EDSON CAORU KITANI

ANÁLISE DE DISCRIMINANTES LINEARES PARA MODELAGEM E RECONSTRUÇÃO DE IMAGENS DE FACES

Dissertação apresentada ao Centro Universitário da FEI como parte dos requisitos necessários para a obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. Carlos Eduardo Thomaz

São Bernardo do Campo

2007

Kitani, Edson Caoru

Análise de discriminantes lineares para modelagem e reconstrução de imagens de faces. / Edson Caoru Kitani. – São Bernardo do Campo, 2007

165f: il.

Dissertação de Mestrado em Engenharia Elétrica – Centro Universitário da FEI.

Orientador Prof. Dr. Carlos Eduardo Thomaz.

1. Análise de Imagens de Face. 2. Análise de Componentes Principais. 3. Análise de Discriminantes Lineares. 4. Reconhecimento de Padrões. I. Thomaz, Carlos Eduardo, orient. II. Titulo.

CDU 681.32

EDSON CAORU KITANI

ANÁLISE DE DISCRIMINANTES LINEARES PARA MODELAGEM E RECONSTRUÇÃO DE IMAGENS DE FACE

Dissertação de Mestrado

Centro Universitário da FEI

Departamento de Engenharia Elétrica

Banca Examinadora

_____

Prof. Dr. Carlos Eduardo Thomaz (Presidente) - FEI

_____

Prof. Dr. Roberto Cesar Marcondes Junior – IME-USP

_____

Prof. Dr. Paulo Eduardo Santos - FEI

São Bernardo do Campo

12 de Março de 2007

Dedico este trabalho a minha amada esposa Yara e aos meus filhos Adriano e Gizele que acompanharam esta minha longa jornada e mantiveram acesa a chama da esperança nos momentos mais difíceis.

AGRADECIMENTOS

Para que estes agradecimentos descrevam o seu real valor é necessário antes contar a história de como eu iniciei, cursei e conclui o meu mestrado e o papel de todas as pessoas que acompanharam este projeto. No entanto, isto não é possível pois demandaria pelo menos igual número de páginas desta dissertação. Portanto, eu pediria para os que lêem esta página que considerem que os agradecimentos representam muito mais do que escrito. Hoje, eu compreendo melhor a frase "*o conhecimento transforma as pessoas*". Este curso me abriu novos horizontes, me rejuvenesceu (apesar dos novos cabelos brancos) e tornou um sonho em realidade. Mas fundamentalmente, acredito que o agradecimento maior será a possibilidade de eu poder ajudar a transformar a vida de outras pessoas, assim como fizeram aqueles que transformaram a minha.

Ao grande amigo, professor e orientador, na plenitude da palavra, **Prof. Carlos Thomaz** cuja participação neste trabalho foi de vital importância. Pelos ensinamentos, sugestões, correções, opiniões, reuniões, conselhos, e principalmente pela paciência que permitiram a realização desta obra.

Aos professores **Bianchi**, **Paulo** e **Flávio** pelos ensinamentos, apoio, e oportunidades, aos amigos do mestrado, **Nelson, Sergião, Rodolfo, Murilo, Luizão, Julião, Leandro e Marcel**, pelo companheirismo e momentos de descontração ao longo do curso. Ao **Leo** pelo extenso trabalho no Banco de Faces da FEI, sem o qual esta obra não seria possível.

À minha amada esposa **Yara** pela compreensão nas minhas ausências durante mais esta jornada, aos meus filhos **Adriano** e **Gizele**, por compreenderem a minha necessidade de realizar mais este sonho. Aos meus **pais** por compreenderem a minha impossibilidade de visitá-los às vezes aos finais de semana. Aos meus **amigos** e **familiares** por perdoarem as minhas eventuais ausências ou atrasos.

À **MAHLE** por financiar parcialmente este estudo, a todos os **amigos e companheiros** da **MAHLE,** aos meus **superiores**, pelo apoio incondicional e profunda compreensão.

Ao meu amigo **Paulo** da National Instruments por ter cedido uma versão do Labview *Student*, a minha prima **Kelly** por acessar através da USP alguns artigos raros, ao meu amigo **Ricardo** e ao seu colega **Rondinelli**, por também conseguirem alguns artigos do IEEE que muito enriqueceram este trabalho.

Não posso me furtar de agradecer ao **Prof. Dr. Mario Ricci** (INPE), **Prof. Luiz Fiorani** (FEI) e ao **Prof. Rômulo** (UNIA) por acreditarem e me apoiarem na realização deste curso.

À **FEI** por me acolher, abrir este espaço e principalmente por me fazer sentir em casa.

Aos membros da banca, **Prof. Roberto Marcondes** e **Prof. Paulo Santos**, cujas sugestões, questionamentos e críticas durante a qualificação contribuíram muito para melhorar a qualidade deste trabalho.

Finalmente agradeço a **Adriana** e **Rejane** da secretaria do mestrado, aos **voluntários** do Banco de Faces da FEI, ao **pessoal da biblioteca** e **a todos aqueles** que direta ou indiretamente colaboraram para a conclusão desta obra e que não foram diretamente citados. A todos, o meu **muito obrigado!**

"Um bêbado está de joelhos, examinando o chão à volta do poste. Um guarda passa e pergunta: "Que está fazendo ai?"

- Procurando minhas chaves, seu guarda.

- Foi embaixo deste poste que as perdeu?

- Não, seu guarda, foi lá no fim da rua, no escuro.

- Mas então por que está procurando aqui, debaixo da lâmpada.

- Porque aqui tem luz bastante para eu poder vê-las.

Joseph Weizenbaum

...em ciência, não sabemos que as chaves procuradas estão lá longe, no escuro. Não sabemos se as chaves existem. De fato não sabemos nem que o escuro existe (...). Assim, o que procuramos sob a lâmpada do que sabemos não são as chaves, mas uma nova fonte de iluminação."

Ian Stewart

RESUMO

O reconhecimento de faces é uma área de pesquisa que tem recebido grande atenção nos últimos anos, dada a sua abrangência e multidisciplinaridade. Entretanto, apesar dos avanços muitos problemas ainda não foram solucionados mantendo vivo o interesse da comunidade científica nesta área. Fundamentalmente, este trabalho aborda o estudo das imagens de face como um problema de reconhecimento de padrões e investiga o domínio de faces, baseado nas projeções vetoriais dessas faces no hiper-espaço, como um problema de estatística multivariada. A partir desta hipótese, estudam-se quais características visuais são capturadas pelos modelos estatísticos lineares, a capacidade de generalização, e a possibilidade de predizer informações que não necessariamente pertencem a um conjunto de treinamento. Ainda no contexto da estatística multivariada, estudou-se a reconstrução visual dessas informações, cujos resultados comprovaram que um classificador linear pode ser utilizado também para extrair informações e predizer novas. Discute-se ainda o modelo de representação das imagens de faces e como uma alteração no modelo poderia ser transferida para uma imagem de face qualquer, de modo que esta incorporasse as novas informações do modelo. Complementando a pesquisa, desenvolveu-se uma nova interpretação das informações discriminantes fornecidas pelas abordagens de análise de discriminantes lineares, e também um nova forma de interpretação das componentes principais para fins de classificação. Os resultados deste trabalho indicaram o potencial de representação e generalização contidos nas bases vetoriais geradas pelo PCA e pelo classificador baseado no método de Fisher.

Palavras chaves: Análise de Imagens de Faces, Análise de Componentes Principais (PCA), Análise de Discriminantes Lineares (LDA), Reconhecimento de Padrões.

ABSTRACT

Face recognition has motivated several research studies in the last years due to its applicability and multidisciplinary inherent characteristics. Despite the advances achieved so far, several problems related to this topic of research remain challenging keeping the interest of the scientific community in this application high. The aim of this dissertation is to consider the problem of interpreting and reconstructing face images as a pattern recognition task, considering each image as a high dimensional vector and using multivariate statistical techniques. This work has studied which discriminant information can be captured by a linear statistical model, its generalization ability, and whether it is possible to predict statistically differences between sample groups that are not necessarily present in the training sets. In such multivariate statistical context, our experimental results have shown that a linear classifier can be used not only to extract discriminant information from samples but also to predict new ones. Additionally, we discuss in this work a model for face image representation and investigate what occurs with face images when we modify the statistical model, analysing its corresponding variation on a face image. Furthermore, we develop a new interpretation of the discriminant information captured by the linear discriminant classifier and a new interpretation of the principal components in the context of classification. The experimental results carried out in this dissertation indicate the power of representation and generalization described by the PCA and Fisher discriminant bases vectors.

Keywords: Face Images Analysis, Principal Components Analysis (PCA), Linear Discriminant Analysis (LDA), Pattern Recognition.

SUMÁRIO

RESUMO ..... **VII**

ABSTRACT ..... **VIII**

LISTA DE FIGURAS ..... **XII**

LISTA DE TABELAS ..... **XVI**

LISTA DE SÍMBOLOS ..... **XVII**

LISTA DE ABREVIATURAS ..... **XVIII**

1 INTRODUÇÃO **19**

1.1 Objetivos **21**

1.2 Contribuições **22**

1.3 Organização do trabalho **22**

2 TÉCNICAS LINEARES PARA RECONHECIMENTO DE FACES **23**

2.1 Reconhecimento de Faces sob os Aspectos Local e Global **23**

2.2 Técnicas Lineares Baseadas no PCA **25**

2.2.1 Transformada de Karhunen-Loève 26

2.2.2 PCA das Autofaces 27

2.2.3 PCA Probabilístico para Autofaces 28

2.2.4 PCA Bayesiano para Autofaces 30

2.2.5 Outras Abordagens PCA 31

2.3 Técnicas Lineares Baseadas no LDA **33**

2.3.1 LDA Direto 34

2.3.2 LDA Combinado 34

2.3.3 Outras Abordagens PCA+LDA 36

2.3.4 PCA+LDA para Determinação de Gênero, Etnia, Idade e Identidade 37

2.4 Análise de Componentes Independentes **39**

2.5 Modelos Flexíveis de Forma **41**

2.5.1 Modelo Ativo de Forma 42

2.5.2 Modelo Ativo de Aparência 44

2.6 Considerações Finais 47

3 MÉTODOS ESTATÍSTICOS 49

3.1 Representação e Interpretação de Imagens de Face 49

3.1.1 Representação Discreta de uma Imagem de Face 51

3.1.2 Interpretação dos Modelos 54

3.2 Análise de Componentes Principais 57

3.3 Análise de Discriminantes Lineares 63

3.3.1 Problema do Número Pequeno de Amostras 65

3.4 MLDA 66

3.5 Considerações Finais 68

4 MODELO ESTATÍSTICO DISCRIMINANTE 69

4.1 SDM: Um Classificador de Dois Estágios 69

4.2 Comparação entre as Abordagens AAM, Buchala et al. e SDM 78

4.3 Considerações Finais 80

5 EXPERIMENTOS E RESULTADOS 81

5.1 Banco de Faces e Conjuntos de Treinamento 81

5.2 Experimentos com PCA 84

5.2.1 Experimento 1 com o PCA 86

5.2.2 Experimento 2 com o PCA 88

5.2.3 Experimento 3 com o PCA 91

5.2.4 Experimento 4 com o PCA 93

5.2.5 Considerações sobre os Experimentos com o PCA 94

5.3 Experimentos com SDM 95

5.3.1 Experimento 1 com SDM 96

5.3.2 Experimento 2 com SDM 100

5.3.3 Experimento 3 com SDM 103

5.3.4 Experimento 4 com SDM 105

5.3.5 Determinação da Informação mais Discriminante 108

5.3.6 Transferência da Informação mais Discriminante 113

5.4 Considerações Finais 117

6 CONCLUSÃO E TRABALHOS FUTUROS 119

REFERÊNCIAS 122

APÊNDICE A 131

APÊNDICE B 153

REFERÊNCIAS DOS APÊNDICES 156

ARTIGO DO SIBGRAPI 2006 157

LISTA DE FIGURAS

Figura 1.1 - Diagrama de distribuição das abordagens no domínio de faces. 21

Figura 2.1 - Ilusão de Mueller-Lyer. Adaptado de (MACHADO, 1994). 24

Figura 2.2 - Representação da decomposição de uma imagem de face como uma combinação linear de autovetores. Adaptado de (WEN-YI, 2004). 26

Figura 2.3 - Em (a) a representação do espaço de imagens decomposta em dois sub-espaços ortogonais. Em (b) um gráfico espectral dos níveis de energia que cada componente principal representa no espaço de imagens. Adaptado de (MOGHADDAN; WAHID; PENTLAND, 1998). 30

Figura 2.4 - Representação esquemática da computação de imagens de faces. Em (a) o processo tradicional por autofaces, em (b) o método Bayesiano. Adaptado de (MOGHADDAN; JEBARA; PENTLAND, 2000). 31

Figura 2.5 - Diagrama de blocos do modelo combinado PCA+LDA. Adaptado de (ZHAO et al., 1998). 36

Figura 2.6 - Síntese de imagens de face variando-se em (a) a terceira componente principal e em (b) a quarta componente principal. Adaptado de (BUCHALA et al., 2005). 38

Figura 2.7 - No gráfico (a) vemos a representação do espaço vetorial de *pixel* onde a base vetorial são as imagens. No gráfico (b) temos a representação do espaço vetorial de imagens onde a base vetorial são os *pixels*. Adaptado de (BARTLETT; MOVELLAN; SEJNOWSKI,. 2002). 40

Figura 2.8 - Representação da composição e decomposição das imagens de face pelo método ICA. Adaptado de (DUDA et al., 2004). 41

Figura 2.9 - A figura (a) representa o conjunto de treinamento, cujas formas foram obtidas a partir de 72 marcos referenciais. Em (b) temos a reconstrução da formas utilizando–se a equação (2.19) e variando-se os valores do vetor b_f . Adaptado de (COOTES et al., 1994), página. 47. 44

Figura 2.10 - Imagem de face com 122 marcos de referência anotadas manualmente. Adaptado de (COOTES et al., 1998). 45

Figura 2.11 - Exemplo de formas faciais do conjunto de treinamento. Adaptado de (COOTES; TAYLOR, 2004) página 19. 45

Figura 2.12 - Exemplos dos modos de variação em forma (a) e em textura (b) das imagens de faces variando-se o parâmetro γ dentro dos limites de $\pm 3sd$ (*standard deviation*). Adaptado de (COOTES; TAYLOR, 2004) página 32. 47

Figura 3.1 - Na seqüência das figuras (a), (b), (c) e (d) observamos o efeito da sub-amostragem da imagem original (f). Em (e) temos uma representação de como as regiões da face são distribuídas para análise. Adaptado de (YANG; KRIEGMAN; AHUJA, 2002) páginas 36, 37. 50

Figura 3.2 - Representação do espaço de imagens $\mathbf{Z}_x$ 51

Figura 3.3 - Gráfico do vetor de imagem de uma face. No canto superior direito da figura temos a imagem da face que é representada por este gráfico. 52

Figura 3.4 - Gráfico de duas faces distintas. No canto superior direito as imagens de face que representam cada gráfico. 53

Figura 3.5 - Imagem de uma face média. 53

Figura 3.6 - Gráfico representativo na forma de sinal da face média (Figura 3.5) concatenada. A escala *y* está normalizada dentro dos limites de 0 a 1. 54

Figura 3.7 - Na figura (a) temos amostras de faces com expressão sorrindo, variando de um sorriso leve ao mais intenso. Em (b) temos as expressões neutras das amostras de (a).....55

Figura 3.8 - Em (a) temos a representação gráfica de uma imagem de face com expressão neutra, em (b) o gráfico da imagem de face com expressão sorrindo. No canto superior direito, as imagens de face que são representadas pelos gráficos. 56

Figura 3.9 - Representação gráfica de uma base vetorial e da variação um par de pontos distribuídos nessa base. 59

Figura 3.10 - Representação de uma nuvem de pontos em 3D e os eixos encontrados pelo PCA. Adaptado de (OSUMA, 2004a). 60

Figura 3.11 - Intensidade do *SNR* conforme a componente principal. Adaptado de (MOGHADDAN; WAHID; PENTLAND, 1998). 62

Figura 3.12 - Na figura (a) vemos um hiper-plano representado pela direção do vetor w_1 e em (b) um hiper-plano representado pela direção do vetor w_2 . Adaptado de (OSUMA, 2004b). 64

Figura 4.1 - Representação esquemática da fase de treinamento e formação do espaço de características PCA. Adaptado de (THOMAZ et al., 2006). 71

Figura 4.2 - Representação esquemática da fase de treinamento e formação do espaço de características MLDA. Adaptado de (THOMAZ et al., 2006). 73

Figura 4.3 - Representação esquemática do fluxo de dados durante a fase de reconstrução da característica mais discriminante. Adaptado de (THOMAZ et al., 2006). 76

Figura 4.4 - Representação esquemática da fase de incorporação e síntese de características discriminantes em uma imagem de face. Adaptado de (KITANI; THOMAZ; GILLIES, 2006b) 77

Figura 5.1 - Amostra do banco de faces da FEI com os nomes de arquivo. 81

Figura 5.2 - Figura (a) representa a face que foi utilizada como padrão para alinhamento frontal, na figura (b) vemos a foto original com resolução de 640×480 *pixels* e na figura (c) a foto padrão reduzido na resolução para 260×360 *pixels*. Adaptado de (OLIVEIRA JR.; THOMAZ, 2006). 82

Figura 5.3 - Na seqüência da figura (a) até a figura (e) observamos todo o processo de alinhamento manual que foi realizado sobre a foto da figura (a). Adaptado de (OLIVEIRA JR.; THOMAZ, 2006). 83

Figura 5.4 - No grupo "a", temos amostras das 100 faces femininas e 100 masculinas, no grupo "b", amostras das 100 faces masculinas neutras e 100 faces masculinas sorrindo. 83

Figura 5.5 - Gráfico que representa uma densidade de distribuição constante formado pelas duas primeiras componentes principais. Adaptado de (FUKUNAGA, 1990). 85

Figura 5.6 - Reconstruções visuais variando-se as três primeiras componentes principais utilizando-se o conjunto de treinamento do grupo "a". Do alto da figura para baixo e na seqüência, a primeira, segunda e terceira componentes principais. 87

Figura 5.7 - Gráfico tipo *scatter-plot* do grupo "a" homens e mulheres, em relação a primeira e segunda componentes principais (*Principal Components* PC). 88

Figura 5.8 - Gráfico da contribuição das 100 primeiras componentes principais para fins de discriminação do grupo "a". 88

Figura 5.9 - Reconstruções visuais variando-se as três primeiras componentes principais utilizando-se o conjunto de treinamento do grupo "b". Do alto da figura para baixo e na seqüência, a primeira, segunda e terceira componentes principais. 89

Figura 5.10 - Gráfico tipo *scatter-plot* do grupo “b” homens sorrindo e não sorrindo, em relação a primeira e segunda componentes principais. 90

Figura 5.11 - Gráfico da contribuição das 100 primeiras componentes principais para fins de discriminação do grupo “b”. 90

Figura 5.12 - Reconstruções visuais variando-se as três primeiras componentes principais utilizando-se o conjunto de treinamento do grupo “c”. Do alto da figura para baixo e na seqüência, a primeira, segunda e terceira componentes principais. 91

Figura 5.13 - Gráfico tipo *scatter-plot* do grupo “c” homens não sorrindo e mulheres sorrindo, em relação a primeira e segunda componentes principais. 92

Figura 5.14 - Gráfico da contribuição das 100 primeiras componentes principais para fins de discriminação do grupo “c”. 92

Figura 5.15 - Gráfico tipo *scatter-plot* do grupo “d” em relação a primeira e segunda componentes principais...93

Figura 5.16 - Reconstruções visuais variando-se as três primeiras componentes principais utilizando-se o conjunto de treinamento do grupo “d”. Do alto da figura para baixo e na seqüência, a primeira, segunda e terceira componentes principais. 94

Figura 5.17 - Gráfico da contribuição das 100 primeiras componentes principais para fins de discriminação do grupo “d”. 94

Figura 5.18 - Área de distribuição constante definida pelos autovetores PCA e a distribuição de duas classes....96

Figura 5.19 - Histograma da distribuição das projeções de 100 faces femininas e 100 masculinas no espaço MLDA. 97

Figura 5.20 - Reconstrução visual da imagem de face média para o grupo “a”. 97

Figura 5.21 - Gráfico que representa em valores absolutos a projeção das amostras do espaço PCA no espaço MLDA para o grupo “a”. 99

Figura 5.22 - Reconstrução visual da componentes principais de números 19, 24 e 1 do grupo “a”, do alto para baixo. 99

Figura 5.23 - Histograma da distribuição das projeções de 100 faces masculinas não sorrindo e 100 faces masculinas sorrindo, no espaço MLDA. 101

Figura 5.24 - Reconstrução visual da imagem de face média para o grupo “b”. 101

Figura 5.25 - Gráfico que representa em valores absolutos a projeção das amostras do espaço PCA no espaço MLDA para o grupo “b”. 102

Figura 5.26 - Reconstrução visual da componentes principais de números 18, 16 e 22 do grupo “b”, do alto para baixo 102

Figura 5.27 - Histograma da distribuição das projeções de 100 faces masculinas não sorrindo e 100 face femininas sorrindo, no espaço MLDA. 104

Figura 5.28 - Reconstrução visual da imagem de face média para o grupo “c”. 104

Figura 5.29 - Gráfico que representa em valores absolutos a projeção das amostras do espaço PCA no espaço MLDA para o grupo “c”. 105

Figura 5.30 - Reconstrução visual da componentes principais de números 19, 14 e 1 do grupo “c”, do alto para baixo. 105

Figura 5.31 - Histograma da distribuição das projeções de 100 faces masculinas e 100 face femininas, alternadas de 10 em 10, no espaço MLDA. 107

Figura 5.32 - Reconstrução visual da imagem de face média para o grupo “d”. 107

Figura 5.33 - Gráfico que representa em valores absolutos a projeção das amostra do espaço PCA no espaço MLDA para o grupo “d”.....108

Figura 5.34 - Reconstrução visual da componentes principais de números 24, 29 e 74 do grupo “d”, do alto para baixo. 108

Figura 5.35 - Reconstrução da imagem de face média do grupo “a” quando se navega para além dos limites convencionais. 109

Figura 5.36 - No gráfico (a) temos o vetor δ_{img} reconstruído do ponto $y_i^{MLDA} = +1sd$ para o grupo “a” homens e mulheres. No gráfico (b) vemos as taxas de variação do gráfico ao longo do vetor δ_{img} , porém com os valores das taxas ordenadas crescentemente.....111

Figura 5.37 - Histograma dos pesos do vetor δ_{img} 112

Figura 5.38 - Reconstrução da imagem de face média do grupo “b” quando se navega para além dos limites convencionais. 112

Figura 5.39 - Reconstrução da imagem de face média do grupo “c” quando se navega para além dos limites convencionais. 113

Figura 5.40 - Reconstrução da imagem de face média do grupo “d” quando se navega para além dos limites convencionais. 113

Figura 5.41 - Imagem de uma pessoa com expressão neutra. 114

Figura 5.42 - Imagens sintetizadas de uma pessoa quando se navega ao longo do hiperplano discriminante de homens não sorrindo e sorrindo. 115

Figura 5.43 - Imagens sintetizadas de uma pessoa quando se navega ao longo do hiperplano discriminante de homens e mulheres. 118

LISTA DE TABELAS

Tabela 1 - Tabela comparativa entre as abordagens AAM, Buchala et al. e SDM 80

Tabela 2 – Comparativo entre face média e a face transformada (grupo “b”).....114

Tabela 3 – Comparativo entre face média e a face transformada (grupo “b”).....116

Tabela 4 – Comparação entre a face média e a face transformada (grupo “a”).....116

Tabela 5 – Comparação entre a face média e a face transformada (grupo ”a”).....117

LISTA DE SÍMBOLOS

Σ	Matriz de covariância.
Φ	Matriz de autovetores da matriz de covariância Σ.
Λ	Matriz de autovalores da matriz de covariância Σ.
N	Número de exemplos de um conjunto de treinamento.
n	Dimensionalidade do espaço de faces ou características.
C	Classe de faces.
c_i	*i-ésima* face da classe C.
P	Probabilidade discreta.
$\hat{P}$	Probabilidade estimada discreta.
$\|\Delta\|$	Norma de um vetor.
$\Re^n$	Espaço de dimensão n.
F	Sub-espaço de características dos p primeiros autovetores.
$\overline{F}$	Sub-espaço complementar de F com $N - p$ autovetores.
S	Medida de similaridade entre faces.
S_W	Matriz de covariância intra-classes.
S_b	Matriz de covariância inter-classes.
$\mathbf{x}$	Vetor randômico.
X	Vetor.
x_i	*i-ésimo* elemento do vetor X.
p	Dimensionalidade do espaço PCA.
$\mathbf{Z}_\mathbf{x}$	Matriz de faces concatenadas.
$\mathbf{Z}_\mathbf{x}^*$	Matriz de faces concatenadas com média zero.
$\overline{\mathbf{x}}$	Vetor média das faces.
sd	Standard deviation (desvio padrão).

