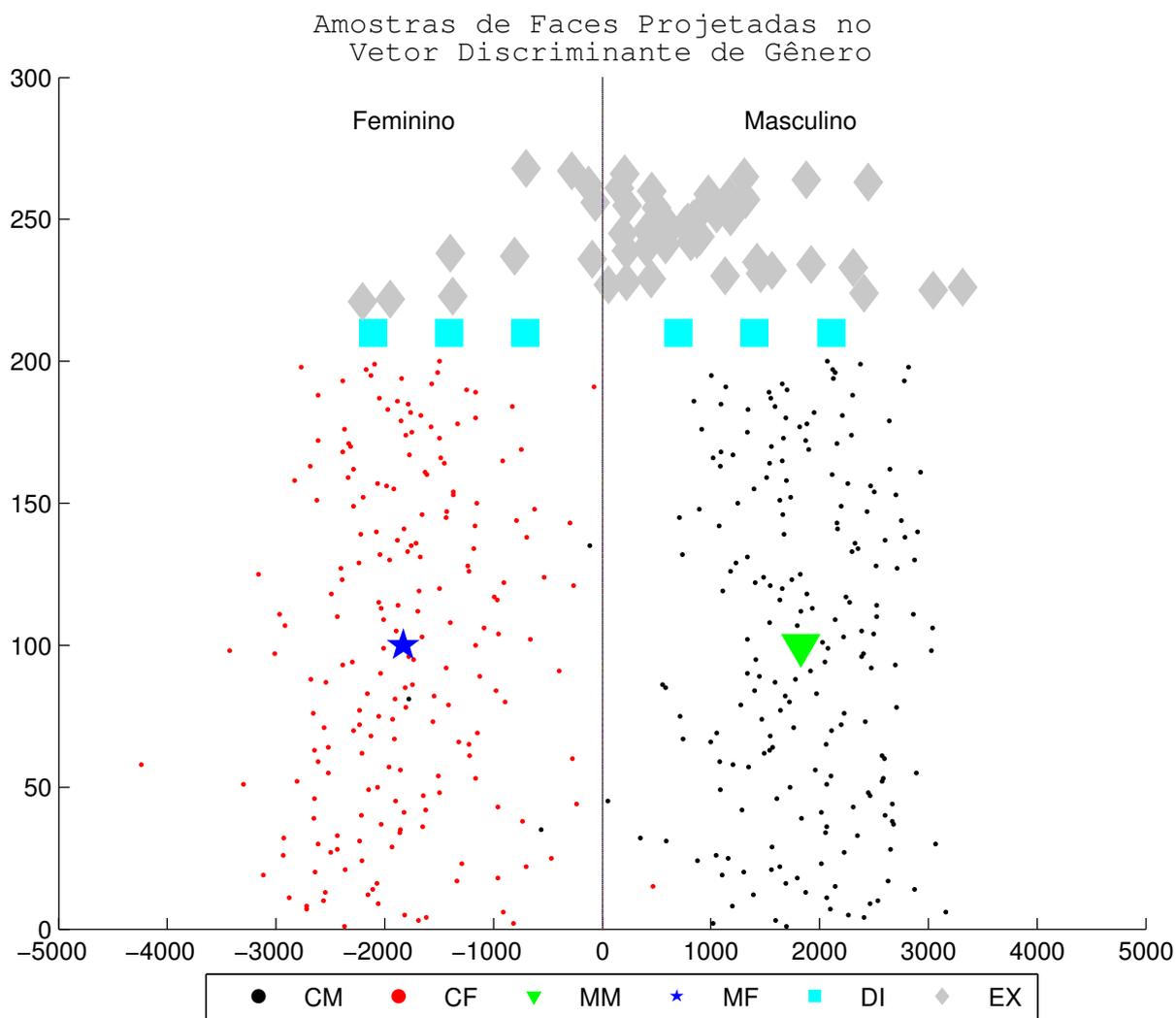


Figura 24 – Projeção das amostras no vetor discriminante de gênero: (CM) - amostras masculinas usadas no cálculo da reta discriminante; (CF) - amostras femininas usadas no cálculo da reta discriminante; (MM) - média das amostras masculinas; (MF) - média das amostras femininas; (DI) - faces geradas artificialmente usando o método supervisionado; (EX) - faces geradas artificialmente pelo método não-supervisionado.