LISTA DE ABREVIATURAS

AdaBoost	Adaptive Boosting.
AAM	Active Appearance Model.
ASM	Active Shape Model.
DIFS	Distance In Face Space.
DKL	Discriminant Karhuren-Loève.
DFFS	Distance From Face Space.
FEI	Fundação Educacional Inaciana.
FERET	FacE REcognition Technology.
ICA	Independent Component Analysis.
IFDA	Inverse Fisher Discriminant Analysis.
IFFACE	Inverse Fisher Face.
LDA	Linear Discriminant Analysis.
MAP	Maximum A Posteriori.
MDF	Most Discriminant Features.
MEF	Most Expressive Features.
ML	Maximum Likelihood.
MLDA	Maximum uncertainty Linear Discriminant Analysis.
MLP	Multi Layer Perceptron.
PCA	Principal Components Analysis.
PCA_S	Principal Components Analysis with Selection.
RGB	Red Green Blue.
SDM	Statistical Discriminant Model.
SSS	Small Sample Size.
SVM	Support Vector Machine.

1 INTRODUÇÃO

O reconhecimento de faces é uma área de pesquisa em visão computacional que tem motivado muitos pesquisadores dada a sua enorme abrangência e complexidade. Abrangência pela sua vasta aplicabilidade, principalmente nas áreas de segurança e monitoramento de pessoas. Em um cenário mundial altamente globalizado, torna-se necessário desenvolver mecanismos de identificação eficazes, robustos e principalmente discretos, capazes de permitir a identificação de indivíduos sem a necessidade de abordagens diretas e muitas vezes constrangedoras (THOMAZ, 1999).

Ainda neste contexto de aplicação abrangente um outro aspecto a ser considerado é o fato de que, entre todos os canais de relacionamento humano, ou seja, voz, audição e tato, o da visão é o que traz a mais rica e completa das informações. É através da visão que podemos inferir sobre o estado emocional das pessoas, visto que as expressões da face e as movimentações corporais transmitem estes estados emocionais. Portanto, um sistema computacional que possa compreender estes estados emocionais, poderia interagir de maneira mais flexível com um usuário humano, sem a frieza dos sistemas atuais (PANTIC; ROTHKRANTZ, 2000), (CESAR JR., 2001). Faces são consideradas elementos biométricos únicos e individualizados, permitindo substituir ou complementar os atuais sistemas de identificação, garantindo uma maior confiabilidade sem a necessidade de sistemas invasivos, como são os sistemas de identificação de digitais e de íris (CAMPOS, 2001), (ZHAO et al., 2000).

A complexidade dos sistema de reconhecimento de faces refere-se aos inúmeros desafios que têm de ser enfrentados, tais como a enorme dimensionalidade dos dados, a segmentação do objeto de interesse (no caso a face) em relação aos objetos de fundo, a localização de uma face em imagens com múltiplas faces, e etc. Problemas relacionados com a escala, cor, iluminação, ângulos, oclusões, mudanças nas expressões faciais, mudanças por envelhecimento, detecção de faces em imagens de vídeo são muito comuns neste tipo de aplicação (YANG; KRIEGMAN; AHUJA, 2002). Busca-se, atualmente, a construção de sistemas de reconhecimento de faces robustos e altamente confiáveis, visto que o ser humano faz estes reconhecimentos de forma natural e rápida. Porém estes mecanismos ainda não estão completamente compreendidos, fazendo com que as pesquisas trilhem por diversos caminhos na busca de uma solução.

Na maioria dos trabalhos baseados em imagens de faces, os pesquisadores procuram meios de representar e interpretar as imagens de faces adequadamente, e assim permitir que sistemas autônomos possam localizar, descrever, e reconhecer faces. Essas pesquisas tomam

diferentes caminhos dependendo de como o pesquisador crê que o processo de visão ocorre em seres humanos. De uma forma geral inicia-se pela escolha de uma abordagem, onde a visão computacional pode ser estudada sob o aspecto da percepção, cognição ou então uma combinação dessas duas abordagens (BURTON; BRUCE; HANCOCK, 1999).

Abordagens que consideram percepção tendem a trabalhar com sistemas que extraem características (*feature based*), utilizando técnicas de reconhecimento de padrões para aprender as características significativas que representam uma imagem de face (TURK; PENTLAND, 1991). No entanto, as abordagens cognitivas partem do conhecimento de um especialista, que ensina um sistema baseado em regras a analisar como reconhecer uma face (*knowledge based*). Em outras palavras, os sistemas cognitivos buscam na imagem características estruturais e geométricas, tais como olhos, nariz, boca, que permitam identificar a presença de uma face e depois reconhecê-la (BRUNELLI; POGGIO, 1993), (MAURO; KUBOVY, 1992) (KOBER; SCHIFFERS; SCHIMIDT, 1994).

Como descrito no parágrafo anterior as abordagens perceptivas utilizam técnicas de reconhecimento de padrões, e muitas dessas técnicas derivam da estatística multivariada. Entende-se como estatística multivariada o estudo de técnicas estatísticas que investigam as relações interdependentes que existem entre duas ou mais variáveis (JOHNSON; WICHERN, 2002). Observa-se ainda que as técnicas de estatística multivariada aplicadas na área de reconhecimento de faces são também conhecidas como máquinas de aprendizado. Esta denominação se deve ao fato de que essas técnicas extraem o conhecimento através de exemplos apresentados a elas, e cujo aprendizado pode ainda ser classificado como *supervisionado* ou *não supervisionado*. Em reconhecimento de padrões, aprendizado supervisionado refere-se às técnicas que adquirem conhecimento baseado em uma estrutura de padrões previamente conhecida, enquanto que no aprendizado não supervisionado as técnicas extraem essas estruturas a partir dos exemplos fornecidos (RUSSEL & NORVIG, 2004). Várias técnicas com estas características são conhecidas atualmente, tais como: *Principal Components Analysis* (PCA), *Linear Discriminant Analysis* (LDA), *Support Vector Machine* (SVM), e Redes Neurais (HAYKIN, 2001), (JAIN; DUIN; MAO, 2000), (DUDA; HART; STORK, 2001).

O diagrama da figura 1.1 ilustra didaticamente uma taxonomia para se estudar as imagens de faces conforme o tipo de problema proposto. Este diagrama apresenta as áreas em que o domínio do problema “faces” recai, dependendo do que se deseja estudar ou resolver. Porém todas as abordagens são de uma certa forma correlacionadas, isto porque se desejamos reconhecer uma face precisamos previamente localizar a face na imagem (FERIS, 2001), (TURK, 2004). Apesar desta dissertação trabalhar no domínio de faces, ela não poderia ser classificada isoladamente como pertencente a qualquer um dos grupos de estudo, ou seja, lo-

calização, reconhecimento de faces ou de expressões vistos no diagrama da figura 1.1. No entanto, consideramos que os estudos e contribuições deste trabalho serão úteis para todos aqueles que investigam qualquer uma das áreas do domínio de faces.

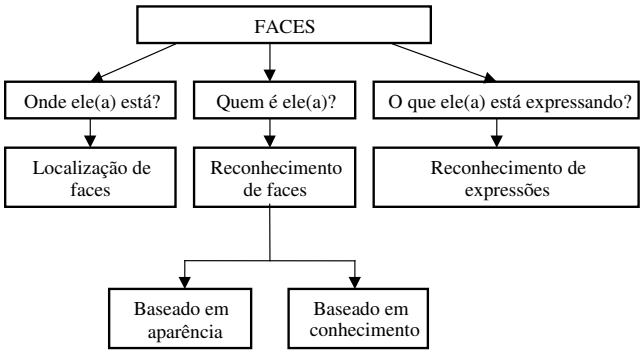


Figura 1.1 - Diagrama de distribuição das abordagens no domínio de faces.

1.1 Objetivos

O objetivo fundamental deste trabalho é investigar como um classificador linear descreve e interpreta as informações que são discriminantes e que ao mesmo tempo são sutis do ponto de vista visual. É proposto um modelo de interpretação dessas informações discriminantes não somente sob o aspecto da classificação mas também como informação para fins de caracterização das diferenças entre dois grupos de treinamento.

Muitos trabalhos já apresentaram a extração de características utilizando técnicas de estatística multivariada, mas com o propósito de se reduzir apenas a dimensionalidade dos dados e criar um espaço de faces computacionalmente[1] manipulável. Neste trabalho veremos como uma técnica de análise de discriminantes lineares, utilizando um novo método de Fisher, não somente pode extrair e classificar características de uma imagem de face como também pode descrever e predizer novas variações dessas características e sintetizá-las visualmente. Considera-se que essas variações representam o grau de generalização do classificador. A síntese visual das imagens de face fornece um conjunto de informações úteis sobre como se distribuem as variações entre dois conjuntos de treinamento e permite determinar visualmente quais são as informações que o classificador considerou como sendo as melhores para fins de classificação.

Este trabalho concentra-se na análise das informações obtidas pelo classificador linear. O reconhecimento de faces ou análise de expressões faciais não são o objetivo primário deste

[1] Refere-se ao consumo de memória e tempo de processamento.

trabalho, mas algumas referências às duas áreas são feitas de modo a facilitar a compreensão desta pesquisa dentro do contexto do domínio de faces e fornecer subsídios para aplicações futuras dos resultados desta pesquisa.

1.2 Contribuições

Como contribuições relevantes deste trabalho pode-se destacar:

- Revisão bibliográfica atualizada sobre os principais trabalhos relacionados com as abordagens que utilizam estatística multivariada em aplicações no domínio de imagens de faces.
- Uma nova interpretação das informações discriminantes fornecidas pelas abordagens baseadas na análise de discriminantes lineares e cujos resultados iniciais foram publicados em (KITANI; THOMAZ; GILLIES, 2006).
- Um modelo de reconstrução e síntese de imagens de faces, para fins de interpretação visual, baseado nas informações mais discriminantes extraídas por um classificador juntamente com uma métrica para medir a taxa de variação das transformações.
- Uma nova interpretação das componentes principais para fins de classificação.
- Uma nova abordagem de síntese de imagens de face a partir da navegação no hiper-plano dos discriminantes lineares para a criação de avatares.

1.3 Organização do trabalho

A estrutura de capítulos desta dissertação está organizada da seguinte forma. No próximo capítulo, capítulo 2, apresenta-se uma revisão bibliográfica dos principais trabalhos já publicados relacionados com a área de reconhecimento de faces e que utilizaram abordagens lineares. No capítulo 3, discute-se uma breve revisão das técnicas de estatística multivariada que serão utilizadas ao longo do trabalho de modo a contextualizar a abordagem nesta dissertação, e sua aplicação na modelagem e síntese de imagens de face. No capítulo 4 descreve-se a abordagem *Statistical Discriminant Model* (SDM) que foi desenvolvida durante as pesquisas para esta dissertação. No capítulo seguinte, capítulo 5, são discutidos os experimentos e resultados obtidos neste trabalho. Finalmente, no capítulo 6, conclui-se a dissertação e discute-se possibilidades de trabalhos futuros.

2 TÉCNICAS LINEARES PARA RECONHECIMENTO DE FACES

2.1 Reconhecimento de Faces sob os Aspectos Local e Global

O reconhecimento de faces exerce um papel muito importante dentro das relações humanas. No contexto social é através do reconhecimento de pessoas que o homem identifica desde pequeno quem são os seus pais, as faces amigas e os grupos sociais as quais pertence (BRUNELLI; POGGIO, 1993). Este mecanismo de busca e reconhecimento de faces nos seres humanos é extremamente robusto, pois opera sobre condições muitas vezes adversas, tais como oclusão, variação na iluminação, múltiplas faces, mudança na expressão e envelhecimento. E o que muitos pesquisadores têm se perguntado é se o sistema de reconhecimento de faces em humanos opera por características locais[2] ou globais. Em seu artigo, Chellappa *et al*. (CHELLAPPA; WILSON; SIROHEY, 1995) sugerem que ambas as características locais e globais podem ser utilizadas.

Um exemplo do uso de informações globais e locais pode ser ilustrado na seguinte pergunta: há algum rosto conhecido na multidão e como poderíamos descrevê-lo? Na primeira parte da resposta, a característica global é a mais utilizada e nos momentos em que temos a impressão de ter encontrado um rosto conhecido fazemos o refinamento com as características locais. Já na descrição de uma face conhecida partimos sempre das características locais e não das globais (CHELLAPPA; WILSON; SIROHEY, 1995).

Muitos pesquisadores concordam que não é possível analisar a visão humana somente sobre os aspectos da percepção, estudando-se somente os mecanismos fisiológicos e muito menos os aspectos cognitivos da representação do que é visto. Chellapa *et al*. (CHELLAPA; WILSON; SIROHEY, 1995) descrevem várias pesquisas na área de neurociência e psicologia relacionadas com a problemática do reconhecimento de visão, e destacam justamente se *"o reconhecimento de faces não seria um processo dedicado"* (CHELLAPA et al., 1995, pág. 710). As seguintes evidências são apresentadas:

1) Faces são mais fáceis de serem lembradas do que objetos.

---

[2] No contexto de imagens de face, características locais são as partes principais que caracterizam a face, tais como olhos, nariz e boca, enquanto que, características globais é o conjuntos de todas as características locais.

2) Pacientes com quadro de Prosopagnosia[3] não conseguem reconhecer faces familiares, e somente conseguem reconhecer pessoas através de outras percepções, tais como: voz, características biométricas e detalhes pessoais.
3) Pesquisas indicam que bebês nascem com o instinto primário de serem atraídos por faces em movimento.

Relativamente à evidência 2, relatos médicos indicam dois tipos de agnosia visual denominadas aperceptivas e associativas. No primeiro caso o paciente não consegue identificar faces conhecidas, apesar de identificar elementos estruturantes como nariz, boca e olhos. Na classe associativa, os pacientes não conseguem relacionar as informações visuais com dados memorizados ou, como descrito por Leme *et al.* (LEME et al., 1999), "*não evocam senso de familiaridade*". Um recente estudo comprova que as regiões cerebrais afetadas pela Prosopagnosia estão relacionadas com a capacidade humana de reconhecer faces e cuja representação é armazenada holisticamente (SCHILTZ; ROSSION, 2006).

Observa-se que não há ainda uma teoria universal que mostre como se processa o reconhecimento de faces em seres humanos e principalmente como se processa o sistema de visão em humanos. Várias hipóteses descrevem seu funcionamento, mas muitos autores concordam que a melhor definição para a visão humana é aquela descrita por Marr (MARR; NISHIHARA, 1977), onde os autores afirmam que "*a visão é uma representação do mundo sobre os aspectos que são relevantes para o observador*."

Esta relevância seria o fato de extrairmos informações de uma imagem, desconsiderando tudo aquilo que não representaria o nosso objeto de interesse. Porém, alguns tipos de ilusão de ótica parecem contradizer isto, pois a nossa percepção parece ser afetada por partes que não seriam relevantes para o nosso interesse. A ilusão de Mueller-Lyer (MACHADO, 1994), ilustrada na figura 2.1, é um exemplo onde observando-se as duas linhas horizontais (superior e inferior) parece-nos que a linha inferior é menor que a superior, apesar de apresentarem a mesma medida.

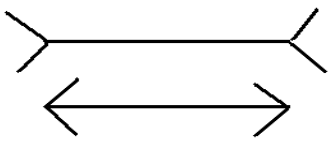


Figura 2.1 - Ilusão de Mueller-Lyer. Adaptado de (MACHADO, 1994).

[3] Lesões cerebrais bilaterais na área ventromedial do giro occipito-temporal que provoca distúrbio no reconhecimento de faces humanas (LEME et al., 1999).

Esta ilusão deve-se a uma influência das linhas oblíquas que fazem com que a nossa percepção seja afetada pelas mesmas, mesmo que a cognição tente dizer o contrário (MACHADO, 1994).

Neste capítulo são abordados os principais trabalhos que estudaram o modelo holístico[4] de faces utilizando técnicas de estatística multivariada tanto para representação quanto para reconhecimento de faces, e que serviram de base para este trabalho. Abordam-se também alguns trabalhos que estudaram métodos para melhorar o desempenho das técnicas estatísticas *Principal Component Analysis* (PCA) e *Linear Discriminant Analysis* (LDA). Nas sub-seções seguintes as principais técnicas estatísticas lineares para reconhecimento de faces serão brevemente discutidas, bem como os trabalhos de Cootes *et al*. que operam com modelos ativos de forma ou *Active Shape Models* (ASM) (COOTES et al., 1994). Apesar do ASM não estar relacionado diretamente com o reconhecimento de faces, eles utilizaram uma técnica estatística que serviu de base para os estudos deste trabalho. Existem no entanto, diversos trabalhos que apontam para outras teorias sobre como se processa o reconhecimento de faces em seres humanos. O leitor interessado em conhecer outras abordagens deverá consultar inicialmente (CHELLAPPA et al., 1995), (ZHAO et al., 2000) e (LI; JAIN, 2005) onde são descritas as várias abordagens de detecção e reconhecimento de faces.

2.2 Técnicas Lineares Baseadas no PCA

Atualmente, considera-se o PCA como a mais antiga e bem sucedida técnica de estatística multivariada. Proposta inicialmente por Karl Pearson (PEARSON, 1901) para estudos sobre dados físicos e biológicos, ele procurava uma representação geométrica mais simples para dados de alta dimensão. Posteriormente Harold Hotteling, em 1933 (HOTTELING, 1933), estudou as representações geométricas de Pearson sob o aspecto matricial e introduziu o termo "*componentes principais*" (JOLLIFFE, 2002).

A aplicação do PCA em imagens de face visa fundamentalmente reduzir a dimensionalidade dos dados, isto porque as imagens discretas são representadas na forma matricial cujas resoluções podem variar muito, dependendo da qualidade da imagem que se deseja estudar. Em uma imagem discreta de 640×480 *pixels*, o total de células dessa matriz é equivalente a 307200. Logo, qualquer análise envolvendo dados dessa magnitude toma um esforço computacional elevado. Isto sem considerar o espaço de $N \times n$ *bytes* de memória que ocupa-

[4] O termo holístico refere-se às abordagens que utilizam toda a imagem de face para fins de representação e análise.

ria uma matriz de faces com essa resolução, onde N representa o número de amostras e n é a dimensionalidade do espaço de imagens. Assim fica clara a necessidade da redução da dimensionalidade para que os dados possam ser manipulados e armazenados de forma mais eficiente. Nas sub-seções seguintes são apresentadas as várias abordagens que utilizaram o PCA no domínio das imagens de faces.

2.2.1 Transformada de Karhunen-Loève

A Transformada de Karhunen-Loève (FUKUNAGA, 1990), que também é conhecida como *Principal Components Analysis* (PCA), teve sua aplicação em imagens de faces proposta inicialmente por (SIROVICH; KIRBY, 1987). A idéia básica apresentada por Sirovich e Kirby era trabalhar a representação das imagens de faces em um sub-espaço de dimensão menor, mas que ainda assim guardasse o máximo de informação da imagem original. Isto é particularmente útil quando se trata de imagens de faces pois, além do problema da alta dimensionalidade, há muita informação redundante. Este sub-espaço é normalmente conhecido como sub-espaço de faces de modo a diferenciá-lo do espaço de imagens. Baseados nesta hipótese os autores apresentaram a proposta de que qualquer imagem de face poderia ser decomposta eficientemente como uma combinação linear de vetores em uma base de menor dimensão. E para se determinar essa nova base eles utilizaram o método da Transformada de Karhunen-Loève para projetar as faces de um conjunto de treinamento em uma base de menor dimensão, mas que ainda retém uma grande quantidade de informações das faces. A figura 2.2 ilustra como uma imagem de face pode ser decomposta em uma base de menor dimensão, onde $\mathbf{x_i}$ é uma imagem do conjunto de treinamento, $\bar{\mathbf{x}}$ é a média de todas as faces do conjunto de treinamento, $(y_1, y_2, ..., y_p)$ é o vetor de pesos da base de menor dimensão e $(\varphi_1, \varphi_2, ..., \varphi_p)$ são os vetores de uma base ortonormal.

Figura 2.2 - Representação da decomposição de uma imagem de face como uma combinação linear de autovetores. Adaptado de (WEN-YI, 2004).

2.2.2 PCA das Autofaces

Depois do estudo pioneiro de Sirovich e Kirby (SIROVICH; KIRBY, 1987), diversos trabalhos posteriores passaram então a utilizar o PCA para a caracterização das imagens de pessoas. O primeiro trabalho publicado utilizando o PCA para o reconhecimento de faces propriamente foi apresentado em 1991 por Mathew Turk e Alex Pentland (TURK; PENTLAND, 1991). Como no trabalho apresentado por Turk e Pentland a base vetorial produzida pelo PCA tinha a mesma dimensão do espaço de imagens e se parecia com faces, os autores batizaram esses autovetores de autofaces, também conhecidos como *eigenfaces* (TURK; PENTLAND, 1991) (LI; JAIN, 2005). Hoje, esse trabalho e esse termo são um dos mais referenciados na área de reconhecimento de padrões e biometria.

A idéia fundamental do trabalho de (TURK; PENTLAND, 1991) foi executar o reconhecimento de faces em um espaço de menor dimensionalidade. O reconhecimento era baseado no princípio da menor distância Euclidiana entre uma face de teste Φ_{teste}, projetada no espaço de faces, com cada face Φ_i do conjunto de treinamento do espaço de faces, conforme descrito na equação abaixo

$$\varepsilon^2 = \left\| \Phi_{teste} - \Phi_i \right\|^2, \tag{2.1}$$

onde $i = 1,2,...,N$, e N é o número de faces do conjunto de treinamento.

O resultado da norma ε é então comparado com um valor de limiar θ_ε de modo a classificar a face de teste como sendo conhecida $(\varepsilon < \theta_\varepsilon)$ ou desconhecida $(\varepsilon > \theta_\varepsilon)$. Ainda dentro do mesmo trabalho, Turk e Pentland apresentaram os resultados do uso do PCA para detecção e localização de imagens de face em imagens de seqüências de vídeo. A localização de uma pessoa nas seqüências de vídeo dependia muito de um fundo de imagem estático. Partindo desta hipótese, a tese seria que somente a pessoa se moveria na imagem. Um filtro diferencial determina a região da imagem ao longo de uma seqüência de quadro onde há um movimento e como conseqüência a presença de uma pessoa. Uma vez localizada a cabeça da pessoa, a região que contém a cabeça é projetada no espaço de faces do PCA e em seguida executa-se o algoritmo de reconhecimento.

A limitação deste sistema reside na necessidade do fundo da imagem (*background*) ser estático, de modo que somente a pessoa nas seqüências de vídeo se movimente. Uma outra limitação se refere às variações de pose, que são muito intensas. Como o conjunto de treinamento era formado somente por imagens de face frontais, não se permitia uma generalização do PCA para reconhecer faces com variações de pose.

Apesar do PCA ter se mostrado eficiente na representação e no reconhecimento de faces, os estudos de (TURK; PENTLAND, 1991) demonstraram também que a técnica PCA era afetada por variações na iluminação, variações na pose, e principalmente por variações na escala. A razão para que o PCA seja muito sensível a variação de escala está no fato de que, mesmo com variações na iluminação, as imagens são muito correlacionadas, porém são pouco correlacionadas quando ocorre mudanças na escala (ZHAO et al., 2000). Isto significa que quando temos imagens de face de uma mesma pessoa, porém com diferentes escalas, o PCA encontra diferentes direções para os vetores que representam cada uma dessas imagens.

2.2.3 PCA Probabilístico para Autofaces

Em (MOGHADDAM; WAHID; PENTLAND, 1998) os autores apresentaram uma abordagem probabilística do PCA para autofaces. A idéia básica é agrupar as variações que ocorrem em imagens de faces de uma mesma pessoa, provocadas pelas variações de pose, iluminação e expressão, e separá-las das imagens de diferentes pessoas. Em outras palavras, um conjunto de imagens de faces pode ser composto por variações na aparência de uma mesma pessoa, provocadas por mudanças nas expressões, formando um grupo intra-pessoal Ω_I, e uma outra contendo as variações na aparência produzidas por diferentes pessoas, formando um grupo extra-pessoal Ω_E.