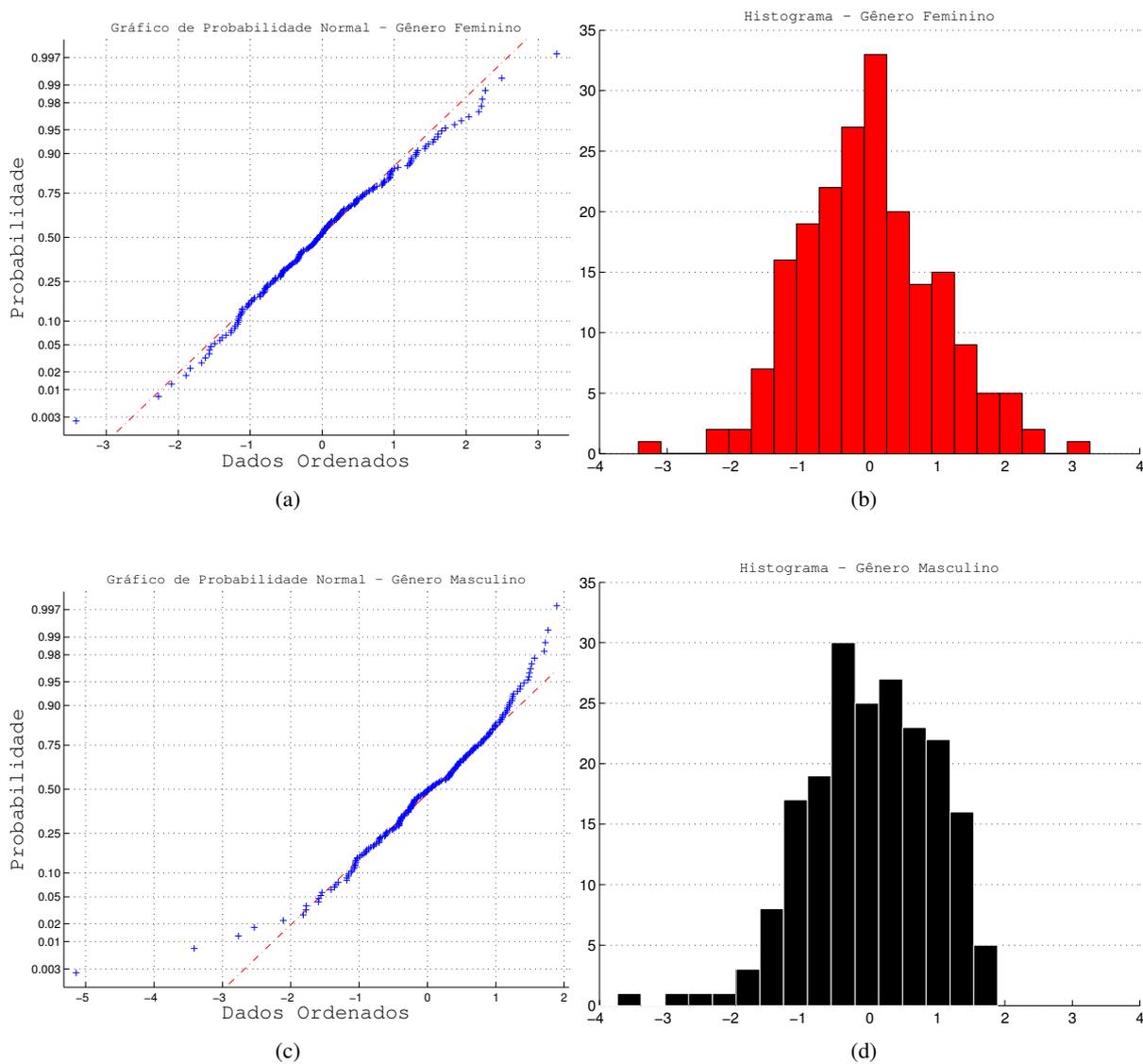


Figura 26 – Gráficos de probabilidade normal e histograma dos grupos de amostras de treinamento de gênero: (a) - gráficos de probabilidade normal das amostras femininas; (b) - histograma das amostras femininas; (c) - gráficos de probabilidade normal das amostras masculinas; (d) - histograma das amostras masculinas.