Consideremos que uma possível medida de similaridade S no espaço de faces pode ser determinada pela norma da distância Euclidiana entre duas imagens de faces (I_1, I_2), tal que $\|I_1 - I_2\|$. Se considerarmos que as imagens I_1 e I_2 são de uma mesma pessoa, a diferença $\Delta = I_1 - I_2$ não leva em consideração as variações nas faces provocadas pelas mudanças nas expressões, pose ou iluminação e poderiam ser consideradas como duas imagens de faces distintas (LI; JAIN, 2005, cap. 7). Entretanto é possível aplicar uma medida de probabilidade *a posteriori* de uma determinada face de teste pertencer ao grupo intra-pessoal Ω_I e ao grupo extra-pessoal Ω_E. Assumindo-se que as distribuições intra-pessoal e extra-pessoal são Gaussianas podemos estimar a densidade de probabilidade $P(\Delta \mid \Omega_I)$ e $P(\Delta \mid \Omega_E)$ para uma certa diferença $\Delta = I_1 - I_2$, e a medida de similaridade *S* pode ser expressa na seguinte forma:

$$S(I_1, I_2) = P(\Delta \in \Omega_I) = P(\Omega_I \mid \Delta). \tag{2.2}$$

A dificuldade, no entanto, reside no fato de que nem sempre temos dados suficientes para computar os parâmetros de densidade de probabilidade. A solução apresentada foi decompor o espaço vetorial $\Re^n$ em dois sub-espaços com auxílio da técnica PCA. O primeiro

sub-espaço foi denominado sub-espaço $F = \{\Phi_i\}_{i=1}^{p}$, onde p representa os p primeiros autovetores associados aos maiores autovalores da matriz de covariância do conjunto de treinamento. O segundo sub-espaço $\overline{F} = \{\Phi_i\}_{i=p+1}^{n}$ representa o restante dos autovetores, considerando-se $p << n$. Consequentemente, uma imagem de face representada como um ponto nesse novo espaço vetorial apresenta uma distância do sub-espaço F, que é definido como DIFS (*Distance In Face Space*), e uma distância em relação ao sub-espaço $\overline{F}$ definido como DFFS (*Distance From Face Space*). A distância DFFS é na realidade o erro de reconstrução provocado pela eliminação dos menores autovalores e os correspondentes autovetores (KITANI; THOMAZ, 2006a), e o DIFS pode ser caracterizado como uma distância de Mahalanobis, uma vez que as imagens de face de uma mesma pessoa são agrupadas (Ω_I)e separadas dos diferentes grupos de pessoas (Ω_E) durante a fase de treinamento. Baseado nesta hipótese, (MOGHADDAN; WAHID; PENTLAND, 1998) utilizaram uma estimativa de densidade marginal $\hat{P}(\Delta \mid \Omega)$ determinada por,

$$\hat{P}(\Delta \mid \Omega) = P_F(\Delta \mid \Omega)\hat{P}_{\overline{F}}(\Delta \mid \Omega), \tag{2.3}$$

onde $P_F(\Delta \mid \Omega)$ é a densidade marginal do espaço F e $\hat{P}_{\overline{F}}(\Delta \mid \Omega)$ é a estimativa da densidade marginal do espaço $\overline{F}$. Desta maneira, dado uma face de teste I, a diferença $\Delta = I - I_i$ é computada para todo $i = 1,2,...,N$, onde N é o número de faces do conjunto de treinamento. A diferença Δ é então projetada nos dois espaços F e $\overline{F}$, e uma medida de classificação, baseada no produto das duas densidades marginais Gaussianas é determinada. O cálculo da equação (2.3) é baseado na seguinte expressão de densidade de probabilidade:

$$\hat{P}(\Delta \mid \Omega) = \left[\frac{\exp\left(-\frac{1}{2}\left(\sum_{i=1}^{p}\frac{\varphi_i^2}{\lambda_i}\right)\right)}{(2\pi)^{p/2}\prod_{i=1}^{p}\lambda_i^{1/2}}\right]\left[\frac{\exp\left(-\frac{\varepsilon^2(\Delta)}{2\rho}\right)}{(2\pi\rho)^{(n-p)/2}}\right], \tag{2.4}$$

onde φ_i e λ_i são os autovetores e autovalores da matriz de covariância do conjunto de treinamento, $\varepsilon(\Delta)$ é o resíduo ou a distância DFFS e ρ é a média dos autovalores relativos aos autovetores do espaço $\overline{F}$, como pode ser visto na equação (2.5) abaixo,

$$\rho = \frac{1}{n-p}\sum_{i=p+1}^{n}\lambda_i. \tag{2.5}$$

A comprovação matemática da equação (2.4) pode ser consultada em (MOGHADDAN; PENTLAND; 1995), onde os autores formulam as bases teóricas do uso

dos autovetores e autovalores para estimar a função densidade de probabilidade de um conjunto de treinamento da base PCA. Na figura 2.3 (a) vemos uma representação gráfica da decomposição do espaço de imagens nos dois sub-espaços F e $\overline{F}$. A figura 2.3 (b) apresenta graficamente a representatividade das componentes principais na formação das imagens de faces e a divisão do espaço F e $\overline{F}$. Observa-se que a fronteira entre os dois espaços é definida pelas p primeiras componentes principais.

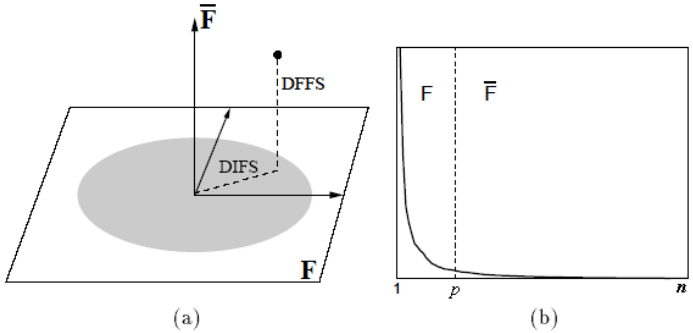


Figura 2.3 - Em (a) a representação do espaço de imagens decomposta em dois sub-espaços ortogonais. Em (b) um gráfico espectral dos níveis de energia que cada componente principal representa no espaço de imagens. Adaptado de (MOGHADDAN; WAHID; PENTLAND, 1998).

2.2.4 PCA Bayesiano para Autofaces

Posteriormente em (MOGHADDAN; JEBARA; PENTLAND, 2000) os autores implementaram um método Bayesiano para o reconhecimento de faces utilizando a densidade de probabilidade calculada pela equação (2.4). Neste método as imagens de teste inicialmente são subtraídas de cada imagem do banco de faces e somente a diferença Δ era projetada em dois sub-espaços de faces denominados intra-pessoal Ω_I e extra-pessoal Ω_E. Depois calcula-se a medida de similaridade S baseado na regra de Bayes, conforme pode ser visto na equação abaixo,

$$S(\Delta) = P(\Omega_I \mid \Delta) = \frac{P(\Delta \mid \Omega_I)P(\Omega_I)}{P(\Delta \mid \Omega_I)P(\Omega_I) + P(\Delta \mid \Omega_E)P(\Omega_E)}. \qquad (2.6)$$

Segundo os autores (MOGHADDAN; JEBARA; PENTLAND, 2000), esta é a primeira abordagem de reconhecimento de faces que não utiliza um discriminador por distância Euclidiana, isto porque a medida de similaridade entre duas imagens de face é determinada pela $P(\Omega_I \mid \Delta)$ e $P(\Omega_E \mid \Delta)$. Desta forma, considera-se que uma imagem de teste pertence a mesma pessoa se $P(\Omega_I \mid \Delta) > P(\Omega_E \mid \Delta)$.

A figura 2.4 (a) ilustra o método tradicional de reconhecimento de faces baseado em (TURK; PENTLAND, 1991), onde as imagens de teste p (*probe*) e treinamento g (*gallery*) são projetadas no espaço de autofaces Ω_u. Nesta abordagem a similaridade $S(p,g)$ é determinada pela norma Euclidiana entre as projeções de p e g no espaço de autofaces, ou seja, $S(p,g) = \|\Delta\|$. A figura 2.4 (b) ilustra o processo Bayesiano de reconhecimento de faces proposto por (MOGHADDAN; JEBARA; PENTLAND, 2000).

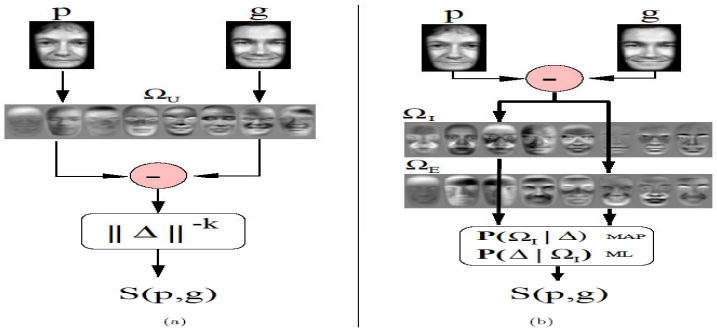


Figura 2.4 - Representação esquemática da computação de imagens de faces. Em (a) o processo tradicional por autofaces, em (b) o método Bayesiano. Adaptado de (MOGHADDAN; JEBARA; PENTLAND, 2000).

A abordagem Bayesina calcula a diferença Δ entre as faces p e g antes de projetá-la no espaço de autofaces, conforme está representado na figura 2.4 (b). Entretanto, no modelo PCA Bayesiano temos dois sub-espaços de autofaces, o primeiro denominado intra-pessoal Ω_I e o segundo denominado extra-pessoal Ω_E. Assim, toma-se a diferença Δ e projeta-se nos dois sub-espaços de autofaces Ω_I e Ω_E, em seguida duas medidas de similaridade podem ser computadas. A partir de cada projeção nos sub-espaços Ω_I e Ω_E determinam-se as probabilidades $P(\Delta \mid \Omega_I)$ e $P(\Delta \mid \Omega_E)$, com as quais então determina-se a máxima probabilidade a *posteriori*, ou *Maximum A Posteriory Probability* (MAP) de $P(\Omega_I \mid \Delta)$ baseado na regra de Bayes, conforme a equação (2.6). Alternativamente é possível determinar uma outra medida de similaridade baseada na probabilidade individual, ou *Maximum Likelyhood* (ML) de $P(\Delta \mid \Omega_I)$. Os autores reportaram uma taxa de reconhecimento de 99% de acerto utilizando-se o banco de faces FERET (PHILLPS et al., 1998).

2.2.5 Outras Abordagens PCA

Diversas abordagens de reconhecimento de faces baseadas em PCA foram apresentadas nos últimos anos, as quais diferem basicamente na forma como os parâmetros do PCA original são trabalhados.

Hiremath e Prabhakar (HIREMATH; PRABHAKAR, 2005) apresentam o estudo sobre uma abordagem denominada PCA Simbólico. A idéia básica é agrupar todas as variações que a imagem de face de uma mesma pessoa sofre, em termos de variação de expressão, iluminação, pose, envelhecimento, em um único vetor de características segundo o seguinte critério

$$x_{ij}^{c} = \frac{\bar{x}_{ij} - \underline{x}_{ij}}{2}, \tag{2.7}$$

onde $\bar{x}_{ij}$ e $\underline{x}_{ij}$ representam respectivamente o maior e o menor valores do *j-ésimo pixel* da *i-ésima* face do grupo de imagens c. Desta maneira, cada nova matriz X^{c} conterá a média da variação de cada *pixel* do grupo de faces *c.* Após a criação de todas as novas matrizes calcula-se o PCA clássico das autofaces. Os autores relatam que o método aumenta a capacidade de generalização do PCA enquanto reduz ainda mais a dimensionalidade porque as primeiras autofaces contêm mais informação do que as autofaces do PCA clássico. Consequentemente a taxa de reconhecimento com poucas autofaces é maior.

Em (JIN et al., 2005) os autores propõem um modelo diferente de PCA Bayesiano em comparação com o modelo apresentado em (MOGHADDAN; JEBARA; PENTLAND, 2000). Considere que temos *c* classes de imagens de faces formando um conjunto $\Omega = \{\omega_1, \omega_2, ..., \omega_c\}$, onde cada matriz ω_i será formada por um conjunto de imagens de faces de uma mesma pessoa. Os autores sugerem calcular os autovetores e autovalores das matrizes de covariância Σ_j para $j = 1,2,...,c$, e então montar uma nova matriz z através de uma transformação linear, tal que

$$z = (\varphi_{11}, \varphi_{12}, ...\varphi_{1p_1}, ...\varphi_{c1}, ..., \varphi_{cp_c})^{T} x, \tag{2.8}$$

onde φ_{ij} são os autovetores de Σ_j e x um vetor do espaço de faces. Como essa nova transformação não garante que a base é ortogonal, os autores aplicaram o Processo de Gram-Schmidt (POOLE, 1995) para construir uma base ortogonal. Desta maneira, a determinação da probabilidade condicional $p(x \mid \omega_j)$ usada para classificar uma face x conforme a equação (2.9)

$$x \rightarrow \omega_i \text{ tal que } i = \arg\max_{j} p(x \mid \omega_j) \tag{2.9}$$

poderia ser substituída pela determinação da probabilidade condicional $p(z \mid \omega_j)$. Portanto a equação (2.9) poderia ser substituída pela equação (2.10), cuja determinação é executada em um espaço de menor dimensionalidade, ou seja,

$$x \rightarrow \omega_i \text{ tal que } i = \arg\max_j p(z \mid \omega_j). \tag{2.10}$$

Os autores fazem uma comparação da abordagem PCA baseada em Bayes com um classificador de mínima distância e relatam que o modelo Bayesiano superou o classificador de mínima distância em todos os testes nos quais o número de imagens de treinamento foi variado.

Por fim, um outro trabalho que investiga em detalhe os aspectos que afetam o desempenho dos algoritmos baseados em PCA, sem considerar nenhuma alteração no modelo clássico, é apresentado em (MOON; PHILLIPS, 2000). Os autores executaram vários experimentos avaliando os aspectos da influência da qualidade das imagens, variações na iluminação, efeitos da compressão das imagens do conjunto de treinamento, número de autovetores retidos e testes com diversos classificadores de distância.

2.3 Técnicas Lineares Baseadas no LDA

Vimos na seção 2.2 que quando aplicamos o PCA em um conjunto de treinamento apenas tratamos com a questão da dimensionalidade, ou seja, extraímos apenas as características mais significativas de cada face, eliminando todas as outras que são comuns ou têm pouca representatividade para explicar cada face. Apesar de termos as informações mais significativas sobre o conjunto de treinamento, não necessariamente temos os dados agrupados por classes, isto porque não necessariamente as componentes principais estarão na direção dos hiperplanos que separam as classes (JOLLIFE, 2002). Se desejarmos determinar uma correspondência linear entre duas ou mais classes de elementos, primeiro precisamos ter as classes, com as quais possamos extrair tais características. Consequentemente, os dados fornecidos pelo PCA não são úteis para uma análise de reconhecimento de padrões, pois os dados estarão espalhados decrescentemente ao longo de uma única super-classe, isto porque o PCA otimiza a extração de características baseado na máxima variância de todas as amostras.

Para resolver este problema, foi proposto por Ronald A. Fisher (FISHER, 1936) um critério estatístico que maximiza a separação entre classes e minimiza o espalhamento dentro das classes, cuja técnica ficou conhecida como *Linear Discriminant Analysis* (LDA). Porém, a aplicação direta do LDA de Fisher em reconhecimento de faces mostrou-se complexa dada as características de alta dimensionalidade dos dados e o baixo número de exemplo do conjunto de treinamento. Portanto a aplicação direta do LDA sempre foi evitada optando-se por mode-

los combinados PCA+LDA (BELHUMER, 1997) ou mesmo alterando-se o modelo LDA original (THOMAZ, 2004).

As sub-seções seguintes descrevem os principais trabalhos que aplicaram a abordagem LDA direto, o modelo combinado PCA+LDA e modificações do modelo LDA original. Finalmente na seção 2.3.4 descrevemos um trabalho que utiliza as abordagens PCA e LDA, cujos objetivos e resultados são muito semelhantes aos que foram obtidos neste trabalho.

2.3.1 LDA Direto

Como descrito no início desta seção, muito autores evitam o uso direto da abordagem LDA em aplicações com faces, pois a dimensionalidade dos dados (n) é muito maior que o número total de exemplos (N). Como, nestes casos, o posto da matriz será sempre $N-c$, onde *c* é o número de classes, a matriz intra-classes S_w será singular (não invertível) impedindo a utilização do LDA padrão (CALLIOLI et al., 1978) (BELHUMER et al., 1997). No entanto alguns autores sugerem abordagens combinadas PCA+LDA ou alterações no LDA padrão com o intuito de resolver este problema. Neste trabalho será aplicado uma abordagem recentemente proposta por (THOMAZ; GILLIES, 2005), que também sugere alterações na matriz S_w com o intuito de estabilizar o cálculo da inversão da matriz. Nas sub-seções seguintes serão discutidos em detalhes as abordagens combinadas e as abordagens que modificam o LDA padrão.

2.3.2 LDA Combinado

A técnica mais usual para se otimizar a extração de informações discriminantes de um conjunto de treinamento é reconhecidamente a análise de discriminantes lineares ou *Linear Discriminant Analysis* (LDA), cujo primeiro trabalho aplicado na área de reconhecimento de faces foi apresentado em (SWETS; WENG, 1996). No trabalho apresentado em (SWETS; WENG, 1996) os autores apresentaram um comparativo de desempenho entre a abordagem da Transformada de Karhunen-Loève, que recebeu a denominação de MEF (*Most Expressive Features*), com a abordagem de análise de discriminantes lineares pelo método de Fisher, que naquele artigo foi denominado MDF (*Most Discriminant Features*).

Como a abordagem MDF era baseada no método de Fisher, e cuja técnica apresenta a conhecida instabilidade provocada pela inversão da matriz S_w, os autores propuseram a construção do sub-espaço linear MDF a partir do sub-espaço MEF. Este novo sub-espaço foi

denominado *Discriminant Karhunen-Loève* (DKL), cuja construção baseava-se nas *p* primeiras componentes principais do espaço MEF, formando um espaço DKL *k*-dimensional, onde *k* segue a restrição $k \leq c-1$, sendo *c* o número de classes do conjunto de treinamento. O número de componentes principais *p* também deve estar restrito a $k+1 \leq c \leq p \leq N-c$, onde *N* representa o número de amostras do conjunto de treinamento. Os autores demonstraram que variações de iluminação e expressão facial não afetavam o desempenho de reconhecimento da abordagem MDF, e que na comparação do MDF e MEF, o MDF sempre superou o MEF.

Posteriormente em (BELHUMER et al., 1997) foi apresentada também uma técnica de reconhecimento de faces baseada no modelo discriminante de Fisher. Em razão disto, os autores denominaram de *Fisherface* o espaço de projeção das imagens de faces. A idéia era maximizar a razão entre as variações intra-classe S_w e inter-classes S_b, assim como foi proposto por (SWETS; WENG, 1996), porém com a restrição do espaço de características vindo do PCA ter $N-c$ componentes principais.

A grande vantagem do LDA sobre o PCA é o fato do LDA executar a extração de características discriminantes com o objetivo de classificar a informação, enquanto que o PCA faz a extração de características com o objetivo de destacar a informação de maior variância. Além do mais, o LDA também promove a redução da dimensionalidade, assim como o PCA, criando um sub-espaço de características discriminantes.

Todavia, em (ZHAO et al., 1998) os autores investigaram o uso do LDA clássico no reconhecimento de faces, e concluíram que o LDA não tem um desempenho satisfatório nas seguintes condições:

1) Quando as faces de teste (*probe*) não são de pessoas do conjunto de treinamento (*gallery*).
2) Quando faces de teste, que pertencem ao conjunto de treinamento, são alteradas artificialmente e depois apresentadas ao LDA.
3) Quando as faces de teste e faces do conjunto de treinamento possuem diferentes imagens de fundo (*background*).

Os autores afirmam que o LDA clássico sofre basicamente de uma baixa capacidade de generalização do conjunto de treinamento, uma vez que em aplicações com imagens de faces raramente o número de faces por classes ultrapassa 4 amostras. Assim, os autores concluem que o modelo combinado PCA+LDA permite resolver o problema da instabilidade da matriz S_w e aumentar o poder de generalização do LDA.

Na figura 2.5 é apresentado um diagrama em blocos da arquitetura combinada PCA+LDA utilizada em muitas das abordagens que trabalham com o PCA+LDA no domínio

de faces. Esta arquitetura baseia-se em duas fases, a primeira fase é formar um conjunto de treinamento, normalizar as imagens em iluminação, escala, rotação e translação, calcular o espaço de projeção PCA, e finalmente, a partir do espaço PCA calcular o espaço LDA. A segunda fase consiste em tomar uma imagem de face de teste e passar pelo processo de normalização, projeção no espaço PCA e no espaço LDA, e finalmente utilizar um discriminador linear através da distância Euclidiana para fornecer o resultado da comparação da imagem de teste com todas as imagens do conjunto de treinamento.

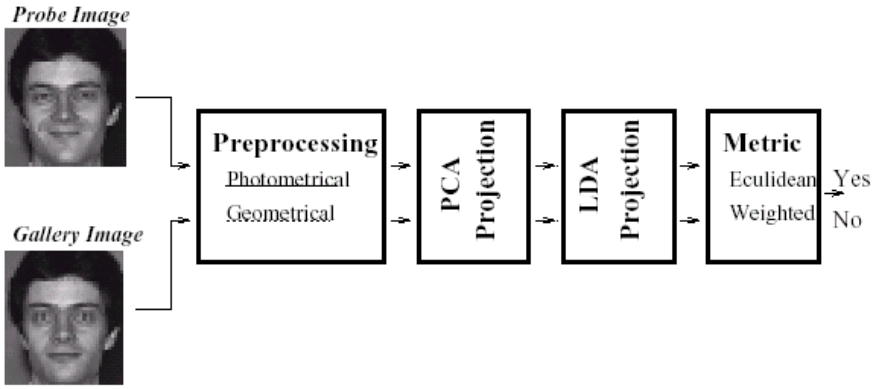


Figura 2.5 - Diagrama de blocos do modelo combinado PCA+LDA. Adaptado de (ZHAO et al., 1998).

2.3.3 Outras Abordagens PCA+LDA

Da mesma maneira que ocorreu com a técnica PCA, foram apresentadas muitas outras abordagens baseadas no LDA, mas que essencialmente alteram a forma como os parâmetros são calculados.

Em (ZHUANG; DAI; YUEN, 2005) os autores discutem as deficiências das abordagens PCA+LDA clássico, e propõem um modelo invertido do LDA (*Inverse Fisher Discriminant Analysis* IFDA), onde ao invés de se calcular o máximo da razão $\det(S_b)/\det(S_w)$, sugerem calcular o mínimo da razão $\det(S_w)/\det(S_b)$. Para que o cálculo seja possível eles também apresentam um novo modelo de PCA, denominado PCA_S, onde os autovetores não são ordenados conforme a energia retida, mas segundo o seguinte critério

$$\begin{aligned} W_{PCA\_S} &= \arg\max\left|W^T \Sigma W\right|, \\ &= [w_1, w_2, ..., w_p], \\ \text{tal que, } & w_i^T S_b w_i > w_i^T S_w w_i. \end{aligned} \tag{2.11}$$

Com esta nova base construída com o PCA, projeta-se nela as matrizes S_b e S_w a fim de se determinar a matriz de transformação do novo espaço. A equação (2.12) mostra o cál-

culo dessa nova base W_{opt}, cujos vetores são denominados pelos autores como *Inverse Fisher Face* (IFFACE).

$$W_{opt} = \arg\min_{w} \frac{\left|W^T W_{proj}^T W_{PCA\_S}^T S_W W_{PCA\_S} W_{proj} W\right|}{\left|W^T W_{proj}^T W_{PCA\_S}^T S_b W_{PCA\_S} W_{proj} W\right|}. \tag{2.12}$$

Segundo os autores esta abordagem permite extrair informações do espaço nulo da matriz S_w, que normalmente é descartado em aplicações convencionais do PCA+LDA. Nos experimentos executados pelos autores, o desempenho do IFFACE supera o *Fisherface* quando o número de exemplos de uma mesma face é maior que 3.

Em (YU; TIAN, 2006) os autores propõem um cálculo da abordagem combinada PCA+LDA que basicamente trabalha com fatores de mistura η e λ de modo a modificar o modelo de Fisher conforme os valores de η e λ. A equação (2.13) mostra como é composta esta nova razão de Fisher. Observa-se que conforme o balanço nos valores de η e λ temos diferentes comportamentos para a equação (2.13). Se $\eta = \lambda = 0$ a equação (2.13) se comporta como o LDA clássico, se $\eta = \lambda = 1$ e equação de comporta como o PCA clássico.

$$W_{opt} = \arg\max_{w} \frac{\left|W^T\left[(1-\lambda)S_b + \lambda\Sigma\right]W\right|}{\left|W^T\left[(1-\eta)S_w + \eta I\right]W\right|}, \tag{2.13}$$

onde *I* é uma matriz identidade.

Como a otimização da equação (2.13) depende dos valores de η e λ e a busca pode ser demorada, os autores implementaram um algoritmo para convergência de classificadores AdaBoost (DUDA; HART; STORK, 2001), de modo a determinar valores ótimos para os parâmetros η e λ. Os resultados apresentados demonstram que o algoritmo denominado *Boosted Hybrid Discriminant Analysis* (B-HDA) apresenta melhor desempenho que as abordagens combinadas PCA+LDA por eles testados.