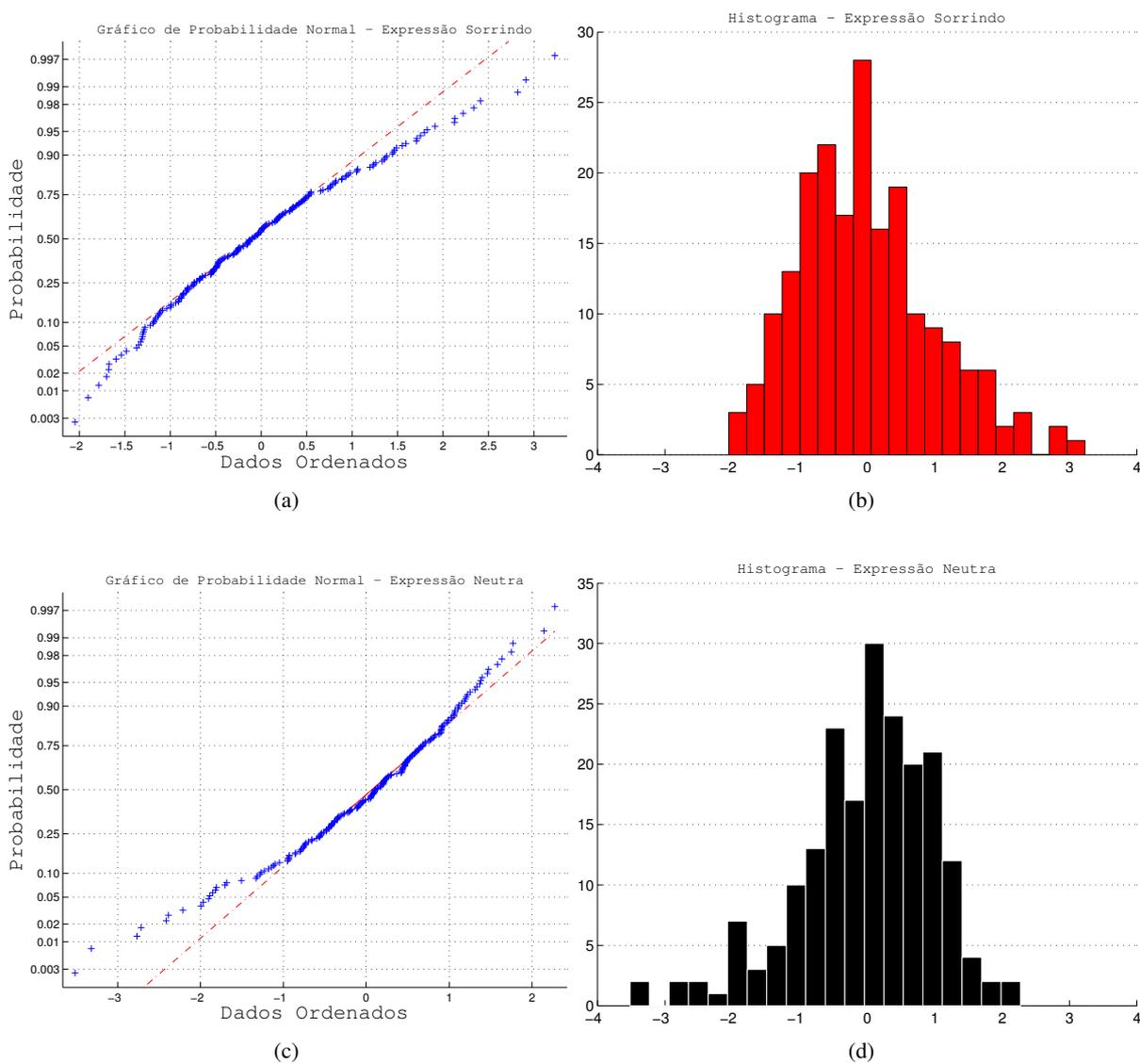


Figura 27 – Gráficos de probabilidade normal e histograma dos grupos de amostras de treinamento de expressão facial: (a) - gráficos de probabilidade normal das amostras sorrindo; (b) - histograma das amostras sorrindo; (c) - gráficos de probabilidade normal das amostras neutras; (d) - histograma das amostras neutras.

5 CONCLUSÃO

Nesta dissertação, explorou-se o alinhamento não-rígido de imagens de faces a um atlas de referência com o objetivo de diminuir os problemas de escala e deformação, e até mesmo de morfologia, entre as amostras de interesse. Por meio de experimentos com a base pública de imagens frontais de face da FEI, composta de 400 amostras, pode-se concluir que a navegação realizada tanto nas direções de maior variação obtidas pelo PCA, quanto nas direções mais discriminantes obtidas pela abordagem PCA+MLDA, produziram-se imagens de face artificiais com realismo promissor. Estes resultados são comparáveis àqueles gerados por outras técnicas de síntese e transformação de faces, tais como em (GIBSON; SOLOMON; BEJARANO, 2003; GIBSON, 2006; TIDDEMAN; STIRRAT; PERRETT, 2005), porém através de uma representação multivariada que engloba forma e textura simultaneamente a partir de grupos de amostras não pares de faces.

Mais especificamente, percebeu-se que normalizar espacialmente as imagens de face em relação a um atlas de referência eliminou a maioria dos artefatos devido ao desalinhamento das características faciais entre as imagens. Mostrou-se, também, que foi possível gerar imagens de face de indivíduos não existentes sem nenhum tipo de retoque ou processamento de imagem adicional. Contudo, estas amostras ainda apresentam uma suavização relativamente atípica da textura facial o que pode levar a uma interpretação mais detalhada de que se tratam, na verdade, de faces geradas artificialmente.

Como esperado, as amostras artificiais geradas pelas componentes principais não garantem variações de um tipo exclusivo de características. Isso foi justificado pela projeção dessas amostras ao longo dos vetores discriminantes e percebe-se, pela sua distribuição, que algumas faces possuem atributos indefinidos por estarem na região da fronteira discriminante. Ou seja, para algumas dessas amostras não se pode afirmar, categoricamente de acordo com o vetor discriminante, se as amostras estão sorrindo ou não, se são masculinas ou femininas.

Por fim, observou-se diferença entre amostras geradas a partir dos conjuntos normalizados não-rigidamente. Tanto nos experimentos não-supervisionados quanto nos experimentos supervisionados, percebeu-se que a inclusão de novos pontos em mais regiões da face contribuiu para criação de faces artificiais com maior resolução em detalhes localizados da face.

Como sugestão para trabalhos futuros, enxerga-se citar as seguintes possíveis ações:

- a) Realizar uma análise multivariada dos vetores de deformação propriamente e aplicar as características mais expressivas e discriminantes desses vetores em imagens de face para

produzir distorções artificiais. Para variações de expressão facial, tal análise pode gerar resultados mais promissores do que aqueles apresentados neste trabalho, investigar métodos de inclusão automática de texturas, tais como, (COOTES; EDWARDS; TAYLOR, 2001) e (TIDDEMAN; BURT; PERRETT, 2001);

- b) Processar outras bases de dados de faces com a abordagem adotada nesta dissertação. Pode-se dar preferência a bases com as faces já anotadas como em (MA; CORRELL; WITTENBRINK, 2015), onde é possível aproveitar as anotações geométricas existentes para marcar as faces automaticamente;
- c) Estender a abordagem deste trabalho para imagens de face 3D. As técnicas e ferramentas computacionais utilizadas neste trabalho não estão restritas a objetos 2D e tal alteração será simples e direta.

REFERÊNCIAS

- AHONEN, T.; HADID, A.; PIETIKÄINEN, M. Face recognition with local binary patterns. In: **Computer vision - ECCV 2004**. Springer, 2004. p. 469–481. Disponível em: <<http://www.masters.dgtu.donetsk.ua/2011/frt/dyrul/library/article8.pdf>>. Acesso em: 5 jun. 2014.
- AMARAL, V. do; THOMAZ, C. E. Normalização espacial de imagens frontais de face. **Periodico científico eletrônico da FATEC São Caetano do Sul (FaSci-Tech)**, 2009. v. 1, n. 1, Outubro 2009. Disponível em: <http://fei.edu.br/~cet/relatorio_tecnico_012008.pdf>. Acesso em: 11 de set. 2014.
- BARTELS, R. H.; BEATTY, J. C.; BARSKY, B. A. **An introduction to splines for use in computer graphics and geometric modeling**. [S.l.]: Morgan Kaufmann, 1987.
- BELHUMEUR, P. N. et al. Localizing parts of faces using a consensus of exemplars. In: IEEE. **2011 IEEE Conference on computer vision and pattern recognition (CVPR)**. 2011. p. 545–552. Disponível em: <<http://dx.doi.org/10.1109/CVPR.2011.5995602>>. Acesso em: 5 jan. 2015.
- BHATIA, K. K. **Analysis of the developing brain using image registration**. Tese (Doutorado) — Imperial College London, March 2007. Disponível em: <<http://www.homes.doc.ic.ac.uk/~dr/pdfs/PhD-thesis-kanwal.pdf>>. Acesso em: 18 nov. 2014.
- CAO, X. et al. Face alignment by explicit shape regression. **International journal of computer vision**, 2014. Springer, v. 107, n. 2, p. 177–190, 2014. Disponível em: <<http://dx.doi.org/10.1007/s11263-013-0667-3>>. Acesso em: 1 mai. 2015.
- COLLIGNON, A. et al. Automated multi-modality image registration based on information theory. In: **Information processing in medical imaging**. [s.n.], 1995. v. 3, n. 6, p. 263–274. Disponível em: <<https://mirc.uzleuven.be/download/public/MIC/publications/929/paper-.pdf>>. Acesso em: 31 jan. 2015.
- COOTES, T. F.; EDWARDS, G. J.; TAYLOR, C. J. Active appearance models. **IEEE Transactions on pattern analysis and machine intelligence**, 2001. IEEE Computer Society, v. 23, n. 6, p. 681–685, 2001. Disponível em: <<https://www.cs.cmu.edu/~efros/courses-/LBMV07/Papers/cootes-pami-01.pdf>>. Acesso em: 18 out, 2014.
- COOTES, T. F. et al. Active shape models-their training and application. **Computer vision and image understanding**, 1995. Elsevier, v. 61, n. 1, p. 38–59, 1995. Disponível em: <http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/Papers/cootes_cviu95.pdf>. Acesso em: 27 abr. 2014.
- COOTES, T. F.; TAYLOR, C. J. et al. **Statistical models of appearance for computer vision**. Technical report, University of Manchester, 2004. Disponível em: <http://www.itu.dk/stud-/projects_f2004/handtracking/referencer/Cootes%20den%20lange%20-%20app_model.pdf>. Acesso em: 2 jun. 2015.
- DEVIJVER, P.; KITTLER, J. **Pattern classification: A Statistical approach**. [S.l.]: Prentice-Hall Englewood Cliffs, NJ, 1982.
- FORSEY, D. R.; BARTELS, R. H. Hierarchical b-spline refinement. In: ACM. **ACM SIGGRAPH computer graphics**. 1988. v. 22, n. 4, p. 205–212. Disponível em: <http://www.cse.iitd.ac.in/~jatin/minor_p/papers/p205-forsey.ps>. Acesso em: 25 mai. 2015.