2.3.4 PCA+LDA para Determinação de Gênero, Etnia, Idade e Identidade

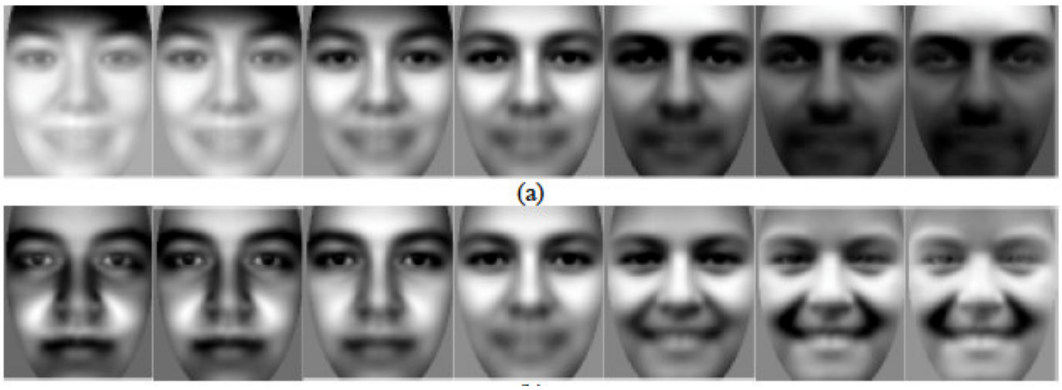
Em (BUCHALA et al., 2005) foi apresentado um trabalho sobre o uso das componentes principais na determinação do gênero, etnia, idade e identidade a partir das imagens de face. Assim como proposto nesta dissertação, Buchala *et al.* (BUCHALA et al., 2005) também investigaram a hipótese de que as componentes principais poderiam ser utilizadas para extrair informações discriminantes, e não somente caracterizar as informações mais expressivas. Trabalhos anteriores já apontavam a possibilidade de as componentes principais descre-

verem variações de sexo e idade (O´TOOLE et al., 1994). No entanto, em (BUCHALA et al., 2005) os autores descreveram e apresentaram os resultados visuais da importância de determinadas componentes principais na classificação das imagens de faces em termos de gênero, idade, etnia e identidade. Utilizando o banco de faces FERET (PHILLIPS et al., 1998) os autores investigaram o que as componentes principais efetivamente representavam em termos de informação visual e quais eram mais importantes para fins de caracterização de gênero, idade, etnia e identidade.

Inicialmente o PCA foi aplicado para construir a base vetorial do conjunto de treinamento e o sub-espaço de faces. Em seguida o LDA foi aplicado de uma forma não convencional para se determinar um índice de importância de cada componente principal na determinação de cada uma das características estudadas. A estimação da importância de cada componente principal foi calculada baseado na seguinte equação,

$$P_i = \frac{\left(Diag\{S_b\}\right)_i}{\left(Diag\{S_w\}\right)_i}, \qquad (2\ 14)$$

onde S_b é a matriz de covariância inter-classes, S_w a matriz de covariância intra-classes, P_i é a medida do grau de discriminação da *i-ésima* componente principal, e $i = 1,2,...,p$.



(b)

Figura 2.6 - Síntese de imagens de face variando-se em (a) a terceira componente principal e em (b) a quarta componente principal. Adaptado de (BUCHALA et al., 2005).

A figura 2.6 (a) apresenta a reconstrução de uma face média variando-se a terceira componente principal dentro dos limites de $\pm 6sd$ (*standard deviation*). Observa-se que a reconstrução mostra que a terceira componente principal modela o gênero do conjunto de treinamento. Na figura 2.6 (b) variou-se a quarta componente principal dentro dos limites de $\pm 6sd$ (*standard deviation*). Neste caso, a quarta componente principal modelou principalmente a expressão sorrindo/não sorrindo. Os autores enfatizam que as componentes principais de menor peso modelam principalmente a identidade e que algumas componentes principais,

como a segunda, podem modelar mais de uma informação, visto que o índice de representatividade aparece alto em todas as características estudadas (etnia, gênero, idade e identidade).

2.4 Análise de Componentes Independentes

Em (BARTLETT; SEJNOWISK, 1997) foi apresentado o reconhecimento de faces baseado na aplicação da técnica *Independent Component Analysis* (ICA). Fundamentalmente, o ICA, assim como o PCA, faz a transformação linear de um espaço de alta dimensão para uma outra de menor dimensão, contudo faz isto procurando minimizar a correlação de alta ordem e maximizar a variância de alta ordem que existem entre os *pixels* de uma imagem de face. O PCA faz esta minimização e maximização em espaços de segunda ordem, ou seja, faz uso da matriz de covariância. Talvez a diferença mais significativa entre o PCA e o ICA seja o fato da base vetorial do ICA não ser necessariamente ortogonal. Uma outra característica do ICA é que esta técnica possui dois modos de operação distintos, denominados arquitetura I e arquitetura II (DRAPER et al., 2003).

Na arquitetura I cada imagem de face forma uma linha da matriz de entrada $\mathbf{X}$. E a matriz $\mathbf{X}$ é considerada a base vetorial do espaço de *pixel* . Explicando melhor, o ICA arquitetura I considera que uma matriz de *pixels* $\mathbf{U}$ no espaço de *pixels* é a combinação linear de cada *pixel* de cada imagem de entrada pela matriz de pesos $\mathbf{W}$. E a matriz $\mathbf{U}$ é considerada a fonte independente de variações do espaço de faces. Da mesma maneira, uma imagem de face x_i também pode ser escrita como uma combinação linear da matriz de pesos $\mathbf{W}$ com um vetor de faces independentes $\mathbf{U}$. Portanto o ICA procura uma matriz de pesos $\mathbf{W}$ que transforme as imagens de face de entrada em imagens de base estatisticamente independentes. A aplicação do ICA arquitetura I em imagens de faces busca destacar as características locais, ou seja, exatamente aquelas características que mais influenciam uma determinada região da imagem.

Na arquitetura II o ICA considera que os *pixels* da matriz $\mathbf{X}$ formam a base vetorial do espaço de imagens. Assim a matriz de componentes independentes $\mathbf{X}$ é agora formada pela combinação linear de vetores de pesos $\mathbf{W}$ e pelas variações dos *pixels* das imagens de face $\mathbf{X}$. Diferentemente da arquitetura I, a arquitetura II trabalha sobre as características globais das imagens de face. Portanto, como descrito por (DRAPER et al., 2003), antes de se definir qual das arquiteturas usar é necessário determinar que tipo de características se deseja investigar.

Na figura 2.7 temos representado graficamente a diferença entre as arquiteturas I e II. Observa-se que na a arquitetura I (a), as imagens formam a base vetorial para se representar

cada *pixel*, enquanto que na arquitetura II (b) os *pixels* formam a base vetorial para se representar cada imagem. Em outras palavras, na arquitetura I, os *pixels* são formados pela combinação linear das imagens enquanto que, na arquitetura II, as imagens são formadas pela combinação linear dos *pixels.*

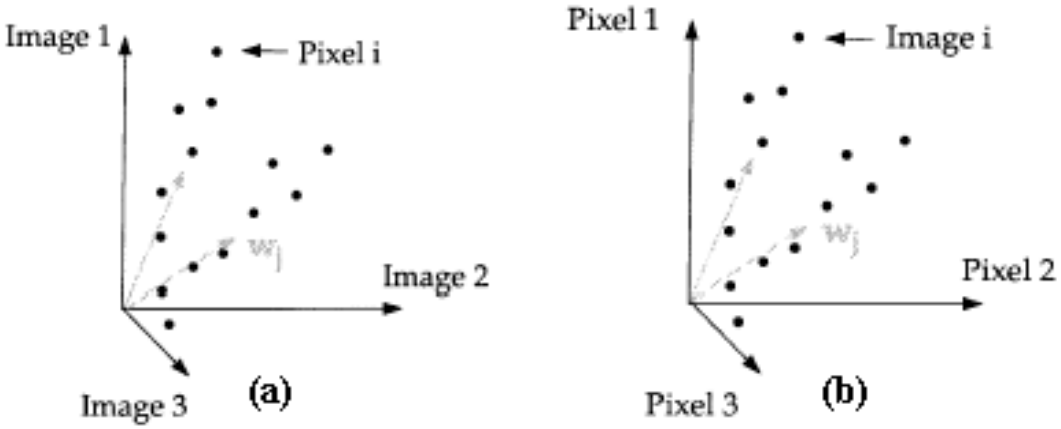


Figura 2.7 - No gráfico (a) vemos a representação do espaço vetorial de *pixel* onde a base vetorial são as imagens. No gráfico (b) temos a representação do espaço vetorial de imagens onde a base vetorial são os *pixels*. Adaptado de (BARTLETT; MOVELLAN; SEJNOWSKI,. 2002).

No trabalho publicado por (BARTLETT; SEJNOWISK, 1997) foram apresentados os resultados comparativos do desempenho de reconhecimento de faces entre as abordagens PCA e ICA. Os autores concluíram que o ICA superou o PCA nos dois conjuntos de testes, com variação de pose e variação de iluminação. No entanto, (DRAPER et al., 2003) discordam em comparar diretamente as abordagens PCA e ICA, uma vez que o ICA tem diferentes parâmetros e algoritmos que devem ser ajustados para cada aplicação.

A figura 2.8 apresenta de forma didática como o ICA considera a formação de uma imagem de face. O ICA considera que uma imagem de face é uma combinação linear de sinais básicos com uma matriz de pesos **U**. De forma análoga, os sinais básicos podem se encontrados decompondo-se as imagens de faces a partir de uma matriz de pesos ***W***. Apesar do ICA poder trabalhar diretamente no espaço de imagens, muitos pesquisadores (BARTLETT; MOVELLAN; SEJNOWSKI,. 2002) têm optado por trabalhar em um espaço de dimensões menor, como aquele produzido pelo PCA. O objetivo principal de se executar um pré-processamento com o PCA é a redução do esforço computacional.

Esta última abordagem (ICA) encerra a revisão relativa às abordagens estatísticas lineares. Na próxima sub-seção abordaremos dois trabalhos que utilizaram a técnica PCA não para fins de reconhecimento de face, mas para explorar como utilizar as componentes principais para modelar as variações de um modelo capturadas pelo PCA. As pesquisas produzidas pelos autores também serviram de referência para esta dissertação.

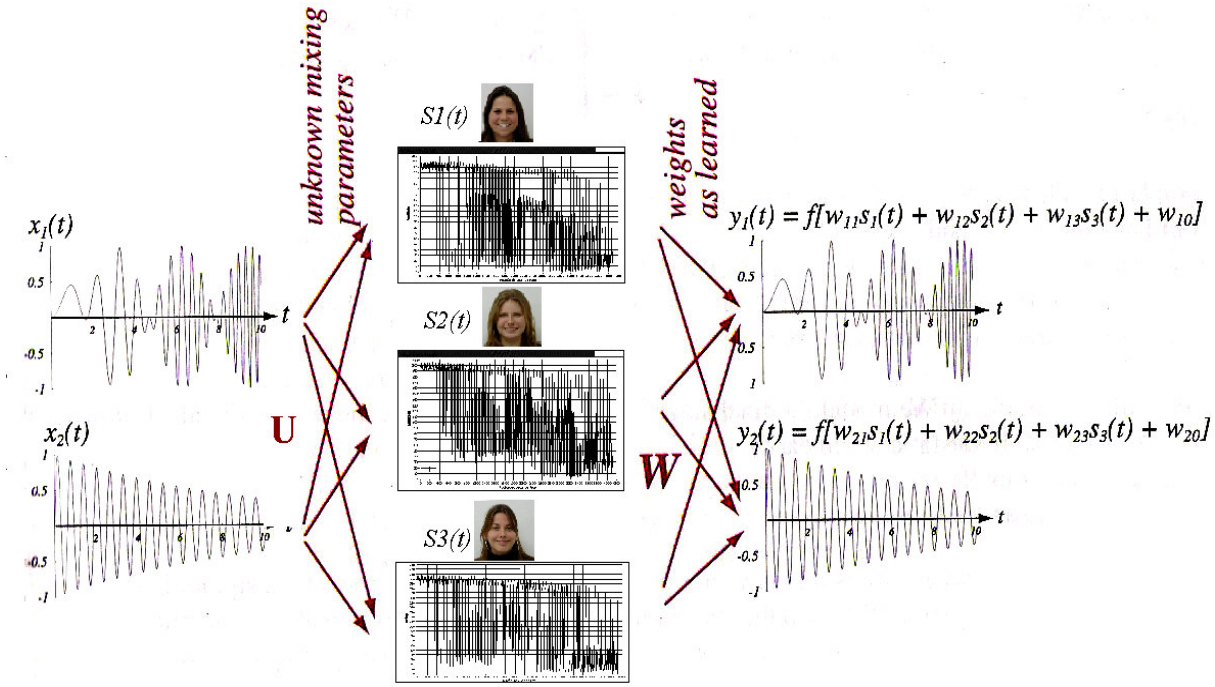


Figura 2.8 - Representação da composição e decomposição das imagens de face pelo método ICA. Adaptado de (DUDA et al., 2004).

2.5 Modelos Flexíveis de Forma

Apesar dos modelos flexíveis de forma não necessariamente estarem relacionados com o reconhecimento de faces, o trabalho apresentado por Cootes *et al.* em (COOTES et al., 1994) demonstraram o uso da técnica PCA para se estudar as variabilidades das formas de objetos 3D através de observações de imagens estáticas sob diferentes pontos de vista. Posteriormente em (LANITIS et al., 1995) foi apresentado um trabalho de reconhecimento de faces utilizando a técnica de modelos ativos de forma e textura. Os resultados do reconhecimento não foram melhores do que outras técnicas lineares e ainda, em contrapartida, o tempo de reconhecimento era muito alto.

No entanto, as técnicas de modelos flexíveis de forma demonstraram ser úteis para modelar e predizer formas que não necessariamente pertencem ao conjunto de treinamento. Isto é equivalente a dizer que as abordagens modelam e apresentam visualmente variações de forma 3D intermediárias a partir de exemplos aprendidos. Nas sub-seções seguintes são descritas a abordagem *Active Shape Model* (ASM) (COOTES et al., 1994), que modela forma, e a abordagem *Active Appearence Model* (AAM) (LANITIS et al., 1995), que modela forma e textura.

2.5.1 Modelo Ativo de Forma

Conforme descrito anteriormente, apesar de o PCA capturar as características e variações mais significativas de um conjunto de treinamento, não há garantias de que as informações capturadas representem exatamente a informação que desejamos. Por exemplo, se desejarmos que o PCA capture as variações da forma de um objeto, somente com o PCA não teremos sucesso, isto porque as variações da formas podem ser tão sutis que uma variação de iluminação nas imagens pode ser considerada como mais significativa pelo PCA.

Para superar este problema em (COOTES et al., 1994) os autores propuseram uma abordagem denominada *Active Shape Models* (ASM), onde os autores capturaram as variações que imagens de objetos em 3D poderiam produzir, e depois modelaram e apresentaram visualmente as variações intermediárias das formas que não necessariamente foram capturadas. Em outras palavras, eles conseguiram produzir e apresentar visualmente uma modelagem contínua das formas a partir de um conjunto de imagens estáticas.

A idéia fundamental da metodologia ASM é de que pode haver uma correlação estatística entre pontos equivalentes de duas formas distintas e que pertencem ao mesmo tipo de objeto. A metodologia ASM parte do princípio de que a forma de um objeto é invariante a transformações Euclidianas (translação, rotação, escala), portanto, se estudarmos as variações da forma dissociadas das transformações Euclidianas, poderemos modelar estatisticamente as variações que a forma sofre. Em outras palavras, o que se deseja é um modelo que represente a forma de um objeto e todas as suas variações típicas (COOTES et al., 1994).

Fundamentalmente, a abordagem ASM trabalha com formas dos objetos mapeados no plano 2D, logo qualquer forma de um objeto pode ser descrita também como um vetor de *m*-pares de pontos (x, y) que contornam a forma do objeto, como pode ser visto na equação abaixo

$$x = (x_1,...,x_m, y_1,..., y_m). \tag{2.15}$$

A equação (2.15) exprime a relação dos pares de coordenadas (x, y) para apenas uma forma. Se ao contrário, tivermos mais exemplos do objeto de estudo, teremos então um vetor $\mathbf{x}_i$ para cada diferente forma $(i = 1,2,...,N)$, formando assim uma matriz de dados $\mathbf{X}$. Se considerarmos que um determinado ponto (x, y) de uma imagem $\mathbf{x}_i$ está correlacionado com os outros pares (x, y) de todas as imagens da matriz de formas $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2,..., \mathbf{x}_N)$, então podemos aplicar a análise de componentes principais (PCA) para determinar a variância do conjunto de formas em $\mathbf{X}$.

Porém, antes da matriz $\mathbf{X}$ ser calculada pelo PCA, é necessário um alinhamento das formas em escala, rotação e posição, isto porque o que se deseja conhecer é como se distribuem os pontos em $\Re^n$ em relação a uma forma base. Se não alinharmos corretamente o conjunto de treinamento, capturaremos variações que não necessariamente representam a variabilidade da forma. Portanto o que se deseja com o alinhamento é minimizar o erro quadrático entre pontos equivalentes de diferentes formas. Em (COOTES et al., 1994) o leitor encontrará uma descrição detalhada sobre o alinhamento de pares de formas pelo método de Procrustes (GOODALL, 1991).

Como desejamos estudar as variações que as diferentes formas produzem, devemos trabalhar a partir de um referencial, desta maneira devemos remover a forma média global $\bar{x}$ da matriz $\mathbf{X}$. Calcula-se antes a matriz de covariância Σ da matriz $\mathbf{X}$ conforme a equação abaixo,

$$\Sigma = \frac{1}{N}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T, \tag{2.16}$$

onde

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i. \tag{2.17}$$

Assim, como descrito por Fukunaga (FUKUNAGA, 1990), para calcularmos os autovalores e autovetores de Σ, devemos resolver a seguinte equação:

$$\Sigma\Phi = \Phi\Lambda. \tag{2.18}$$

Ao determinarmos os autovetores e autovalores de Σ, poderemos ordenar os autovetores Φ conforme os maiores autovalores de Λ. Fazendo isto, ordenamos as projeções de Σ conforme sua importância na representação das variações da forma do conjunto $\mathbf{X}$. Desta maneira, qualquer novo exemplo de forma do conjunto $\mathbf{X}$ pode ser representado pela equação a seguir,

$$x = \bar{\mathbf{x}} + \Phi_f b_f, \tag{2.19}$$

onde Φ_f representa a matriz de autovetores das formas e b_f o vetor de parâmetros do modelo de formas.

Portanto, qualquer nova forma x é uma deformação da forma média $\bar{\mathbf{x}}$, dado pela combinação linear dos autovetores Φ_f e o vetor de pesos b_f. Variando-se agora os valores do vetor b_f, podemos modelar linearmente diferentes outras formas em relação ao conjunto

de treinamento. Observe-se ainda que este modelo permite encontrar as variações de forma de modo contínuo, pois os parâmetros do vetor b_f podem variar em $\Re$ restritos apenas ao limite de $\pm 2\sqrt{\lambda_i}$ (COOTES et al., 1994).

Na figura 2.9 vemos o exemplo de um conjunto de treinamento (a) e as variações que ocorrem sobre uma imagem média (b) ao variarmos os parâmetros b_f onde $f = 1,2,...,p$. Observa-se que quando o parâmetro b_1 é variado dentro dos limites de $\pm 2\sqrt{\lambda_1}$, e o modelo de forma é reconstruído, a figura da mão sintetiza os modelos capturados e alguns modos que não necessariamente existiam no conjunto de treinamento.

(a) (b)

Figura 2.9 - A figura (a) representa o conjunto de treinamento, cujas formas foram obtidas a partir de 72 marcos referenciais. Em (b) temos a reconstrução da formas utilizando–se a equação (2.19) e variando-se os valores do vetor b_f. Adaptado de (COOTES et al., 1994), página. 47.

2.5.2 Modelo Ativo de Aparência

A reconstrução de novas formas executada pela abordagem ASM era sempre feita pelo contorno do objeto de estudo. Não era possível obter então a variação do objeto considerando-se a sua textura. Além do que, o contorno de um objeto leva em consideração a forma externa do objeto e não os detalhes internos.

Posteriormente em (COOTES et al., 1998), foi apresentado um novo método baseado no ASM que permitia modelar não somente formas, mas também a aparência das faces, considerando-se a sua textura. Esta abordagem ficou conhecida como *Active Appearance Model* (AAM). Os autores demonstraram que era possível modelar não somente a forma, mas também a textura, e assim poder combinar a variação de forma e de textura para modelar faces humanas.

Antes de continuar, deve-se definir o termo textura no contexto deste trabalho. Em processamento de imagens a textura pode ser considerada como um conjunto de padrões or-

ganizados de forma muito regular (FORSYTH; PONCE, 1990), porém no contexto do reconhecimento de faces, a textura é a intensidade do *pixel* sobre o objeto de estudo, no caso a face (STEGMANN et al., 2000). Da mesma maneira que na abordagem ASM, na abordagem AAM é necessário anotar a forma ou aparência dos conjuntos de treinamento. Isto é conseguido através dos marcos de referência (*landmarks*) posicionados manualmente sobre as imagens de face antes de se fazer o alinhamento e extração através do PCA. Na figura 2.10 vemos um exemplo dessa anotação manual.

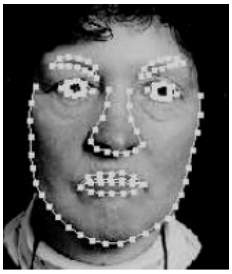


Figura 2.10 - Imagem de face com 122 marcos de referência anotadas manualmente. Adaptado de (COOTES et al., 1998).

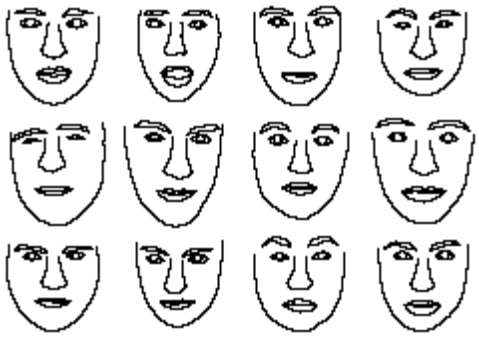


Figura 2.11 - Exemplo de formas faciais do conjunto de treinamento. Adaptado de (COOTES; TAYLOR, 2004) página 19.

Observa-se ainda que a técnica AAM trabalha sobre um modelo de caricatura obtido pelos marcos de referência, mas que na realidade representa a forma da face. Porém utilizar somente esses marcos referenciais não traduz fielmente o que se espera de uma modelagem de faces, isto porque a imagem de uma face requer informações sobre a sua textura para que se possam trazer informações mais relevantes. A figura 2.11 ilustra como ficam as caricaturas capturadas pelos marcos de referência. Observa-se que cada caricatura representa não somente uma imagem distinta de face de diferentes pessoas, mas também diferentes poses e expressões faciais.

Cootes *et al.* (COOTES et al., 1998) propuseram então combinar os modelos de forma juntamente com os modelos de textura, permitindo assim a visualização não somente das va-

riações da forma, mas também da textura. Como o modelo de textura poderia ser capturado do mesmo modo que o modelo de forma, teríamos assim um vetor t cujos elementos seriam as intensidades de cada *pixel* da imagem de face, conforme pode ser visto na equação, a seguir

$$t = (t_1, t_2, ..., t_n), \tag{2.20}$$

onde n é o número de *pixels* da imagem e t_i, para $i = 1,2,...,n$, representa a intensidade em nível de cinza de cada *pixel*.

A questão que surge agora é que não podemos montar uma matriz com todos os *pixels* das imagens do conjunto de treinamento porque o modelo de forma não contempla a imagem toda, isto porque existem mais pontos de textura do que de forma. Em outras palavras, as anotações dos marcos de referência somente contemplam regiões das imagens que possuem correspondência com as mesmas regiões das outras imagens. Logo o perímetro definido pelos marcos de referência é sempre menor que toda a imagem, consequentemente devemos somente considerar a textura das regiões de imagem contidas dentro do perímetro dos marcos de referência.

A solução para este problema é utilizar um algoritmo de triangularização (SHEWCHUCK, 1998) para ajustar a imagem em escala de cinza de cada exemplo do conjunto de treinamento, dentro do modelo de forma média previamente determinada. Após o ajuste de cada imagem dentro do modelo de forma média, normaliza-se cada imagem e então aplica-se o PCA na matriz de covariância Σ da textura para capturar a influência das variações dos níveis de cinza, e assim obtermos um modelo de textura, da mesma maneira que obtivemos com o modelo de forma. A equação (2.21) a seguir representa esse modelo

$$t = \bar{t} + \Phi_t b_t, \tag{2.21}$$

onde Φ_t é a matriz de autovetores de Σ e b_t é o vetor de parâmetros.