FRIEDMAN, M. **There's no such thing as a free lunch**. [S.l.]: Open Court La Salle, IL, 1975.

FUKUNAGA, K. **Introduction to statistical pattern recognition**. 2. ed. [S.l.]: Academic press, 1990.

GIBSON, S. J. **EigenFIT: A Statistical learning approach to facial composites**. Tese (Doutorado) — University of Kent, July 2006. Disponível em: <[https://kar.kent.ac.uk/id-eprint/35646](https://kar.kent.ac.uk/id/eprint/35646)>. Acesso em: 26 abr. 2015.

GIBSON, S. J.; SOLOMON, C. J.; BEJARANO, A. P. Synthesis of photographic quality facial composites using evolutionary algorithms. In: **British machine vision conference**. [s.n.], 2003. p. 1–10. Disponível em: <<http://www.bmva.org/bmvc/2003/papers/120-FacialCompKent.pdf>>. Acesso em: 26 abr. 2015.

HAHMANN, S. et al. Volume-preserving ffd for programmable graphics hardware. **The Visual computer**, 2012. Springer, v. 28, n. 3, p. 231–245, 2012. Disponível em: <<http://link.springer.com/content/pdf/10.1007%2Fs00371-011-0608-5.pdf>>. Acesso em: 27 abr. 2014.

JAIN, A. K.; CHANDRASEKARAN, B. 39 dimensionality and sample size considerations in pattern recognition practice. **Handbook of statistics**, 1982. Elsevier, v. 2, p. 835–855, 1982.

JESORSKY, O.; KIRCHBERG, K. J.; FRISCHHOLZ, R. W. Robust face detection using the hausdorff distance. In: SPRINGER. **Audio and video based biometric person authentication**. 2001. p. 90–95. Disponível em: <http://dx.doi.org/10.1007/3-540-45344-X_14>. Acesso em: 5 jan. 2015.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. [S.l.]: Prentice hall Englewood Cliffs, NJ, 2007.

KIRBY, M.; SIROVICH, L. Application of the Karhunen-Loeve procedure for the characterization of human faces. **IEEE Transactions on pattern analysis and machine intelligence**, 1990. IEEE, v. 12, n. 1, p. 103–108, 1990. Disponível em: <<http://dx.doi.org/10.1109/34.41390>>. Acesso em: 7 mai. 2014.

KITANI, E. C.; THOMAZ, C. E.; GILLIES, D. F. A statistical discriminant model for face interpretation and reconstruction. In: IEEE. **19th Brazilian symposium on computer graphics and image processing, 2006. SIBGRAPI'06**. 2006. p. 247–254. Disponível em: <<http://fei.edu.br/~cet/sibgrapi06b.pdf>>. Acesso em: 20 nov. 2014.

KOELSTRA, S.; PANTIC, M. Non-rigid registration using free-form deformations for recognition of facial actions and their temporal dynamics. In: IEEE. **8th IEEE international conference on automatic face & gesture recognition, 2008. FG'08**. 2008. p. 1–8. Disponível em: <<http://dx.doi.org/10.1109/AFGR.2008.4813361>>. Acesso em: 27 abr. 2014.

KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. **The annals of mathematical statistics**, 1951. JSTOR, p. 79–86, 1951. Disponível em: <<http://www.jstor.org/stable/2236703>>. Acesso em: 31 mai. 2015.

LANITIS, A.; TAYLOR, C. J.; COOTES, T. F. Automatic face identification system using flexible appearance models. **Image and vision computing**, 1995. Elsevier, v. 13, n. 5, p. 393–401, 1995. Disponível em: <[http://dx.doi.org/10.1016/0262-8856\(95\)99726-H](http://dx.doi.org/10.1016/0262-8856(95)99726-H)>. Acesso em: 29 abr. 2015.

- LEE, S. et al. Image metamorphosis with scattered feature constraints. **IEEE transactions on visualization and computer graphics**, 1996. IEEE, v. 2, n. 4, p. 337–354, 1996. Disponível em: <<http://dx.doi.org/10.1109/2945.556502>>. Acesso em: 27 abr. 2014.
- LEE, S.; WOLBERG, G.; SHIN, S. Y. Scattered data interpolation with multilevel b-splines. **IEEE transactions on visualization and computer graphics**, 1997. IEEE, v. 3, n. 3, p. 228–244, 1997. Disponível em: <<http://dx.doi.org/10.1109/2945.620490>>. Acesso em: 27 abr. 2014.
- LEE, S.-Y.; CHWA, K.-Y.; SHIN, S. Y. Image metamorphosis using snakes and free-form deformations. In: ACM. **Proceedings of the 22nd annual conference on computer graphics and interactive techniques**. 1995. p. 439–448. Disponível em: <<http://dx.doi.org/10.1145/218380.218501>>. Acesso em: 27 abr. 2014.
- MA, D. S.; CORRELL, J.; WITTENBRINK, B. The chicao face database: A free stimulus set of faces and norming data. **Behavior research methods**, 2015. Springer, p. 1–14, 2015. Disponível em: <<http://link.springer.com/content/pdf/10.3758%2Fs13428-014-0532-5.pdf>>. Acesso em: 10 jun. 2015.
- OLIVEIRA JR, L.; THOMAZ, C. Captura e alinhamento de imagens: Um banco de faces brasileiro. **Relatório de iniciação científica, depto. eng. elétrica da FEI, São Bernardo do Campo, SP**, 2006. v. 10, 2006. Disponível em: <http://fei.edu.br/~cet-/iniciacaocientifica_LeoOliveira_2006.pdf>. Acesso em: 27 abr. 2014.
- RUECKERT, D. et al. Nonrigid registration using free-form deformations: application to breast MR images. **IEEE Transactions on Medical Imaging**, 1999. IEEE, v. 18, n. 8, p. 712–721, 1999. Disponível em: <<http://dx.doi.org/10.1109/42.796284>>. Acesso em: 4 mai. 2014.
- SANDBACH, G. et al. A dynamic approach to the recognition of 3d facial expressions and their temporal models. In: IEEE. **2011 IEEE international conference on automatic face & gesture recognition and workshops (FG 2011)**. 2011. p. 406–413. Disponível em: <<http://dx.doi.org/10.1109/FG.2011.5771434>>. Acesso em: 27 abr. 2014.
- _____. Recognition of 3D facial expression dynamics. **Image and vision computing**, 2012. Elsevier, v. 30, n. 10, p. 762–773, 2012. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0262885612000157>>. Acesso em: 4 mai. 2014.
- SARAGIH, J. M.; LUCEY, S.; COHN, J. F. Deformable model fitting by regularized landmark mean-shift. **International journal of computer vision**, 2011. Springer, v. 91, n. 2, p. 200–215, 2011. Disponível em: <<http://dx.doi.org/10.1007/s11263-010-0380-4>>. Acesso em: 27 abr. 2014.
- SEDERBERG, T. W.; PARRY, S. R. Free-form deformation of solid geometric models. In: ACM. **ACM SIGGRAPH computer graphics**. 1986. v. 20, n. 4, p. 151–160. Disponível em: <<http://doi.acm.org/10.1145/15922.15903>>. Acesso em: 27 abr. 2014.
- SINHA, P. et al. Face recognition by humans: Nineteen results all computer vision researchers should know about. **Proceedings of the IEEE**, 2006. IEEE, v. 94, n. 11, p. 1948–1962, 2006. Disponível em: <<http://dx.doi.org/10.1109/JPROC.2006.884093>>. Acesso em: 9 jun. 2015.
- SONKA, M.; FITZPATRICK, J. M. **Handbook of medical imaging, volume 2. Medical image processing and analysis**. [S.l.]: Academic Press, 2000.

STUDHOLME, C.; HILL, D. L.; HAWKES, D. J. An overlap invariant entropy measure of 3d medical image alignment. **Pattern recognition**, 1999. Elsevier, v. 32, n. 1, p. 71–86, 1999.

SUN, Y.; WANG, X.; TANG, X. Deep convolutional network cascade for facial point detection. In: IEEE. **2013 IEEE conference on computer vision and pattern recognition (CVPR)**. 2013. p. 3476–3483. Disponível em: <<http://dx.doi.org/10.1109/CVPR.2013.446>>. Acesso em: 1 mai. 2015.

SWETS, D. L.; WENG, J. J. Using discriminant eigenfeatures for image retrieval. **IEEE Transactions on pattern analysis and machine intelligence**, 1996. IEEE Computer Society, v. 18, n. 8, p. 831–836, 1996. Disponível em: <<http://doi.ieeecomputersociety.org/10.1109/34.531802>>. Acesso em: 2 jun. 2015.

TENORIO, E. Z.; THOMAZ, C. E. Análise multilinear discriminante de formas frontais de imagens 2d de face. **Proceedings of the X simpósio brasileiro de automação inteligente SBAI 2011**, 2011. p. 266–271, September 2011. Disponível em: <http://fei.edu.br/~cet-/sbai11_ElianaTenorio.pdf>. Acesso em: 27 abr. 2015.

THOMAZ, C. E. et al. A multi-linear discriminant analysis of 2d frontal face images. In: IEEE. **2009 XXII brazilian symposium on computer graphics and image processing (SIBGRAPI)**. 2009. p. 216–223. Disponível em: <http://sibgrapi.sid.inpe.br/col/sid-.inpe.br/sibgrapi@80/2009/09.09.21.33/doc/57160_finalversion.pdf>. Acesso em: 27 abr. 2014.

THOMAZ, C. E.; GILLIES, D. F.; FEITOSA, R. Q. A new covariance estimate for bayesian classifiers in biometric recognition. **IEEE transactions on circuits and systems for video technology**, 2004. IEEE, v. 14, n. 2, p. 214–223, 2004. Disponível em: <<http://fei.edu.br/~cet/csvt04.pdf>>. Acesso em: 26 out. 2014.

THOMAZ, C. E.; GIRALDI, G. A. A new ranking method for principal components analysis and its application to face image analysis. **Image and vision computing**, 2010. Elsevier, v. 28, n. 6, p. 902–913, 2010. Disponível em: <<http://dx.doi.org/10.1016/j.imavis.2009.11.005>>. Acesso em: 27 abr. 2014.

THOMAZ, C. E.; KITANI, E. C.; GILLIES, D. F. A maximum uncertainty lda-based approach for limited sample size problems-with application to face recognition. **Journal of the brazilian computer society**, 2006. SciELO Brasil, v. 12, n. 2, p. 7–18, 2006. Disponível em: <http://fei.edu.br/~cet/jbcs_finalversion.pdf>. Acesso em: 26 out. 2014.

TIDDEMAN, B.; BURT, M.; PERRETT, D. Prototyping and transforming facial textures for perception research. **IEEE computer graphics and applications**, 2001. IEEE, v. 21, n. 5, p. 42–50, 2001. Disponível em: <<http://dx.doi.org/10.1109/38.946630>>. Acesso em: 9 jun. 2015.

TIDDEMAN, B.; PERRETT, D. Moving facial image transformations based on static 2d prototypes. In: SKALA, V. (Ed.). **The 9-th international conference in central europe on computer graphics, visualization and computer vision 2001**. University of West Bohemia, 2001. p. 260–266. Disponível em: <http://wscg.zcu.cz/wscg2001/Papers_2001/R186.pdf>. Acesso em: 9 nov. 2014.

TIDDEMAN, B.; STIRRAT, M.; PERRETT, D. I. Towards realism in facial image transformation: Results of a wavelet mrf method. In: WILEY ONLINE LIBRARY. **Computer graphics forum**. 2005. v. 24, n. 3, p. 449–456. Disponível em: <<http://archive.cs.st-andrews.ac.uk/papers/download/TSP06b.pdf>>. Acesso em: 27 abr. 2014.

TODOROV, A.; OOSTERHOF, N. N. Modeling social perception of faces [social sciences]. **IEEE signal processing magazine**, 2011. IEEE, v. 28, n. 2, p. 117–122, 2011. Disponível em: <<http://dx.doi.org/10.1109/MSP.2010.940006>>. Acesso em: 2 jun. 2015.

TODOROV, A. et al. Understanding evaluation of faces on social dimensions. **Trends in cognitive sciences**, 2008. Elsevier, v. 12, n. 12, p. 455–460, 2008. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1364661308002350>>. Acesso em: 2 jun. 2015.