A combinação dos modelos de forma e de textura é feita removendo-se a correlação que existe entre elas, isto porque pode haver uma correlação entre o vetores b_f e b_t. Consequentemente, procuramos um modelo combinado e um parâmetro único que altere simultaneamente os dois modelos, e isto é conseguido aplicando-se o PCA novamente. Como os parâmetros b_f e b_t representam as possíveis variações de cada exemplo do conjunto de treinamento, seria razoável considerar que podemos montar uma matriz com os dois vetores concatenados, onde cada linha da matriz representaria um exemplo do conjunto, como pode ser visto na equação (2.22) a seguir,

$$b_g = \begin{pmatrix} W_f b_f \\ b_t \end{pmatrix} = \begin{pmatrix} W_f \Phi_f^T (x - \bar{x}) \\ \Phi_t^T (t - \bar{t}) \end{pmatrix} = Q\gamma, \tag{2.22}$$

onde W_f é uma matriz diagonal que ajusta a escala dos parâmetros de b_f em relação aos parâmetros de b_t, uma vez que cada um deles traduz uma grandeza distinta e diferente entre si e a variável b_g é um parâmetro global que incorpora as variações combinadas de b_f e b_t.

A variável γ será então o parâmetro comum aos dois modelos que modelará um exemplo tanto em forma quanto em textura e Q é a matriz de autovetores dos parâmetros combinados. Aplicando-se o PCA na equação (2.22), podemos escrever o parâmetro de cada modelo de forma e textura em relação a um parâmetro único γ, como está descrito na equação (2.23) para o modelo de forma e a equação (2.24) para o modelo de textura.

$$x = \bar{x} + \Phi_f W_f^{-1} Q_f \gamma. \tag{2.23}$$

$$t = \bar{t} + \Phi_t Q_t \gamma. \tag{2.24}$$

(a) (b)

Figura 2.12 - Exemplos dos modos de variação em forma (a) e em textura (b) das imagens de faces variando-se o parâmetro γ dentro dos limites de $\pm 3sd$ (*standard deviation*). Adaptado de (COOTES; TAYLOR, 2004) página 32.

A figura 2.12 (a) apresenta a reconstrução da face média variando-se o parâmetro γ da equação (2.23), onde observa-se as variações da forma da face (figura (a) inferior), pose e possivelmente o gênero (figura (a) superior). Na figura 2.12 (b) observarmos a reconstrução da face média variando-se o parâmetro γ da equação (2.24), e observa-se alterações nas expressões (figura (b) superior) e possivelmente na idade (figura (b) inferior). Essas reconstruções demonstram que o PCA capturou as diferenças entre os conjuntos de treinamento, e que essas diferenças podem ser modeladas sinteticamente variando-se uma determinada componente principal durante a fase da reconstrução.

2.6 Considerações Finais

Observa-se no conjunto dos trabalhos apresentados neste capítulo a utilização intensa das abordagens PCA e LDA, demonstrando que esses dois métodos estatísticos estão bem

consolidados nas aplicações do domínio das imagens de faces. Foram também descritos trabalhos que manipularam os modelos básicos do PCA e do LDA de modo a superar eventuais obstáculos ou limites dessas abordagens. E finalmente, duas abordagens distintas que investigaram diferentes formas de se interpretar as informações contidas nos autovetores, e não somente sob o ponto de vista da redução da dimensionalidade. No próximo capítulo veremos os modelos estatísticos e os modelos de representação das imagens de faces em mais detalhes, mas principalmente com sua interpretação voltada para o domínio das imagens de faces.

3 MÉTODOS ESTATÍSTICOS

Neste capítulo serão descritos os modelos de representação de imagens de faces, e em seguida a forma de se interpretar o modelo. Nas seções seguintes são ainda descritas as principais técnicas de estatística multivariada que serão utilizadas neste trabalho. A idéia fundamental é mostrar os caminhos percorridos durante o desenvolvimento deste trabalho e assim facilitar a compreensão da abordagem proposta.

3.1 Representação e Interpretação de Imagens de Face

Representar uma imagem de face através de um modelo significa encontrar um modelo matemático que possa se adequar aos parâmetros da face. Quando um modelo genérico descreve uma imagem de face, os parâmetros descrevem as variações de diferentes faces que ocorrem em torno desse modelo. Em outras palavras, representar uma face é encontrar um meio de se descrever formalmente uma face humana típica. Na interpretação dos modelos, tenta-se compreender suas variações, explicar como ocorrem essas variações e porque ocorrem.

As abordagens baseadas em modelos são aquelas em que as características da face são descritas por variações na forma (COOTES et al., 1995), variações na textura (TURK; PENTLAND, 1991), variações combinadas de forma e textura (COOTES; EDWARDS; TAYLOR, 1998), ou baseadas em características geométricas e adaptativas (BRUNELLI; POGGIO, 1983). Porém, de um modo geral, o que todas as abordagens fazem é extrair e isolar as fontes dessas variações nas imagens de faces e, assim, poder descrevê-las e interpretá-las. Isto é particularmente útil na estimação de uma imagem que não existe no conjunto de treinamento.

Os primeiros modelos de representação de faces para fins de extração utilizavam inicialmente as técnicas de processamento de imagens, tais como a segmentação (*edge detection*), para pré-processar as imagens (GONZALEZ; WOODS, 1992). Depois, padrões deformáveis sob certas restrições de rotação e translação eram aplicados nas imagens pré-processadas para se localizar características faciais (olhos, nariz, boca, rosto, sobrancelhas) que melhor se ajustassem aos padrões (SAKAI, et al., 1969). O maior problema deste tipo de modelagem por forma é o seu alto custo computacional, uma vez que a localização das características faciais é mais um processo de operação lógica do que numérica (CHELLAPPA et al., 1995).

Um outro exemplo típico de representação baseado no conhecimento são as abordagens que fazem uma sub-amostragem da imagem e depois procuram regiões na imagem que

apresentam maior intensidade em nível de cinza de *pixels* agrupados. Supondo-se *a priori* que a imagem contenha uma face, a região com a maior intensidade tem maior probabilidade de representar a região central de uma face (YANG; KRIEGMAN; AHUJA, 2002).

Na figura 3.1 é apresentado um exemplo da técnica de localização de faces baseada no conhecimento. Observa-se que a região do rosto é mais clara que o restante da imagem, como é representado na figura 3.1 (e), onde um agrupamento de células tem uma intensidade diferente do restante da matriz. É um modelo de representação muito simples e de baixo custo computacional, entretanto não é robusto o suficiente para localizar faces em imagens com múltiplas faces, ou com quando há uma face em um fundo mais branco que a face, ou então quando não se tem certeza de que a imagem em questão possui alguma face.

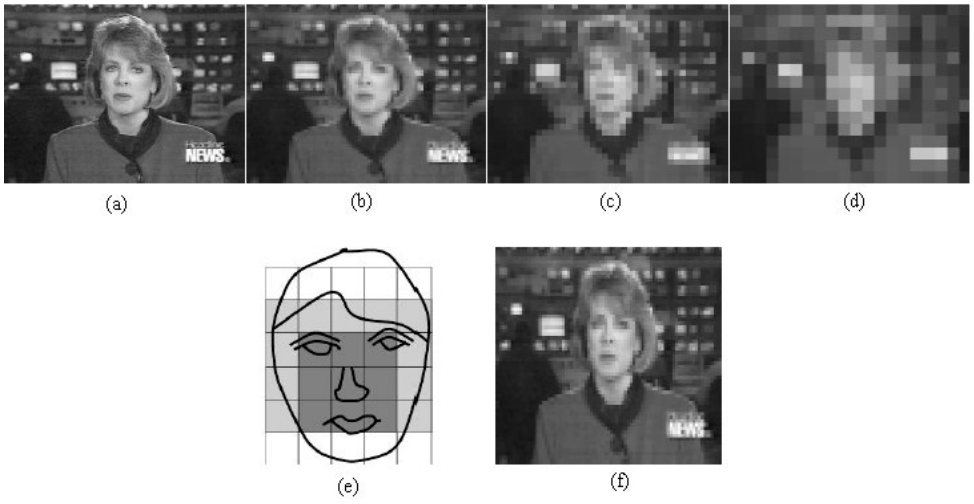


Figura 3.1 - Na seqüência das figuras (a), (b), (c) e (d) observamos o efeito da sub-amostragem da imagem original (f). Em (e) temos uma representação de como as regiões da face são distribuídas para análise. Adaptado de (YANG; KRIEGMAN; AHUJA, 2002) páginas 36, 37.

A importância de se discutir um modelo adequado para representar uma imagem de face decorre do fato de que o modelo é a representação simbólica do objeto de interesse do estudo, no nosso caso as faces. Os modelos permitem que separemos os objetos em classes e desta maneira possibilitar a identificação e separação dos diferentes objetos do mundo real. O modelo é uma representação abstrata do mundo real, porém o modelo não contempla todas as variações que possam ocorrer em torno dele. Portanto, encontrar um modelo que possa representar as imagens de face de forma precisa e que, ao mesmo tempo, seja flexível para conter todas as variações dessa representação, tornaria o processamento computacional mais confiável (BASSANEZI, 2002).

Neste trabalho são estudados os métodos baseados em aparências (*Appearance-Based Methods*), que trabalham as imagens de face como um todo e consideram a extração de ca-

racterísticas baseadas em modelos estatísticos. Os modelos baseados em aparência também são conhecidos como modelos holísticos. O leitor interessado em conhecer outras abordagens sobre detecção e localização de imagens deve consultar inicialmente (YANG; KRIEGMAN; AHUJA, 2002).

3.1.1 Representação Discreta de uma Imagem de Face

Uma imagem discreta de face pode ser representada como uma matriz $\mathbf{X}$ de dimensão $lin \times col$, onde lin é o número de linhas e col é o número de colunas da matriz $\mathbf{X}$. No entanto, é possível representar também essa mesma matriz como um vetor $\mathbf{x}^T$ n-dimensional, onde $n = lin \times col$, e que na realidade representa a matriz $\mathbf{X}$ da imagem de face concatenada. Considere que a imagem de uma face pode ser descrita como um vetor no espaço $\Re^n$, então essa face é um ponto nesse espaço $\Re^n$. Portanto um espaço de faces $\mathbf{Z}_\mathbf{x}$ é um espaço multi-dimensional que contém todas as faces de um determinado conjunto de treinamento (TURK; PENTLAND, 1991). Isto é válido considerando-se a operação em um sistema de coordenadas ortogonais e cujos vetores unitários $\varphi_1, \varphi_2, ..., \varphi_n$ representam essa base, e assim um ponto nesse espaço pode ser descrito como uma combinação linear de vetores ortonormais e das projeções do vetor $\mathbf{x}$ de uma dada face no sistema de coordenadas ortogonais (CALLIOLI et al., 1978), isto é:

$$\mathbf{x}_i = x_{i,1}\varphi_1 + x_{i,2}\varphi_2 + ... + x_{i,n}\varphi_n \,, \tag{3 25}$$

onde $\mathbf{x}_i$ é uma face no espaço de imagens $\mathbf{Z}_\mathbf{x}$.

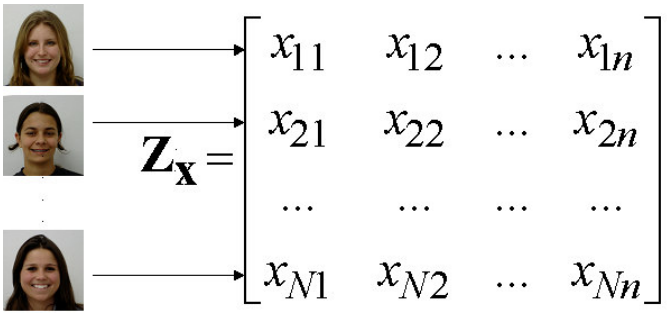


Figura 3.2 - Representação do espaço de imagens $\mathbf{Z}_\mathbf{x}$.

Considere-se que agora existe um conjunto de N imagens discretas de faces. Se as imagens são concatenadas e agrupadas em uma matriz $\mathbf{Z}_\mathbf{x}$, a matriz será então formada por $N \times n$ elementos. Deste modo, cada linha da matriz $\mathbf{Z}_\mathbf{x}$ será a representação de uma face. A figura 3.2 ilustra didaticamente um espaço de imagens, cuja dimensão é de N linhas por n colunas, onde N é o número de faces e n é o produto $linhas \times colunas$ de uma matriz de

face, e que representa a dimensionalidade desse espaço. A razão de se concatenar a matriz de face foi baseada no primeiro trabalho apresentado por (SIROVICH; KIRBY, 1987), onde os autores apresentaram a hipótese da redução da dimensionalidade da matriz de face utilizando a técnica da Transformada de Karhunen-Loève que normalmente é aplicada em sinais que variam no domínio do tempo e que são representados como vetores aleatórios (FUKUNAGA, 1990).

A figura 3.3 ilustra como seria o vetor imagem de uma face de dimensão 1×4096, visto como se fosse um sinal no domínio do tempo, sendo que neste caso o eixo do tempo é na realidade a posição de cada peso do vetor imagem. Observa-se no gráfico da figura 3.3 que muita informação é redundante, e que a face poderia ser representada somente pela variância que ocorre em torno de uma média (THOMAZ, 1999).

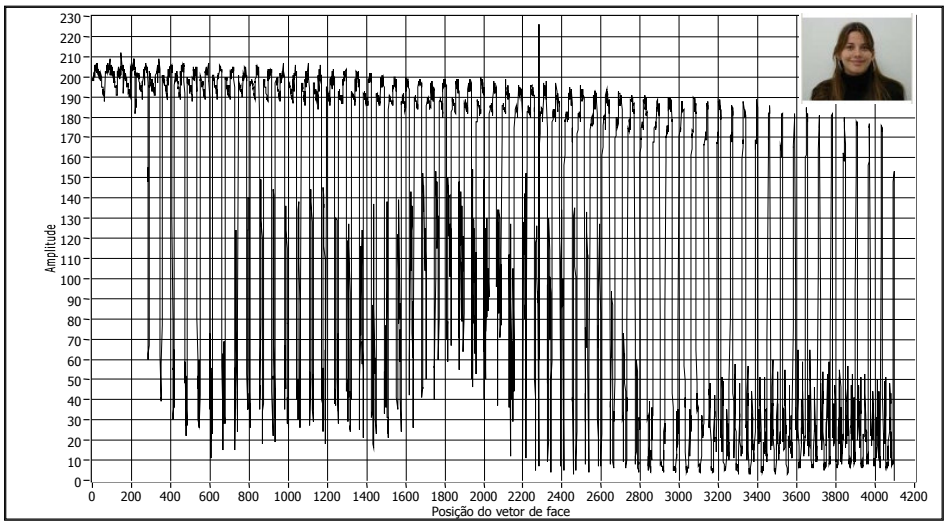


Figura 3.3 - Gráfico do vetor de imagem de uma face. No canto superior direito da figura temos a imagem da face que é representada por este gráfico[5].

A partir desta hipótese construiu-se a tese de que uma imagem de face poderia ser representada com poucas componentes principais, pois muito da informação contida em uma imagem de face é redundante. Observa-se que o trabalho de (SIROVICH; KIRBY, 1987) estava baseado apenas na hipótese da representação de uma face em um espaço *p*-dimensional menor que o espaço original, onde *p* representa o número mínimo de componentes principais e consequentemente $p << n$.

Com o intuito de detalhar esse processo de caracterização de imagens por componentes principais, apresenta-se na figura 3.4 a seguir o gráfico de duas faces distintas. Observa-se no gráfico que alguns pontos são comuns, principalmente com relação ao fundo da imagem.

[5] Apesar da imagem ser colorida, todo o trabalho foi executado com as imagens de faces transformadas para monocromático.

Pode-se observar também que há uma densidade de variação em algumas regiões, o que pode significar que capturando-se somente essas variações é possível representar cada face com um mínimo de informação e sem perda qualitativa da informação.

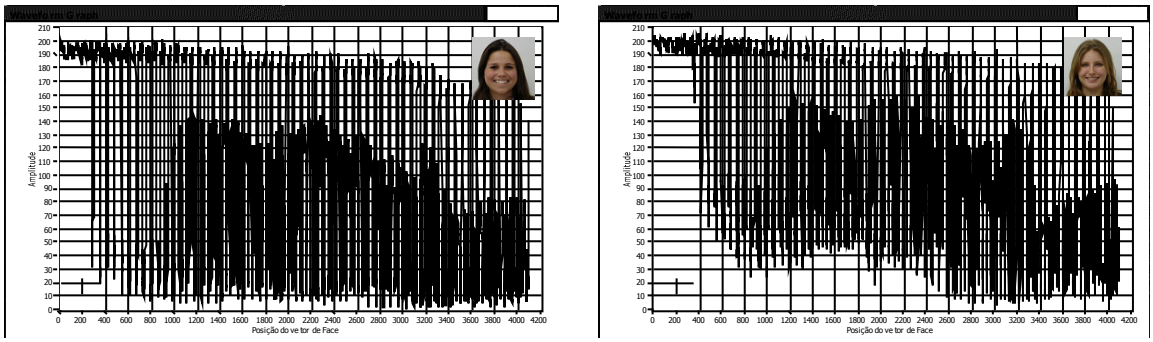


Figura 3.4 - Gráfico de duas faces distintas. No canto superior direito as imagens de face que representam cada gráfico.

Partindo-se da hipótese de que há muita informação redundante e ruído, e como foi dito anteriormente, se a variância de uma face pode ser descrita em torno de uma média, é razoável estudar essas variâncias em torno de uma média global. Portanto, tomando-se a matriz do espaço de faces $\mathbf{Z}_x$ e calculando o vetor média dessa matriz, isto é:

$$\overline{\mathbf{x}} = \left[\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i \right], \tag{3.26}$$

onde N é o número total de faces do conjunto de treinamento, determina-se o vetor média. O vetor média $\overline{\mathbf{x}}$ resultante é conhecido como "face média", e representa tudo aquilo que é comum a todas as faces. Um exemplo da imagem de uma "face média" pode ser visto na figura 3.5. Esta é a face média global e cada face $\mathbf{x}_i$ deverá variar em torno dessa média conforme:

$$\mathbf{x}_i^* = \mathbf{x}_i - \overline{\mathbf{x}}, \tag{3.27}$$

onde $i = 1,2,...,N$.

Figura 3.5 - Imagem de uma face média.

Este novo vetor $\mathbf{x}_\mathbf{i}^*$ contém todas as variações de uma determinada face i em torno da média $\overline{\mathbf{x}}$. Se então subtrairmos de todas as faces da matriz $\mathbf{Z}_\mathbf{x}$ o vetor média $\overline{\mathbf{x}}$, teremos então uma nova matriz $\mathbf{Z}_\mathbf{x}^*$,conforme a equação (3.28), que contém somente as variações de cada face em torno da média, e como conseqüência, uma matriz cuja média é zero e cuja a variância é máxima.

$$\mathbf{Z}_{\mathbf{x}}^{*} = [\mathbf{x}_1^{*}, \mathbf{x}_2^{*}, ..., \mathbf{x}_N^{*}]^T \ . \tag{3.28}$$

No gráfico da figura 3.6 vemos a representação da imagem média da figura 3.5 na forma de um sinal no domínio tempo. Considerando-se que este é o modelo, então todas as outras faces são variações deste modelo. Portanto, este trabalho considera a face média como sendo o modelo, e a principal vantagem desta consideração é o fato de que o modelo de face média está contido em qualquer conjunto de treinamento.

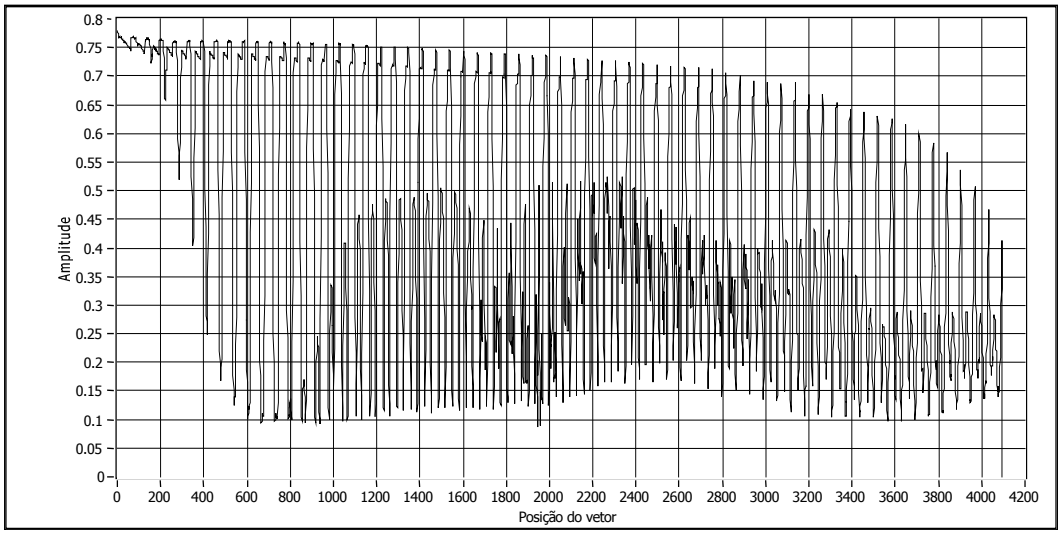


Figura 3.6 - Gráfico representativo na forma de sinal da face média (Figura 3.5) concatenada. A escala *y* está normalizada dentro dos limites de 0 a 1.

É relevante destacar que várias outras técnicas para análise e processamento de sinais já foram aplicadas nos estudos do domínio de imagens de faces, e isto se deve ao fato de uma imagem de face poder ser observada também como um sinal variável no tempo. Técnicas como a transformada de Fourier (SERGENT, 1986) (ZANA; CESAR JR., 2006) foram aplicadas para se determinar a influência de sinais de baixa e alta freqüência na determinação automática do gênero (sexo) de imagens de face. A transformada de Gabor também já foi aplicada para decomposição e redução de dimensionalidade (CHELLAPPA et al., 1995), (FERIS, 2001).

3.1.2 Interpretação dos Modelos

Uma vez definido o modelo de representação de faces, que este trabalho considera a face média como o modelo, deve-se discutir a interpretação das variações que ocorrem em torno do modelo. Essas variações são as diferentes representações que o modelo pode trazer do mundo real. Por exemplo, se temos um modelo que representa uma face, as variações da

cor da pele, cor do cabelo, comprimento do cabelo, tipo racial, forma dos olhos, da boca, etc., são as diversas variações que o modelo deve contemplar, e essas variações seriam descritas através das variações nos parâmetros do modelo.

Em termos matemáticos, a interpretação do modelo poderia ser compreendida como um classificador que pudesse extrair essas variações do nosso modelo. Diversos trabalhos na área de reconhecimento de faces apresentaram resultados com classificadores, porém apenas sob o aspecto da classificação para fins de reconhecimento. Poucos foram os trabalhos que discutiram as transições que ocorrem dentro do modelo. Essas transições pode ser entendidas como um grau de generalização e interpretação que um classificador pode fornecer baseado nas informações aprendidas do conjunto de treinamento.

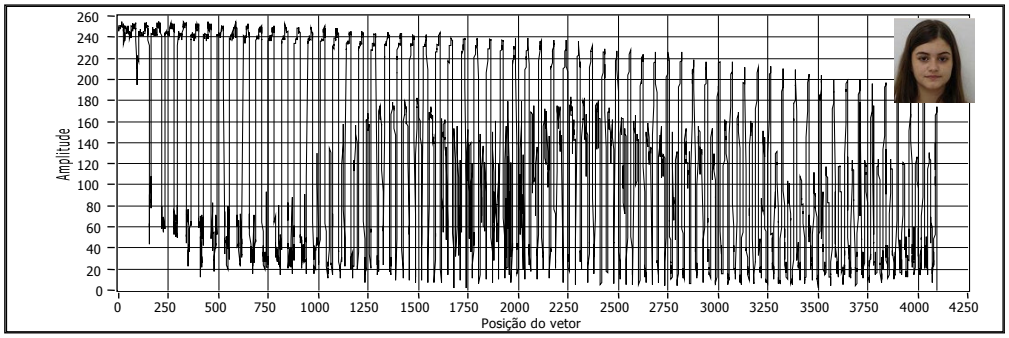
Por exemplo, se construirmos um classificador que classifique imagens de faces com as expressões sorrindo e não sorrindo, com certeza teríamos dois grupos distintos. Esta afirmação é baseada na suposição de que a distribuição dos dois grupos é Gaussiana, suportada pela teoria do limite central (JOHNSON; WICHERN, 2002). Assim, se apresentarmos uma imagem de uma pessoa que não necessariamente estivesse sorrindo, mas também não totalmente séria, a imagem seria considerado um *outlier* ou seria classificada erroneamente. Agora, considere que os classificadores normalmente extraem a informação de conjuntos finitos e não necessariamente bi-modais. Logo, o classificador poderia falhar no julgamento, ao interpretar uma determinada imagem de entrada que estivesse no cruzamento entre as duas curvas de distribuição.