TURK, M.; PENTLAND, A. Eigenfaces for recognition. **Journal of cognitive neuroscience**, 1991. MIT Press, v. 3, n. 1, p. 71–86, 1991. Disponível em: <<http://www.academia.edu/download/30894770/jcn.pdf>>. Acesso em: 27 abr. 2014.

VALSTAR, M. et al. Facial point detection using boosted regression and graph models. In: IEEE. **2010 IEEE conference on computer vision and pattern recognition (CVPR)**. 2010. p. 2729–2736. Disponível em: <<http://dx.doi.org/10.1109/CVPR.2010.5539996>>. Acesso em: 27 abr. 2014.

VUKADINOVIC, D.; PANTIC, M. Fully automatic facial feature point detection using gabor feature based boosted classifiers. In: IEEE. **2005 IEEE international conference on systems, man and cybernetics**. 2005. v. 2, p. 1692–1698. Disponível em: <<http://dx.doi.org/10.1109/ICSMC.2005.1571392>>. Acesso em: 29 abr. 2015.

XAVIER, I. R. R.; THOMAZ, C. E. Effect of nonrigid alignment using free-forma deformation in frontal face images. **Proceedings of the 10th workshop de visão computacional, WVC'14**, 2014. v. 1, p. 91–96, October 2014.

ZITOVA, B.; FLUSSER, J. Image registration methods: a survey. **Image and vision computing**, 2003. Elsevier, v. 21, n. 11, p. 977–1000, 2003. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0262885603001379>>. Acesso em: 13 mai. 2015.

APÊNDICE A – PCA COM NÚMERO PEQUENO DE AMOSTRAS

Sejam $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ ($N < n$) amostras n -dimensionais de um padrão de interesse. A matriz de covariância S dessas amostras é

$$S_{n \times n} = \frac{1}{(N-1)} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T = \frac{1}{(N-1)} (UU^T)_{n \times n}, \quad (\text{A.1})$$

onde $U = [\mathbf{x}_1 - \bar{\mathbf{x}} \dots \mathbf{x}_N - \bar{\mathbf{x}}]_{n \times N}$.

De acordo com a solução proposta em Fukunaga (1990), nos casos em que o número total de amostras N é menor que a dimensionalidade destas, pode-se optar por calcular os autovalores e autovetores da matriz $(U^T U)_{N \times N}$, tal que

$$\frac{1}{(N-1)} (U^T U)_{N \times N} \Phi_{N \times N} = \Phi_{N \times N} \Lambda_{N \times N}. \quad (\text{A.2})$$

Multiplicando-se ambos os lados da Equação A.2 pela esquerda por $U_{n \times N}$, tem-se

$$\frac{1}{(N-1)} (UU^T)_{n \times n} (U\Phi)_{n \times N} = (U\Phi)_{n \times N} \Lambda_{N \times N}. \quad (\text{A.3})$$

Portanto, $(U\Phi)_{n \times N}$ e $\Lambda_{N \times N}$ são os autovetores e autovalores da matriz de covariância $S_{n \times n}$. Os outros $(n - N)$ autovalores são iguais a zero e seus correspondentes autovetores são indefinidos. A vantagem deste cálculo é que utilizou-se uma matriz $N \times N$ para calcular N autovetores e autovalores, em vez da matriz original $n \times n$.

No entanto, a matriz $(U\Phi)_{n \times N}$ representa um conjunto de vetores ortogonais, mas, não ortonormais, tanto que

$$(U\Phi)^T (U\Phi) \neq I. \quad (\text{A.4})$$

Para obter um conjunto de vetores ortonormais V_i , divide-se cada vetor coluna de $(U\Phi)_{n \times N}$ por $((N-1)\lambda_i)^{1/2}$, tal que

$$V_i = \frac{1}{((N-1)\lambda_i)^{1/2}} U\phi_i \quad (\text{A.5})$$

ou

$$V_{n \times N} = \frac{1}{(N-1)^{1/2}} (U\Phi\Lambda^{-1/2})_{n \times N} \quad (\text{A.6})$$

Pode-se mostrar que o conjunto de vetores $V_{n \times N}$ é ortonormal substituindo a Equação A.2 na Equação A.6 tal que

$$\begin{aligned} V^T V &= \frac{1}{(N-1)} \Lambda^{-1/2} \Phi^T U^T U \Phi \Lambda^{-1/2} \\ &= \Lambda^{-1/2} \Phi^T \Phi \Lambda \Phi^T \Phi \Lambda^{-1/2} \\ &= \Lambda^{-1/2} \Lambda \Lambda^{-1/2} \\ &= I \end{aligned} \tag{A.7}$$