(a)

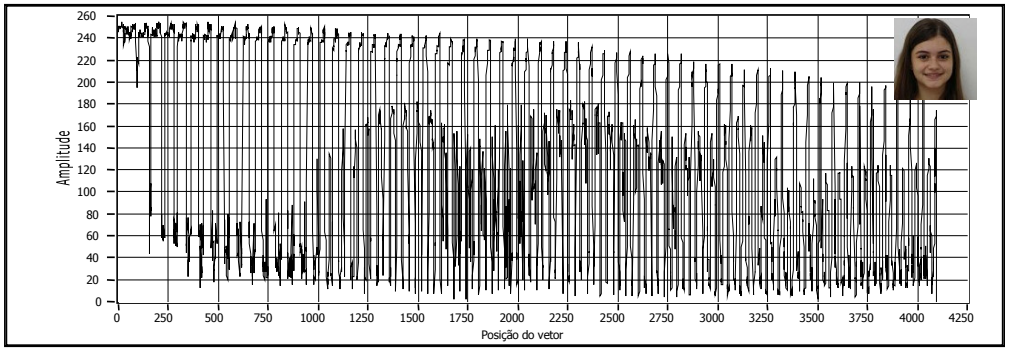
Figura 3.7 - Na figura (a) temos amostras de faces com expressão sorrindo, variando de um sorriso leve ao mais intenso. Em (b) temos as expressões neutras das amostras de (a).

Na figura 3.7 temos uma amostra visual das variações que podem ocorrer no modelo de face. Observa-se também que não somente as expressões são diferentes, mas também outras características biométricas. Observa-se que o sorriso varia de uma intensidade leve a um sorriso de intensidade mais alta. Na seção 5.3.5 descreveremos como determinamos os níveis

de intensidade do sorriso sem considerar fatores subjetivos na análise. A figura 3.8 representa graficamente esta dificuldade, onde a imagem de uma mesma pessoa é representada de forma análoga à forma de sinal no domínio tempo, contudo em duas expressões distintas, neutra e sorrindo apenas.



(a)



(b)

Figura 3.8 - Em (a) temos a representação gráfica de uma imagem de face com expressão neutra, em (b) o gráfico da imagem de face com expressão sorrindo. No canto superior direito, as imagens de face que são representadas pelos gráficos.

Observa-se pela figura 3.8 que existem diferenças entre os dois gráficos, entretanto a variação é muito pequena e até poderia ser interpretada como um ruído. Sabe-se que o sorriso de uma pessoa é uma expressão facial provocada pelos músculos da face. E as variações que os músculos faciais provocam não são discretas, mas são distensões e contrações contínuas no tempo. Porém as imagens de faces estáticas não capturam essas transições, logo, como seria possível predizer ou sintetizar os intervalos entre essas transições? Acredita-se que este trabalho propõe uma solução para esta questão.

Nas sub-seções seguintes são abordados os estudos sobre a extração de informações em imagens de faces, de modo a permitir a interpretação, parametrização e a síntese de uma imagem de face.

3.2 Análise de Componentes Principais

Nesta seção discute-se a técnica de análise de componentes principais ou *Principal Components Analysis* (PCA) sob os aspectos que são relevantes para o domínio de faces. Várias publicações e trabalhos já apresentaram de forma muito clara as bases teóricas e os algoritmos do PCA (JOLLIFE, 2002), (JOHNSON; WICHERN, 2002), (SHLENS, 2005), (TURK; PENTLAND, 1991). Em (FUKUNAGA, 1990) o leitor encontrará detalhes sobre a demonstração da optimalidade do PCA na redução da dimensionalidade dos dados e em (KITANI; THOMAZ, 2006a) uma adaptação daquela demonstração para o domínio de faces. Entretanto, nesta seção são discutidos as seguintes questões:

1) O que se deseja extrair com a aplicação do PCA no domínio de faces?
2) Como interpretar os autovetores e autovalores no contexto do domínio de faces?
3) Como ocorre efetivamente a redução do espaço de imagens para o espaço de faces?

Supondo-se que temos uma matriz de imagens de faces **Z** composta de *N* faces concatenadas de dimensionalidade *n* e com média 0. Se cada linha da matriz **Z** representa uma imagem de face e esta linha pode ser interpretada como um vetor *n*-dimensional, deve existir uma base vetorial *W* que permita escrever esse vetor como um ponto no espaço $\Re^n$, conforme pode ser observado na equação (3.29) abaixo,

$$\mathbf{x}_i^* = Wx_i, \tag{3.29}$$

onde *W* é uma base ortonormal , x_i um vetor de características da face *i* no espaço de faces, e $\mathbf{x}_\mathbf{i}^*$ o ponto que representa a face *i* no espaço $\Re^n$.

Portanto depara-se com o seguinte problema: como determinar essa base vetorial, uma vez que temos apenas os vetores de características do espaço de faces? De um modo geral poderíamos encontrar várias bases vetoriais que podem representar o conjunto **Z**. Mas desejamos encontrar uma base que represente melhor o conjunto **Z** em termos de variabilidade das amostras. Como o conjunto **Z** é composto de diferentes imagens de faces, é possível presumir que há uma distribuição das faces no espaço *n*-dimensional.

Esta variabilidade do conjunto **Z** pode ser eficientemente determinada pela matriz de covariância Σ da matriz **Z**. Porém, utilizando-se a matriz de covariância deve-se considerar que trabalha-se somente em espaços lineares, isto porque a covariância mede o grau de relação que existe entre os pares de variáveis. Consequentemente, calculando-se o produto da

matriz **Z** pela sua transposta, pode-se eficientemente determinar a matriz de covariância de **Z** (THOMAZ, 1999), como está proposto na equação (3.30) a seguir.

$$\Sigma = \frac{\mathbf{Z}\mathbf{Z}^{\mathbf{T}}}{N-1}. \tag{3.30}$$

A equação (3.30) permite a reflexão sobre dois aspectos fundamentais, que também foram levantados por (SIROVICH; KIRBY, 1987). Se o número de faces do conjunto **Z** for numericamente maior ou igual à sua dimensionalidade *n*, tem-se uma matriz cuja dimensão é $n \times n$, tornando tanto o cálculo quanto o consumo de memória massivo sob o ponto de vista computacional. Por outro lado se o número de faces do conjunto **Z** é muito menor que a dimensionalidade, $N << n$, a matriz de covariância não considerará todos os pares de características das faces, consequentemente há um espaço de características nulo na matriz de covariância, e essa matriz será considerada degenerada, tendo no máximo ordem $N-1$. As implicações para o primeiro caso são de ordem prática, pois demanda recursos computacionais para se determinar a matriz Σ. No segundo caso, observa-se que a matriz Σ é uma função de *N* ou menos vetores linearmente independentes do conjunto **Z** (FUKUNAGA, 1990). Este problema é conhecido como *Small Sample Size* (SSS), onde o número de amostras é muito menor que o espaço de características. Na seção 3.3.1 é abordado com mais detalhes as implicações teóricas do problema SSS.

Para ilustrar o problema SSS especificamente para o caso do PCA, o que ocorre é que a base vetorial formada pelos autovetores é limitada a $N-1$ vetores linearmente independentes. Apesar da dimensionalidade *n* ser muito maior, as imagens de faces somente podem ser descritas pela combinação linear dos $N-1$ autovetores disponíveis, logo há uma limitação no grau de liberdade da base vetorial. Considere um conjunto de faces $\mathbf{Z} = \{x_1, x_2\}$ e que cada face é representada por um vetor no espaço $\Re^3$. Considere ainda que existem apenas duas classes c_1 e c_2, e que os elementos do conjunto **Z** estejam distribuídos entre essas duas classes, conforme a relação (3.31).

$$x_1 \in c_1 \quad x_2 \in c_2. \tag{3.31}$$

Supondo-se que a base vetorial extraída para representar o conjunto **Z** pertença ao espaço $\Re^3$, e como temos apenas duas classes, somente é possível representar graficamente as variações desses vetores em apenas uma dimensão, como visto na figura 3.9 a seguir

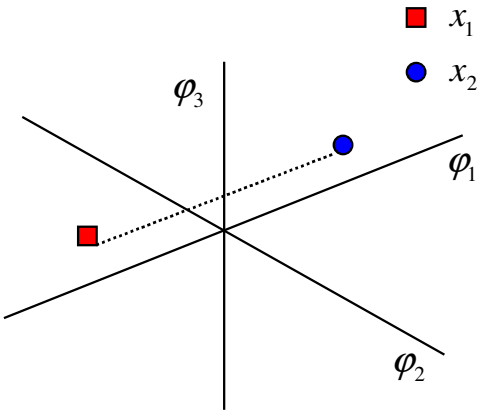


Figura 3.9 - Representação gráfica de uma base vetorial e da variação um par de pontos distribuídos nessa base.

O problema do tamanho das amostras não afeta diretamente o cálculo do PCA, uma vez que a determinação da base vetorial que maximiza a direção das maiores variâncias não utiliza inversão da matriz de covariância, mas é considerado apenas como um problema de maximização, como apresentado na equação (3.32),

$$W_{opt} = \arg\max_{w} \left| W^T \Sigma W \right|. \tag{3.32}$$

O problema é que a matriz Σ de dimensão $N \times N$ não determina os autovetores Φ para os autovalores Λ que são zero, isto porque um número de $(n-p)$ elementos são suprimidos. Isto faz com que os autovetores gerados por essa matriz sejam ortogonais entre si, mas não ortonormais. Fukunaga (FUKUNAGA, 1990) propõe uma solução para esse problema, recalculando a matriz de autovetores conforme indicado na equação (3.33) a seguir,

$$W_{pca} = \frac{Z^T \Phi (\Lambda)^{-\frac{1}{2}}}{\sqrt{N-1}}, \tag{3.33}$$

onde W_{pca} será a nova matriz de autovetores que é ortonormal.

A matriz W_{pca} tem dimensão $N \times p$ onde p é o número de componentes principais retidos e cada vetor φ_i representa uma vetor ortonormal da base que faz rotação do conjunto **Z** para a orientação que representa a maior variância. Portanto é possível agora responder a primeira questão formulada no início desta seção. O que se deseja com o PCA é extrair do conjunto **Z** a base vetorial que melhor representa a variabilidade das amostras. Apesar de muitos textos definirem que a função principal do PCA é reduzir a dimensionalidade do conjunto de treinamento, isto não ocorre de uma forma direta. Antes da redução é necessário determinar a base vetorial W_{pca}.

Como descrito acima, esta base vetorial W_{pca} é formada de autovetores w_i e cada vetor φ_i representa a direção de uma característica. Em aplicações no domínio de faces o número de vetores é normalmente menor que o número de características. Isto significa que cada autovetor pode representar a direção de mais de uma informação. A figura 3.10 representa uma nuvem de pontos distribuída no espaço $\Re^3$, e os vetores w_1, w_2, w_3 apontam as direções de maior variabilidade em ordem decrescente de importância. Desta maneira cada ponto pode então ser descrito como uma combinação linear de suas projeções nos eixos da nova base W_{pca} e, consequentemente, são estatisticamente independentes.

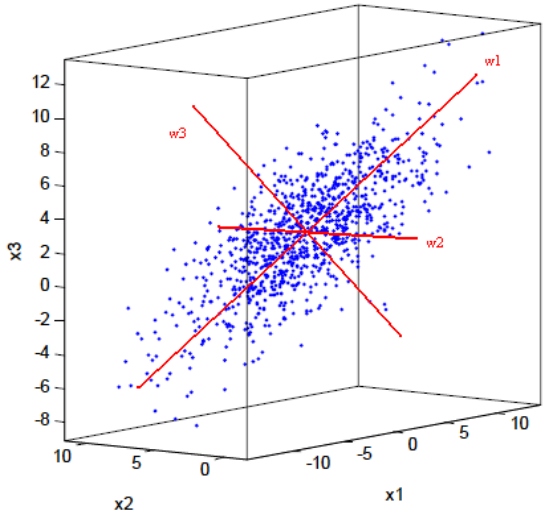


Figura 3.10 - Representação de uma nuvem de pontos em 3D e os eixos encontrados pelo PCA. Adaptado de (OSUMA, 2004a).

Considerando agora o que foi exposto acima, a segunda questão colocada no início da seção pode ser respondida, ou seja, o que representam os autovetores e os autovalores no domínio de faces. Os autovetores representam os vetores da base vetorial W_{pca} que indicam a direção das variâncias da amostra. Os autovalores representam a amplitude dessas variâncias, ou seja, como ocorre o espalhamento das amostras em cada eixo. Consequentemente, isto justificaria o fato de não termos autovetores associados aos autovalores iguais a zero, isto porque não há variação em torno desse autovetores. No contexto do domínio de faces, os autovetores são vetores de características extraídos do conjunto de treinamento, e onde são retidas as informações mais relevantes para fins de interpretação e reconstrução das imagens de face. A interpretação significa quais características são mais relevantes do ponto de vista da máxima variância, e a reconstrução é baseada na minimização do erro quadrático entre uma face *i* do espaço de imagens com a mesma face *i* vinda da reconstrução do espaço de faces.

Sabe-se agora que o PCA pode encontrar uma base vetorial que determina a direção de máxima variância das amostras, logo, se existem uma base vetorial e um conjunto de características que pode ser representado nessa base, então é possível escrever que,

$$\mathbf{Y} = \mathbf{Z}W_{pca}. \tag{3.34}$$

A equação (3.34) exprime uma transformação linear da matriz **Z** para a matriz **Y** através da matriz de transformação W_{pca}. Em outras palavras, a matriz **Y** é uma nova forma de se representar as imagens de face de **Z**.

Considere que a matriz **Z** tenha uma dimensão $N \times n$ e a matriz de transformação W a dimensão $n \times p$, onde $p << n$. O produto das duas matrizes resultará em uma matriz **Y** de dimensão $N \times p$, que é muito menor que a matriz **Z** original. O fenômeno da redução de dimensionalidade ocorre justamente nesta fase. A base vetorial W_{pca} contém e exprime as informações mais relevantes do ponto de vista da representação da matriz **Z**. Logo, qualquer imagem de face *n*-dimensional projetada nessa base será uma combinação linear dos vetores que efetivamente carregam uma informação relevante para fins de representação dessa face projetada. Assim, qualquer imagem de face pode ser eficientemente representada por *p* componentes principais.

Em resposta a terceira questão proposta no início da seção, ou seja, como ocorre a redução da dimensionalidade do espaço de imagens para o espaço de faces, observa-se que ela ocorre por dois fatores. O primeiro é o baixo número de exemplos em relação ao número de características, fazendo com que a matriz de covariância Σ tenha sempre o posto igual a $N-1$ no máximo, limitado pelo número de amostras *N*. Desta maneira, cada imagem é descrita como uma combinação linear de uma base vetorial $(N-1)$-dimensional. Apesar de uma imagem de face ser representada como um vetor *n*-dimensional, ela é descrita em uma base vetorial que apresenta somente $N-1$ vetores linearmente independentes. Caso a matriz **Z** tenha uma dimensão $n \times n$, a base vetorial W_{pca} extraída terá dimensão $n \times p$, porém $p = n-1$. Observa-se que agora a redução da dimensionalidade não será pela limitação do número de amostras, mas sim pela minimização do erro de reconstrução.

Como este estudo trata com informações obtidas através de sensores, neste caso câmeras fotográficas, possivelmente a informação estará contaminada por ruídos, logo a base vetorial provavelmente estará contaminada também pelo ruído. Como o ruído tem uma variância própria, pode-se utilizar a relação sinal/ruído para determinar quanto o ruído afeta o conjunto de amostras (SHLENS, 2005).

$$SNR = \frac{\sigma^2_{sinal}}{\sigma^2_{ruido}} \tag{3.35}$$

Utilizando-se a razão *SRN* (*Signal Noise Ratio*) da equação (3.35) e adaptando-a para medir a razão entre a componente principal PC_i e a PC_{i+1}, é possível definir um limiar θ (*threshold*) tal que se o *SNR* for menor que θ limita-se o número de componentes principais conforme pode ser visto na equação abaixo,

$$\begin{aligned} &SNR = \frac{PC_i^2}{PC_{i+1}^2} \Leftrightarrow SNR > \theta, \\ &para\ i = 1,2,...,n. \end{aligned} \tag{3.36}$$

Observa-se pela curva da figura 3.11 que a intensidade do *SNR* decai a medida que a componente principal aproxima de um ponto *p*. Este poderia ser considerado o ponto de corte, onde $p \leq \theta$.

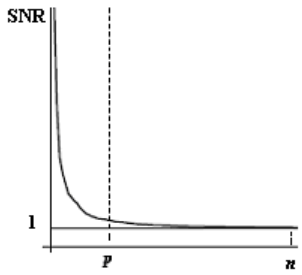


Figura 3.11 - Intensidade do *SNR* conforme a componente principal. Adaptado de (MOGHADDAN; WAHID; PENTLAND, 1998).

No entanto, como nas aplicações no domínio de faces o número de exemplos é sempre menor que o número de características, sempre se trabalhará com uma matriz Σ de dimensão $N \times N$. Portanto, pode-se concluir que durante a utilização da abordagem PCA no domínio de faces deve-se considerar algumas restrições ou características, como descritas a seguir:

- Como a base vetorial é extraída da matriz de covariância Σ essa base estará relacionada com a estatística de segunda ordem, ou seja, dependerá da média e da variância. Portanto, o PCA parte da hipótese de que as amostras têm uma distribuição Gaussiana, apesar de não ser possível afirmar que o espaço das imagens de faces tem este tipo de distribuição (CAMPOS, 2001).
- O PCA é uma metodologia linear e somente pode capturar relações lineares.
- O PCA é uma metodologia não supervisionada para extração de características.

3.3 Análise de Discriminantes Lineares

Foi apresentado na seção 3.2 como o PCA reduz a dimensionalidade através da manutenção da direção de maior variância, no entanto uma limitação do PCA é o fato dele não conseguir separar as amostras em classes. Isto seria particularmente útil na determinação das informações mais discriminantes e que não necessariamente são as informações que têm a maior variância. Para resolver este problema, foi proposto por Ronald A. Fisher (FISHER, 1936) um critério estatístico que maximiza a separação entre classes e minimiza a espalhamento dentro das classes.

A análise de discriminantes lineares ou *Linear Discriminant Analysis* (LDA) é uma técnica de estatística multivariada supervisionada, sendo assim espera-se que os dados apresentados para o LDA já estejam previamente agrupados em c classes. Considere duas medidas relativas aos dados de entrada, a primeira é a matriz de espalhamento inter-classes S_b e a segunda é a matriz de espalhamento intra-classes S_w. O critério de Fisher define uma métrica de separabilidade entre as classes, e a maneira de se determinar esse valor é calcular a razão entre os determinantes de S_b e S_w, conforme pode ser visto na equação (3.37).

$$Fisher\_criterion = \frac{\det(S_b)}{\det(S_w)} = \frac{|S_b|}{|S_w|}. \tag{3.37}$$

Assim, quanto maior a razão do critério de Fisher melhor será a separação, portanto o objetivo do LDA é encontrar uma nova projeção W para as amostras de entrada de modo a maximizar o critério de Fisher, como pode ser visto na equação a seguir,

$$W_{opt} = \arg\max_{w} \left| \frac{W^T S_b W}{W^T S_w W} \right|. \tag{3.38}$$

W é então uma matriz de projeção e ortonormal, e

$$S_b = \sum_{i=1}^{c} N_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T, \tag{3.39}$$

$$S_w = \sum_{i=1}^{c} \sum_{j=1}^{N_i} (x_{i,j} - \bar{x}_i)(x_{i,j} - \bar{x}_i)^T. \tag{3.40}$$

O vetor $x_{i,j}$ é o padrão j da classe i, N_i é o número de padrões da classe i, c é o número de grupos ou classes, $\bar{x}_i$ a média da classe i, e a média global $\bar{x}$ é dada por

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{g} N_i \bar{x}_i = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{Ni} x_{i,j}. \tag{3.41}$$

A solução da equação (3.38) pode ser generalizada como um problema de autovetores/autovalores e resolvido conforme a equação abaixo (ZHAO; CHELLAPPA; KRISHNASWAMY, 1998),

$$S_b \Phi = \Lambda S_w \Phi, \tag{3.42}$$

onde Φ é a matriz de autovetores e Λ a matriz de autovalores de $S_w^{-1} S_b$. Portanto, a maximização do critério de Fisher é obtida construindo-se uma matriz de transformação com a composição dos autovetores cujos autovalores são os maiores.

Considerando a título de exemplo que se trabalha com duas classes distintas de amostras que contém quantidades iguais de elementos. Então existem w_j vetores, para $j \geq 1, j \in \mathrm{N}$, que separam essas duas classes, e cada vetor w_j representará a direção de um plano de projeção onde poderemos projetar ortogonalmente todos os elementos das classes 1 e 2. Consequentemente, trabalhar sobre os dados discriminantes significa procurar um vetor W que maximize a separação entre as projeções das médias μ_1 e μ_2 de cada classe, sobre o vetor W, mas que também considere o espalhamento de cada classe em torno de suas respectivas médias. Então esse vetor W será o máximo da razão entre a matriz inter-classe S_b pela matriz dos elementos intra-classe S_w, conforme descrito na equação (3.38). A figura 3.12 ilustra o processo do busca da direção do hiper-plano discriminante que permite maximizar a separação entre as projeções sobre esse hiper-plano. A direção representada pelo vetor w_1 não consegue separar adequadamente os elementos dos grupos c_1 e c_2, no entanto, a direção apontada pelo vetor w_2 permite uma separação clara entre os dois grupos.

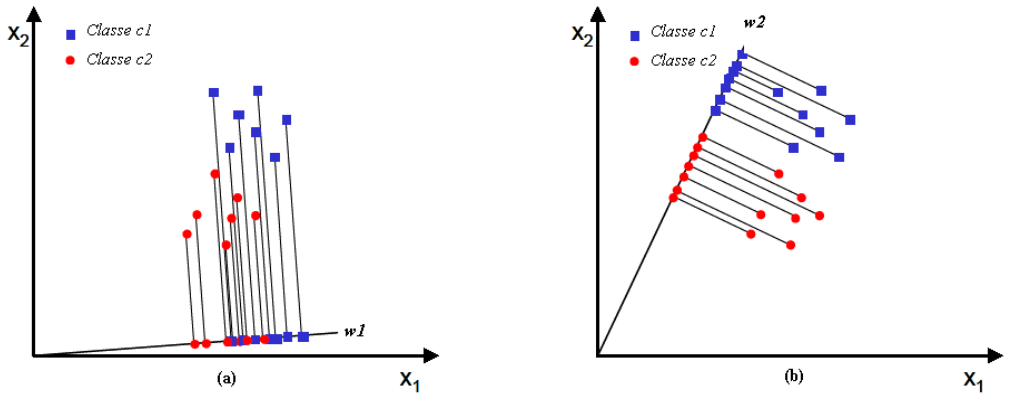


Figura 3.12 - Na figura (a) vemos um hiper-plano representado pela direção do vetor w_1 e em (b) um hiper-plano representado pela direção do vetor w_2. Adaptado de (OSUMA, 2004b).

No entanto, a aplicação direta da abordagem LDA em reconhecimento de faces pode ser problemática, isto porque, normalmente, o número de amostras de treinamento é sempre menor que o número de variáveis de cada amostra. Desta maneira a matriz S_w será singular provocando uma instabilidade na determinação da sua inversa. Em outras palavras, o LDA sofre com o problema do número pequeno de amostras no conjunto de treinamento (*Small Sample Size*), e este é um problema muito conhecido por aqueles que trabalham com o LDA em aplicações para reconhecimento de faces.

Na sub-seção seguinte descreve-se o problema do número pequeno de amostras.

3.3.1 Problema do Número Pequeno de Amostras

Como foi inicialmente descrito na seção 3.2, o problema do número pequeno de amostras, ou *Small Sample Size* (SSS) aparece com freqüência nas aplicações do domínio de imagens de faces, isto porque o número de exemplos normalmente é menor que o número de características. Uma primeira implicação do SSS nas abordagens que trabalham com classificadores é a taxa de erro do classificador. A taxa de erro é uma métrica muito aceita para se medir o desempenho de um classificador, e pode ser determinada uma taxa de erro aparente E_{ap}, fornecida através da seguinte equação,

$$E_{ap} = \frac{N_{erro}}{N_{casos}} 100, \tag{3.43}$$

onde N_{erro} é o número de erros efetivos cometido pelo classificador e N_{casos} é número de casos testados. Considere-se que na taxa de erro aparente estão contidos os erros cometidos pelo classificador, tais como o falso positivo e o falso negativo. A taxa de erro aparente somente considera os erros cometidos pelo classificador baseados nos exemplos do conjunto de treinamento. A estratégia é treinar o classificador com um conjunto de treinamento e depois apresentar esse mesmo conjunto para se avaliar a taxa de erro aparente. Para uma análise mais precisa seria necessário utilizar a taxa de erro verdadeira E_{true}, que leva em consideração os erros cometidos pelo classificador sobre um grande conjunto de amostras diferentes do conjunto de treinamento. No entanto sua aplicação prática é muito difícil, dado ao grande número de exemplos distintos necessários para o teste (BATISTA, 1997), (THOMAZ; GILLIES, 2001). Portanto um número pequeno de amostras no conjunto de treinamento pode produzir uma baixa capacidade de generalização do LDA e consequentemente os erros de classificação tendem a aumentar.

Em (RAUDYS; JAIN, 1991) os autores apresentam os resultados de vários trabalhos que investigaram a relação existente entre a taxa de erro de um classificador e o número de amostras de um conjunto de treinamento, e concluíram que os classificadores são muito sensíveis ao problema SSS, mas sugerem que o número de amostras seja limitado a um número mínimo, dependendo do tipo de classificador utilizado.

Um segundo problema recorrente do SSS refere-se à singularidade da matriz de covariância intra-classes S_w. Como foi descrito anteriormente na seção 2.3.1, a matriz S_w é invertível se e somente se o posto de S_w for maior que $N-c$ ou se $\det|S_w| \neq 0$. Caso o número de exemplos seja menor que a dimensionalidade dos dados, a matriz S_w não será invertível. E mesmo que a matriz S_w permita inversão, o pequeno número de exemplos poderá produzir uma matriz instável, isto porque teremos muitos autovetores associados a autovalores nulos ou de baixíssimo valor, e muitos desses autovetores pertencerão ao espaço de números complexos.

Conclui-se que o LDA é altamente afetado pela tamanho das amostras do conjunto de treinamento em relação à dimensionalidade, no entanto algumas soluções para regularizar a matriz S_W foram propostas. Na seção seguinte abordaremos o modelo apresentado em (THOMAZ; GILLIES, 2005), denominado *Maximum uncertainty* LDA (MLDA), que foi proposto para estabilizar a matriz S_w e permitir a cálculo da razão $S_w^{-1}S_b$.

3.4 MLDA

Como descrito no início da seção 3.3, o objetivo principal da abordagem LDA é encontrar uma matriz de projeção que maximize a razão $S_w^{-1}S_b$, no entanto, todos os algoritmos que trabalham com o modelo LDA clássico sofrem com a instabilidade provocada pelo matriz S_w, uma vez que para se calcular a matriz W é necessário determinar a solução para o seguinte sistema linear,

$$S_b\Phi - S_w\Phi\Lambda = 0. \tag{3.44}$$

Freqüentemente, os algoritmos para calcular a equação (3.44) utilizam a técnica da diagonalização simultânea das matrizes S_w e S_b, mas para se determinar a diagonalização das matrizes é necessário antes calcular a matriz inversa de S_w, como pode ser visto na equação a seguir (FUKUNAGA, 1990)

$$S_w^{-1} S_b \Phi = \Phi \Lambda. \tag{3.45}$$

Se a matriz S_w for singular, não será possível determinar a inversa, portanto o cálculo dos discriminantes lineares não será possível. Sendo assim, em aplicações para reconhecimento de faces, onde o número de características é sempre muito maior que o número de exemplos, este problema sempre ocorrerá. A abordagem *Maximum uncertainty* LDA (MLDA) trata este problema substituindo a matriz S_w por uma outra matriz S_w^* que tem uma expansão dos autovalores que possuem os menores ou zero autovalores, e mantendo aqueles que já são maiores. Esta expansão mantém as maiores variâncias das amostras, pois quando se calcula os autovetores da matriz S_w no LDA clássico, observa-se que, em termos numéricos, os autovetores apresentam um desvio padrão próximo de zero, e consequentemente a magnitude do desvio em relação a média é muito pequena.

Considere que a equação (3.40) possa ser rescrita da seguinte forma,

$$S_w = \sum_{i=1}^{c} (N_i - 1) S_i = \sum_{i=1}^{c} \sum_{j=1}^{N_i} (x_{i,j} - \bar{x}_i)(x_{i,j} - \bar{x}_i)^T, \tag{3.46}$$

onde $x_{i,j}$ é o j*-ésimo* padrão da classe i, S_i a matriz de covariância da classe i, N_i o número total de amostras da classe i, e c o número total de classes ou grupos. Portanto, a matriz S_w é na realidade uma matriz de covariância S_p multiplicada pela média ponderada de todos os grupos de amostras e pode ser determinada por

$$S_p = \frac{1}{N-c} \sum_{i=1}^{c} (N_i - 1) S_i = \frac{(N_1 - 1)S_1 + (N_2 - 1)S_2 + ...(N_g - 1)S_c}{N - c} \tag{3.47}$$

Desta maneira, procuramos uma nova matriz S_w^* que possa ser escrita conforme a equação a seguir,

$$S_w^* = S_p^*(N - c) = (\Phi \Lambda^* \Phi^T)(N - c), \tag{3.48}$$

onde Φ são os autovetores da matriz S_p e Λ^* é uma nova matriz de autovalores que é formada baseada na dispersão dos maiores autovalores λ, conforme a equação a seguir,

$$\Lambda^* = diag[\max(\lambda_1, \bar{\lambda}), ..., \max(\lambda_n, \bar{\lambda})]. \tag{3.49}$$

Na equação (3.49), o autovalor $\bar{\lambda}$ é a média dos autovalores da matriz S_p, definida na equação (3.50) a seguir,

$$\overline{\lambda} = \frac{1}{n}\sum_{j=1}^{n}\lambda_j. \tag{3.50}$$

O algoritmo proposto no MLDA segue então os seguintes passos:

i. Encontre os autovetores Φ e os autovalores Λ da matriz S_p, considerando-se que $S_p = S_w/[N-c]$, isto porque distribuímos a população do conjunto de treinamento por *c* grupos.

ii. Calcule o autovalor médio $\overline{\lambda}$ da matriz Λ utilizando a equação (3.51) abaixo,

$$\overline{\lambda} = \frac{1}{n}\sum_{j=1}^{n}\lambda_j. \tag{3.51}$$

iii. Forme uma nova matriz de autovalores baseado na seguinte dispersão de valores;

$$\Lambda^* = diag[\max(\lambda_1, \overline{\lambda}),..., \max(\lambda_n, \overline{\lambda})]. \tag{3.52}$$

iv. Calcule o novo S_w^* conforme a equação (3.48) , substitua a matriz S_w da equação (3.38) pela nova matriz S_w^* e calcule o LDA clássico (THOMAZ; GILLIES, 2005).

Apesar da abordagem MLDA trabalhar com uma matriz identidade I de dimensão $n \times n$, o cálculo dos autovetores e autovalores da matriz S_p é tratável sob o aspecto computacional e matemático, ao contrário da abordagem LDA clássica. No entanto, se o consumo de memória durante o processamento for problemático, pode-se utilizar a abordagem combinada PCA+MLDA, e assim reduzir o consumo de memória (THOMAZ, 2004).

3.5 Considerações Finais

Ao longo deste capítulo foram descritos os modelos de representação das imagens de faces, as abordagens de estatística multivariada comumente usadas nas aplicações do domínio de faces, bem como um modelo de regularização para o LDA proposto recentemente por (THOMAZ; GILLIES, 2005). Além desta breve revisão, foram também discutidos alguns conceitos que são relevantes para aqueles que trabalham no domínio de faces. Esse conjunto de informações formam a base teórica da abordagem proposta neste trabalho, e permitirá uma melhor compreensão dos resultados obtidos nos experimentos realizados. No próximo capítulo é descrito o modelo estatístico discriminante desenvolvido neste trabalho.

4 MODELO ESTATÍSTICO DISCRIMINANTE

A abordagem proposta nesta dissertação é denominada Modelo Estatístico Discriminante ou *Statistical Discriminant Model* (SDM), e é essencialmente um classificador combinado PCA+MLDA (KITANI; THOMAZ; GILLIES, 2006b). Basicamente o SDM trabalha sobre um conjunto de treinamento previamente construído com média zero, e extrai desse conjunto as características mais discriminantes que determinam a separação entre os grupos, e as apresenta visualmente.

Outros estudos já abordaram modelos combinados PCA+LDA conforme descrito sucintamente na seção 2.3.3, no entanto, na maioria das abordagens o espaço PCA retido é sempre menor que $N-1$, onde N é o número de exemplos. Isto ocorre porque apesar da redução de dimensionalidade que o PCA produz, apenas um número de $N-c$ componentes principais deve ser retido, onde c é o número de classes. Isto é necessário para se evitar que a matriz S_w se torne singular. Entretanto, durante o processo de descarte das menores componentes principais, muita informação importante para fins de discriminação pode ser perdida. Por esta razão, este trabalho propõe o uso de todo o espaço PCA restrito a $N-1$ componentes principais, visto que a face média já foi previamente removida. Este espaço PCA será então enviado à abordagem MLDA para que se faça a extração da base vetorial W que melhor discrimine os autovetores do PCA para fins de classificação. Como a abordagem MLDA estabiliza a matriz S_w, o uso de todo o espaço PCA permite minimizar o erro de reconstrução, fornecendo melhores informações durante a visualização.

Na próxima seção são descritos em detalhes como a abordagem SDM opera para executar a extração de características e síntese visual dessas características.

4.1 SDM: Um Classificador de Dois Estágios

O SDM é um classificador de dois estágios que opera em quatro fases distintas. A primeira fase é a fase do treinamento dos conjuntos de amostras. Nessa fase as amostras devem ser previamente alinhadas e normalizadas. Isto se faz necessário para se evitar que a abordagem SDM capture variações que não necessariamente são relacionadas com as diferenças efetivas que se deseja capturar. Em outras palavras, dados dois conjuntos formados por faces sorrindo e não sorrindo de pessoas distintas, deseja-se capturar as variações provocadas pelo sorriso e não sorriso, e não necessariamente a diferença entre as distintas faces dos con-

juntos de treinamento. Observa-se que as operações de alinhamentos e normalizações são processos que não fazem parte do SDM e devem ser executadas previamente.

Seja $\mathbf{x}_i$ uma face de amostra, o processo de formação do conjunto de treinamento inicia-se lendo as matrizes das imagens de face $\mathbf{x}_i$ para $i = 1,2,...,N$, concatenando-as e montando uma matriz de faces $\mathbf{Z}$, onde cada linha representa uma face $\mathbf{x}_i$. Como o treinamento é supervisionado, a formação das classes de imagens é executada nesta fase. Portanto, lêem-se todas a imagens de uma classe e depois todas as imagens da outra classe. Desse conjunto de treinamento é calculado e removido a face média global $\bar{x}$, conforme a equação (3.26), de cada linha da matriz $\mathbf{Z}$, criando-se uma nova matriz $\mathbf{Z}^*$, cuja média é zero e a variância é máxima. Em seguida, a partir da matriz Z^* aplica-se o PCA para se extrair os autovetores Φ_{PCA} que formam a base vetorial do espaço de imagens. O produto vetorial entre a matriz Z^* e a matriz de autovetores Φ_{PCA} resulta na formação da matriz $\mathbf{Y}_{PCA}$, conforme pode ser visto na equação abaixo,

$$\mathbf{Y}_{PCA} = \mathbf{Z}^* . \Phi_{PCA}^T . \tag{4.53}$$

A nova matriz $\mathbf{Y}_{PCA}$ contém as projeções das imagens de face do conjunto de treinamento $\mathbf{Z}^*$ na nova base vetorial formada pelas autofaces Φ_{PCA}. A figura 4.1 a seguir ilustra didaticamente o fluxo dos dados ao longo desta primeira fase. O processo descrito acima é semelhante à abordagem apresentada em (TURK; PENTLAND, 1991). Apesar de as imagens de face contidas na figura 4.1 serem coloridas, elas são apenas ilustrativas, pois todo o trabalho foi desenvolvido com imagens monocromáticas, além de terem também a sua resolução reduzida para 64×64 *pixels*. A descrição detalhada sobre o pré-processamento das imagens de faces do conjunto de treinamento encontra-se na seção 5.1. O retângulo que contém a palavra PCA representa todas as operações necessárias para se determinar a matriz de autovetores ordenadas conforme os maiores autovalores. Observa-se também que a figura apresenta a dimensão de cada matriz ou vetor durante as várias etapas, onde é relevante destacar a dimensão do espaço de faces $\mathbf{Y}_{PCA}$ como resultado do produto interno descrito pela equação (4.53).

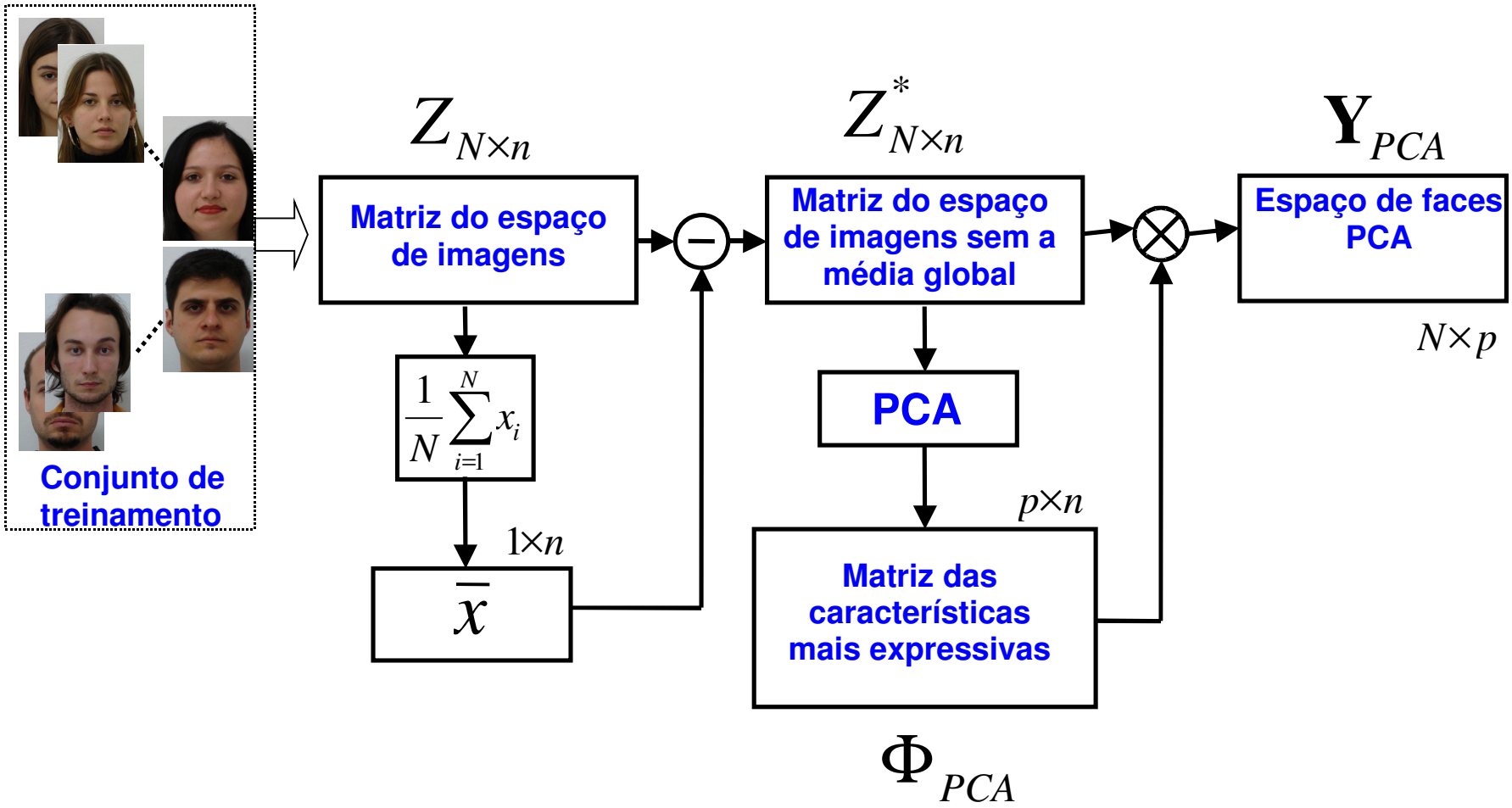


Figura 4.1 - Representação esquemática da fase de treinamento e formação do espaço de características PCA. Adaptado de (THOMAZ et al., 2006).

Apesar da matriz $\mathbf{Y}_{PCA}$ conter as informações que melhor representam cada uma das faces do conjunto de treinamento, não significa que as direções apontadas pelo PCA sejam necessariamente as melhores direções para fins de classificação. Portanto, uma segunda fase consiste em se aplicar um algoritmo de classificação na matriz $\mathbf{Y}_{PCA}$ para se extrair as informações mais discriminantes entre as duas classes do conjunto de treinamento. A solução utilizada neste trabalho foi aplicar o algoritmo MLDA.

Aplicando-se o MLDA na matriz $\mathbf{Y}_{PCA}$ forma-se uma matriz de autovetores Φ_{MLDA} que será a base vetorial de projeção das informações mais discriminantes do conjunto de treinamento. De forma análoga ao algoritmo do PCA, faz-se o produto vetorial da matriz $\mathbf{Y}_{PCA}$ com a matriz Φ_{MLDA} e obtém-se a matriz $\mathbf{Y}_{MLDA}$, que forma na realidade, um espaço de características discriminantes entre as duas classes ou conjuntos de treinamento. A equação (4.54) representa a operação de transformação do espaço PCA para o espaço MLDA,

$$\mathbf{Y}_{MLDA} = \mathbf{Y}_{PCA}\Phi_{MLDA}. \tag{4.54}$$

Como esta dissertação trabalha somente com dois grupos distintos de imagens de face, o hiperplano que separa as duas classes é formado apenas por um vetor que é representado pelo autovetor Φ_{MLDA}. A razão de se trabalhar com apenas dois conjuntos de amostras é que está é a melhor forma de se fazer uma análise de discriminantes lineares, tanto para o PCA quanto para o MLDA, isto considerando-se que as duas populações têm uma distribuição normal multivariada distintas entre si. Se as médias de cada população e as respectivas matrizes de covariância são conhecidas a *priori*, então a estimação de uma função critério de separação é facilitada (JOLLIFFE, 2002).

A figura 4.2 ilustra didaticamente o fluxo dos dados ao longo desta segunda fase, e o fluxograma contido dentro do pontilhado representa a fase de treinamento PCA, descrito anteriormente. Observa-se ainda pela figura 4.2 que todas as faces do conjunto de treinamento estão representadas dentro do vetor $\mathbf{Y}_{MLDA}$. Considere uma face x_i do conjunto de treinamento, a projeção dessa *i*-ésima face no espaço MLDA é representada pelo *i-ésimo* elemento do vetor $\mathbf{Y}_{MLDA}$, ou seja, y_i^{MLDA}. Conclui-se que o vetor $\mathbf{Y}_{MLDA}$ representa a projeção da informação mais discriminante do conjunto de treinamento no espaço MLDA.

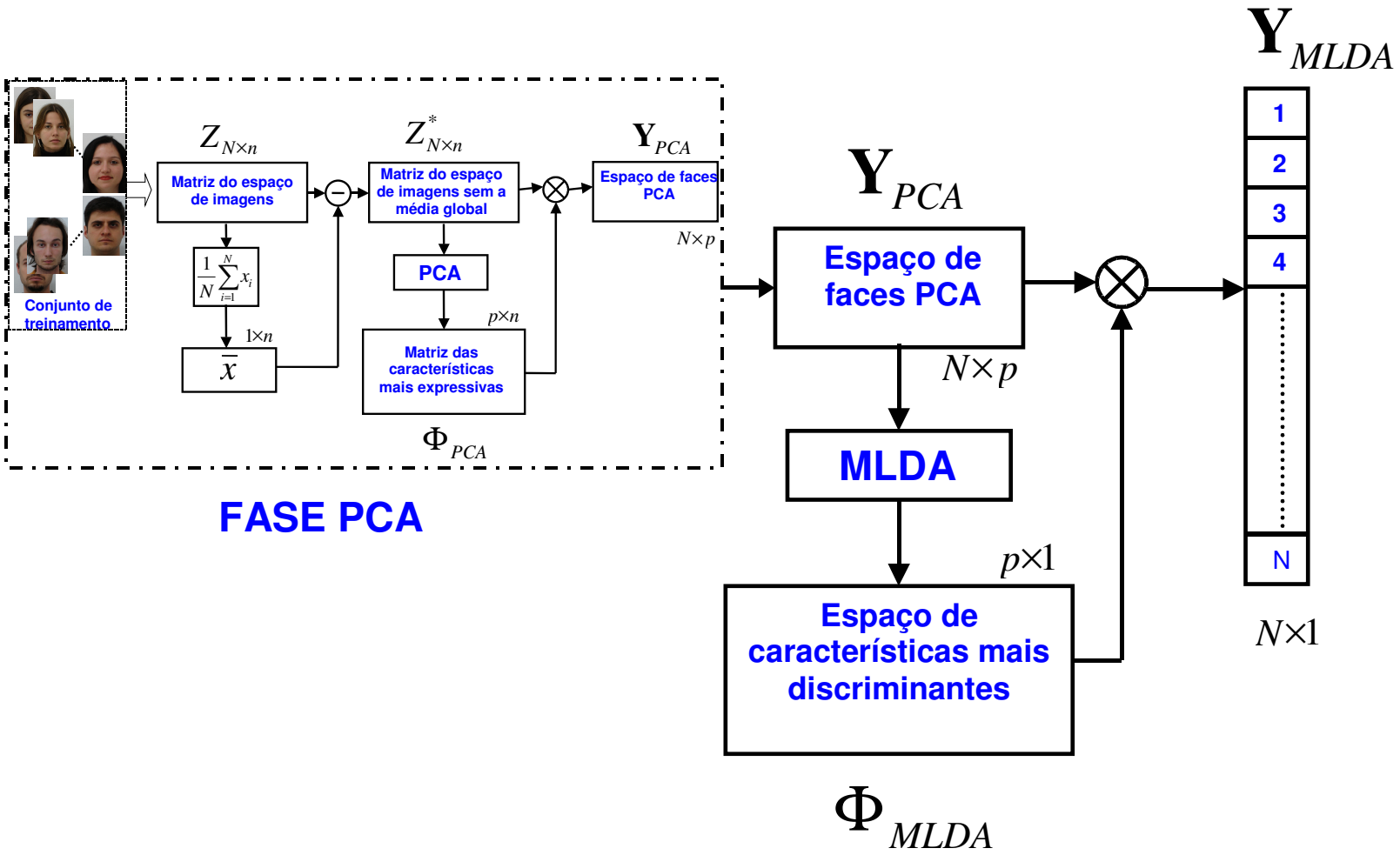


Figura 4.2 - Representação esquemática da fase de treinamento e formação do espaço de características MLDA. Adaptado de (THOMAZ et al., 2006).

Como descrito anteriormente, a formação do conjunto de treinamento é executada de modo supervisionado. No entanto, suponha-se que não são conhecidas a *priori* as características mais discriminantes entre os dois grupos para fins de classificação. Logo, seria útil observar visualmente quais seriam as características discriminantes extraídas pelo classificador e identificar visualmente o que separa essas duas classes. Uma maneira direta de se observar essas características seria reconstruir qualquer um dos elementos do vetor $\mathbf{Y}_{MLDA}$. Isto pode ser conseguido tomando-se um elemento y_i^{MLDA} do vetor $\mathbf{Y}_{MLDA}$ e multiplicando-o pela transposta da matriz de autovetores Φ_{MLDA}, conforme descrito na equação (4.55) a seguir,

$$y_i^{PCA} = \Phi_{MLDA}^T y_i^{MLDA}. \tag{4.55}$$

O resultado do produto vetorial da equação (4.55) sintetiza a *i-ésima* característica do espaço MLDA para o espaço PCA. De maneira análoga, multiplicando-se o vetor sintetizado y_i^{PCA} pela transposta da matriz de autovetores Φ_{PCA}, sintetizaremos essa característica para o espaço de imagens. Como de todo o conjunto de treinamento é removido a média, é razoável somar a média global $\bar{x}$ ao vetor sintetizado x_i, conforme é mostrado na equação (4.56),

$$x_i = \bar{x} + \Phi_{PCA}^T \Phi_{MLDA}^T y_i^{MLDA}. \tag{4.56}$$

Conclui-se que para cada $i = 1,2,...,N$ é possível sintetizar visualmente a característica mais discriminante. Essa síntese visual reconstrói uma imagem de face que incorporou alguma característica considerada discriminante pelo MLDA. Em outras palavras, para cada face projetada no espaço MLDA, o espaço MLDA apenas reterá a informação que classifica cada imagem de face conforme as *c* classes definidas durante a fase de treinamento.

A navegação na direção do autovetor que separa as duas classes e a síntese visual durante a navegação permitiria a observação das características que foram capturadas para a classificação dos dois conjuntos de treinamento. Isto configura a terceira fase da abordagem SDM, ou seja, nesta fase o SDM permite visualizar as características discriminantes capturadas pelo classificador MLDA.

A figura 4.3 a seguir representa o processo de reconstrução visual das características mais discriminantes capturadas pelo MLDA. Observa-se que fluxo é basicamente o processo inverso da fase de treinamento do PCA e do MLDA. Como descrito acima, a síntese visual apenas retorna as características discriminantes que foram capturadas, e não necessariamente a identidade de alguma face do conjunto de treinamento. A informação da identidade de cada face é atenuada durante a fase de cálculo do classificador, uma vez que a determinação do critério de Fisher considera a razão do $\det|S_b|$, que é a distância quadrática entre as médias

das classes *c*, pelo $\det|S_w|$, que é a matriz de covariância das amostras de todas as classes *c* (JOHNSON; WICHERN, 2002). Em outras palavras, o critério de Fisher se dará pela maximização da normalização das médias das classes pela somatória das variâncias de cada classe. Logo, não se consideram informações para fins de reconstrução, mas somente informações que são úteis para a classificação. Assim, o espaço MLDA retém apenas as informações sobre a média de cada classe e as variações em torno de cada média. As imagens à direita da figura 4.3 ilustram a síntese das imagens de face média retidas no espaço MLDA. Maiores detalhes deste experimento estão descritos na seção 5.3.

É importante destacar que durante a fase da reconstrução das informações discriminantes, é possível reconstruir tanto os pontos existentes no vetor $\mathbf{Y}_{MLDA}$ quanto aqueles valores que não necessariamente pertencem ao conjunto de pontos do vetor. Isto permite a reconstrução sintética de um determinado ponto y_k^{MLDA} pertencente ao intervalo $y_i^{MLDA} < y_k^{MLDA} < y_{i+1}^{MLDA}$. Esta reconstrução é equivalente a uma interpolação linear da imagem da face média no espaço de alta dimensão.

A última fase da abordagem SDM trabalha com a transferência de características discriminantes para imagens de faces tanto do conjunto de treinamento quanto daquelas que não pertencem ao conjunto de treinamento. A figura 4.4 ilustra a operação desta quarta fase. Toma-se uma imagem qualquer x_i pertencente ou não ao conjunto $\mathbf{Z}$ do conjunto de treinamento, subtrai-se a média global e depois soma-se a imagem média reconstruída do espaço MLDA. Observa-se que esta operação incorpora na imagem de face com identidade somente as características discriminantes da face média retidas no espaço MLDA, não alterando-se a identidade. Esta face do SDM indica que o modelo de faces que é baseado na face média é a conjunto de todas as características que são comuns a todas as faces. Como cada face é uma variação do modelo, se modificarmos o modelo as variações de cada face acompanharão o modelo, incorporando na face com identidade as características do modelo. Se o modelo sorri, as faces reconstruídas também sorriem.

Na próxima seção são descritas brevemente as diferenças e semelhanças entre o SDM e as duas abordagens (AAM e BUCHALA et al., 2005) que mais se aproximam em termos de objetivos e resultados. Essa comparação objetiva apenas contextualizar o SDM entre todas as abordagens descritas nesta dissertação.

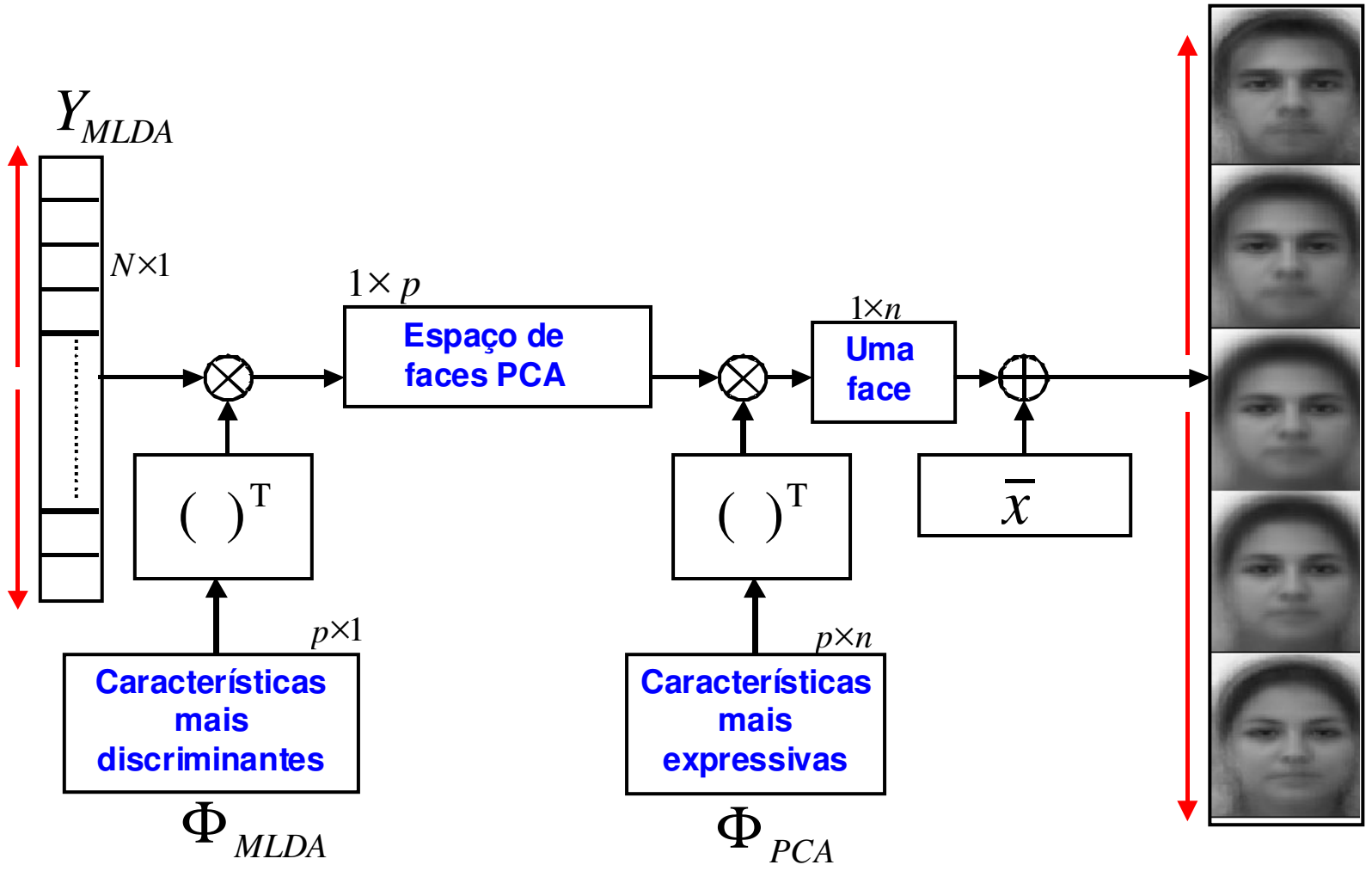


Figura 4.3 - Representação esquemática do fluxo de dados durante a fase de reconstrução da característica mais discriminante. Adaptado de (THOMAZ et al., 2006).

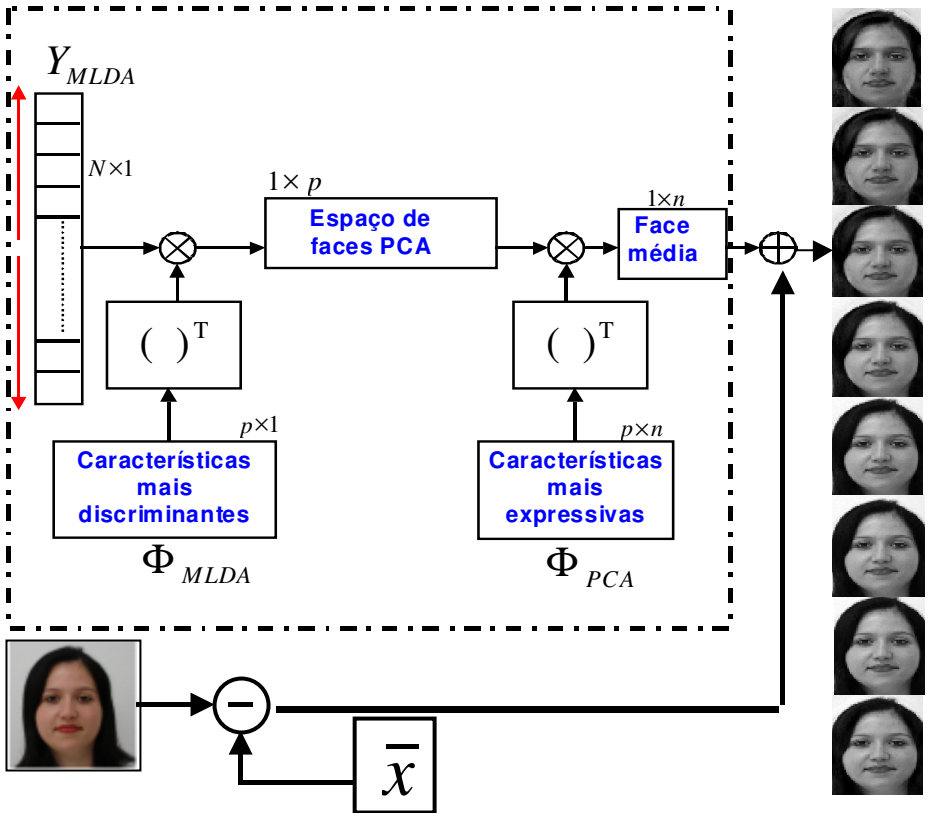


Figura 4.4 - Representação esquemática da fase de incorporação e síntese de características discriminantes em uma imagem de face. Adaptado de (KITANI; THOMAZ; GILLIES, 2006b)

.

4.2 Comparação entre as Abordagens AAM, Buchala et al. e SDM

Como descrito na seção 2.5.2, a abordagem AAM (COOTES et al., 1995) tem por objetivo modelar e sintetizar as variações capturadas de imagens de face e apresentá-las visualmente. Com o mesmo objetivo, o SDM também captura e sintetiza as variações tais como aquelas capturadas pelo AAM. As diferenças fundamentais entre as abordagens AAM (COOTES et al., 1995), Buchala *et al*. (BUCHALA et al., 2005), e o SDM são apresentadas a seguir.

Na abordagem AAM o modelo de forma é extraído a partir de marcas denominadas *landmarks* posicionadas manualmente nas imagens de face. As marcas são desenhadas sobre as fotos do conjunto de treinamento de modo que, em cada foto, cada marca esteja posicionada o mais próximo possível de uma característica comum a todas as imagens de teste. Alinham-se todas as fotos utilizando-se alguma coordenada em comum, calcula-se a imagem média da forma, realinham-se todas as fotos baseadas agora nessa nova forma média, e então repete-se o ciclo, até que todas as imagens convirjam. Aplica-se a técnica PCA para a extração de dados, estuda-se a variação sobre o que ocorre com cada ponto em comum determinando-se a possibilidade de haver alguma correlação entre elas. É importante destacar que as *landmarks* devem ser posicionadas sobre as características da face que melhor descrevam a forma da face. Observa-se ainda que a técnica AAM trabalha sobre um modelo de caricatura obtido pelas marcas de referência o que, na realidade, representa a forma da face. A abordagem SDM, no entanto, não faz uso de nenhuma marca previamente feita nas imagens de treinamento, apenas captura essas variações através da diferença de níveis de cinza entre cada ponto de uma imagem com relação ao mesmo ponto de outras imagens, e esta captura é executada estatisticamente, utilizando-se as autofaces, conforme descrito em (TURK; PENTLAND, 1991).

Na abordagem AAM, o modelo de variação da textura da imagem a partir dos níveis de cinza é obtido fazendo-se uma deformação de cada imagem de treinamento de modo que eles ajustem cada ponto de controle (*landmarks*) com os pontos da forma média. Faz-se uma normalização nos níveis de cinza para se evitar influências da variação de iluminação, e depois aplica-se o PCA para se extrair as características mais significativas da imagem. Na abordagem SDM, também é necessário um alinhamento das imagens de faces. A partir de uma face qualquer que possa ser considerada como sendo um padrão, ajustes de escala, translação e rotação são realizadas em todas as faces, para que elas estejam geometricamente

alinhadas com a face padrão escolhida. Depois aplica-se o PCA para se reduzir a dimensionalidade e extrair as características mais significativas.

A abordagem AAM extrai e interpola as variações de pose e de expressão a partir dos dois modelos (forma e textura) e busca explicar como seria uma nova face gerada a partir da modificação dos parâmetros de forma e textura. Partindo-se de uma nova imagem de face, o qual é previamente marcado nas mesmas coordenadas das *landmarks* das faces de treinamento, obtém-se uma matriz de parâmetros dessa imagem. Combinando-se esses parâmetros com o modelo geral de forma e textura, reconstrói-se uma nova imagem de face, e espera-se que esta nova imagem seja mais próxima possível da imagem real. A abordagem SDM não busca necessariamente modelar perfeitamente as variações de forma e textura, mas parte da hipótese de que pode haver uma correspondência linear entre as variações extraídas de duas classes distintas, não importando qual seja a característica a ser analisada. Assim, a modelagem SDM extrai essa correspondência e depois apresenta visualmente como ocorrem essas variações sobre a face média.

A abordagem proposta em (BUCHALA et al., 2005) busca compreender como as componentes principais representam e modelam as diferentes características de um conjunto de imagens de faces. De maneira semelhante ao SDM, a abordagem de Buchala *et al.* necessita de um alinhamento de todas as imagens de face em relação a um referencial. No entanto, a abordagem de Buchala *et al*. exige também a equalização do histograma, com o objetivo de se reduzir os efeitos das variações na iluminação. A abordagem SDM considera não ser necessária a equalização do histograma, pois informações importantes para discriminação podem ser atenuadas durante esse processo. Entretanto, o aspecto mais relevante é que o trabalho de Buchala *et al.* investiga principalmente o significado das componentes principais para a determinação de gênero, idade e etnia, indicando que determinadas componentes principais podem capturar e caracterizar essas diferenças. De forma análoga, o SDM também permite a exploração das componentes principais para fins de determinação de gênero, idade e etnia. Entretanto o trabalho de Buchala *et al.* não investiga a possibilidade de sintetizar a informação mais discriminante, ao passo que o SDM tem isto como objetivo principal, ou seja, investigar as características mais discriminantes capturadas por um classificador, verificar o poder de generalização do classificador, e sintetizar essas características capturadas.

Durante o desenvolvimento desta dissertação, juntamente com a abordagem SDM foi desenvolvido um aplicativo, denominado FACES2.EXE, para auxiliar nas pesquisas e validar experimentalmente todas as questões levantadas ao longo deste trabalho. No apêndice “A” o leitor encontrará uma breve descrição sobre o aplicativo, bem como parte do código fonte e as *interfaces* com o usuário.

A tabela 1 a seguir resume as principais características consideradas nas três abordagens e mostra uma comparação entre elas.

Tabela 1 - Tabela comparativa entre as abordagens AAM, Buchala et al. e SDM .

CARACTERÍSTICAS	AAM	BUCHALA et al.	SDM
Necessita alinhamento	Sim	Sim	Sim
Entrada supervisionada[6]	Não	Sim	Sim
Necessita *Landmarks*	Sim	Não	Não
Aplica método de Procrustes	Sim	Não	Não
Aplica método de Triangularização	Sim	Não	Não
Modelos distintos de forma e textura	Sim	Não	Não
Sintetiza expressões	Sim	Sim	Sim
Sintetiza as variações capturadas pelas componentes principais	Sim	Sim	Sim
Apresenta as diferentes representações de cada componente principal (gênero, etnia, idade, etc)	Não	Sim	Sim
Captura e sintetiza visualmente a característica mais discriminante	Não	Não	Sim

4.3 Considerações finais

Este capítulo descreveu em detalhes como a abordagem SDM trabalha e que tipo de resultados são esperados. De maneira complementar, foram apresentados as comparações entre as três abordagens que mais se aproximam em termos de objetivos e resultados ao SDM. Observou-se que o SDM incorpora as características das duas abordagens e expande as possibilidades de uso das informações capturadas de um conjunto de treinamento. No próximo capítulo serão apresentados os resultados dos experimentos com o SDM utilizando-se um banco de faces que foi criado e mantido pelo Departamento de Engenharia Elétrica da FEI.

[6] Refere-se à formação dos conjuntos de treinamento em classes distintas, por exemplo masculino e feminino.

5 EXPERIMENTOS E RESULTADOS

Neste capítulo serão discutidos os resultados obtidos durante os experimentos realizados com as imagens de faces utilizando-se a abordagem SDM e comparando-as com os mesmos experimentos realizados com o PCA. Inicialmente descreve-se o banco de faces utilizado e como foram formados os conjuntos de treinamento.

5.1 Banco de Faces e Conjuntos de Treinamento

Para este trabalho utilizou-se um banco de faces criado e mantido pelo Departamento de Engenharia Elétrica da FEI[7], composto de fotos de 200 voluntários, formados por professores, funcionários e alunos daquela instituição. As fotos foram tiradas entre Junho de 2005 a Junho de 2006 no Laboratório de Inteligência Artificial da FEI em São Bernardo do Campo. Este banco de faces é composto de 14 fotos de cada voluntário, totalizando 2800 imagens.

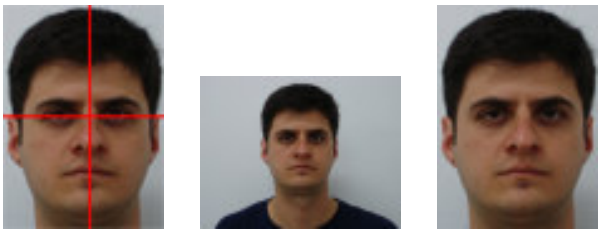
As imagens são coloridas e com uma resolução de 640×480 *pixels*, tiradas contra um fundo branco e homogêneo. As 14 fotos de cada pessoa são fotos frontais e de perfil, variando de 0^o a 180^o. Na figura 5.1 vemos uma amostra do banco de faces da FEI onde cada foto do banco de faces da FEI possui um nome de arquivo, que é formado da seguinte maneira: seja aaa o número da pessoa no banco de faces, bb a pose da pessoa, então um nome de arquivo é formado pela composição do número da pessoa no banco e sua pose, mais a extensão do arquivo. Como pode ser visto na figura 5.1, a foto de uma pessoa é referenciada pelo seu nome de arquivo (aaa-bb.jpg) indicando sua posição no banco de faces (aaa) e a pose (bb) (OLIVEIRA JR; THOMAZ, 2006).

Figura 5.1 - Amostra do banco de faces da FEI com os nomes de arquivo.

[7] Imagens disponíveis sob solicitação para cet@fei.edu.br.

No entanto, para esta dissertação foram selecionadas apenas as fotos masculinas e femininas frontais com expressão neutra e expressão sorrindo. Para se formar os conjuntos de treinamento, as fotos selecionadas foram alinhadas na linha dos olhos e no centro do nariz, ajustadas na mesma escala e removido uma parte do fundo. Com a remoção de uma parte do fundo, observou-se que as imagens ajustadas tinham agora uma dimensão de 260×360 *pixels*, e assim considerou-se que essa seria a dimensão padrão para todo o banco de faces. Os alinhamentos foram executados apenas para corrigir pequenas transformações Euclidianas (translação e rotação) e ajustes de escala, que não necessariamente estavam relacionadas com as diferenças entre as faces, as quais se desejava capturar. Ainda, com o intuito de se reduzir o esforço computacional, as imagens foram reduzidas em sua resolução para 64×64 *pixels* e tornadas monocromáticas em 8 *bits* de nível de cinza. A resolução de 64×64 *pixels* é hoje reconhecidamente a que fornece o melhor equilíbrio entre o custo da qualidade de imagem pelo consumo de memória (CAMPOS, 2001). No entanto, resoluções e profundidades em nível de cinza menores também podem ser utilizadas (THOMAZ; GILLIES, 2005).

A figura 5.2 (b) ilustra a face que foi utilizada como referência para o alinhamento de todas as outras imagens. Considerando a linha dos olhos e a linha central do nariz, conforme pode ser visto na figura 5.2 (a), todas as imagens sofreram transformações Euclidianas para se ajustar a essa imagem de referência.



(a) **(b)** **(c)**

Figura 5.2 - Figura (a) representa a face que foi utilizada como padrão para alinhamento frontal, na figura (b) vemos a foto original com resolução de 640×480 *pixels* e na figura (c) a foto padrão reduzido na resolução para 260×360 *pixels*. Adaptado de (OLIVEIRA JR.; THOMAZ, 2006).

Na figura 5.3 temos todas as fases de alinhamento de uma face baseado no alinhamento da face padrão (*template*). Este processo foi repetido para todas as fotos de imagens frontais masculinas, femininas, sorrindo e não sorrindo. O leitor interessado em obter maiores detalhes sobre o banco de faces da FEI, ou sobre o processo de alinhamento, deve consultar (OLIVEIRA JR.; THOMAZ, 2006).

Para a formação do primeiro conjunto de treinamento (grupo “a”) utilizou-se as fotos frontais com expressão neutra das 100 faces femininas e 100 faces masculinas. O segundo

conjunto (grupo “b”) foi formado por 100 faces masculinas com expressão neutra e 100 faces masculinas com expressão sorrindo, das mesmas pessoas.

(a) **(b)** **(c)** **(d)** **(e)**

Figura 5.3 - Na seqüência da figura (a) até a figura (e) observamos todo o processo de alinhamento manual que foi realizado sobre a foto da figura (a). Adaptado de (OLIVEIRA JR.; THOMAZ, 2006).

O terceiro conjunto de treinamento (grupo “c”) foi formado com 100 faces masculinas com expressão neutra e 100 faces femininas com expressão sorrindo e finalmente o ultimo grupo (grupo “d”) foi formado com 100 faces masculinas de expressão neutra e 100 faces femininas de expressão neutra. Entretanto para este último conjunto de treinamento a leitura das imagens de face foi executada de forma alternada, de 10 em 10. Foram lidas 10 faces masculinas e depois 10 faces femininas, e assim sucessivamente, até completar 200 faces na matriz **Z**. Este conjunto de treinamento servirá para se avaliar o comportamento das abordagens PCA e MLDA em relação a conjuntos que efetivamente não têm uma distribuição Gaussiana. A alteração na distribuição foi executada de modo sintético. Quando se lêem as imagens de face de forma alternada (homens e mulheres), espera-se uma distribuição multi-modal em cada classe.

Grupo a

Grupo b

Figura 5.4 - No grupo “a”, temos amostras das 100 faces femininas e 100 masculinas, no grupo “b”, amostras das 100 faces masculinas neutras e 100 faces masculinas sorrindo.

As fotos que formaram o conjunto de treinamento são principalmente de jovens entre 19 a 30 anos e com uma grande variação na aparência, estilo de cabelo e adornos. Uma pequena amostra dos dois primeiros conjuntos de treinamento pode ser visto na figura 5.4, onde as fotos têm resolução de 64×64 *pixels*. Nas próximas seções são apresentados os experimentos e os respectivos resultados.

5.2 Experimentos com PCA

Os experimentos com o PCA têm por objetivo investigar quais informações visuais são sintetizadas quando navegamos ao longo das três primeiras componentes principais. Assim como descrito em (BUCHALA et al., 2005), alguma outra informação mais significativa pode estar contida nas componentes principais e não somente aquelas relativas à reconstrução de uma imagem de face com o mínimo de erro. Cootes *et al*. (COOTES et al., 1994) também investigaram a capacidade das componentes principais de modelarem variações de forma, desde que o conjunto de variações permitidas pelos exemplos estivessem contidas em um conjunto de treinamento. No entanto, a abordagem de (COOTES et al., 1994) exige um intenso trabalho de pré-processamento, criando previamente os marcos de referência nas imagens de teste. Para os experimentos com o PCA foram utilizados os quatro conjuntos de treinamento formados conforme descrito na seção anterior.

A navegação ao longo de cada componente principal significa variar o parâmetro b_i da equação (5.57) dentro dos limites de $\pm 3\sqrt{\lambda_i}$, onde λ_i é o respectivo autovalor, e reconstruir visualmente essas variações.

$$x_i = \bar{\mathbf{x}} + \varphi_i^T b_i, \tag{5.57}$$

Como as componentes principais estão ordenadas decrescentemente segundo a sua variância, definida pelo autovalor λ_i, reconstruir visualmente cada componente principal pode trazer informações mais claras sobre exatamente o que cada componente capturou como informação mais expressiva. A equação (5.57) expressa como se processa a navegação ao longo do eixo da *i-ésima* componente principal, onde $\bar{\mathbf{x}}$ é a face média global e b_i o parâmetro que será variado para que se percorra o eixo da respectiva componente principal.

Com o intuito de melhorar a compreensão sobre o processo de navegação nos eixos das componentes principais, é apresentado na figura 5.5 um gráfico de uma elipse que representa ilustrativamente uma área de distribuição de pontos constante formada pelas duas primeiras componentes principais, e limitada a $2\sqrt{\lambda_i}$, onde $i = 1, 2$, isto porque a variância original de cada componente principal foi normalizada por $\sqrt{\lambda_i}$ conforme indicado na equação (3.33). Como cada componente principal representa uma autoface, percorrendo o eixo dessa autoface é possível visualizar todos os modos de variação que essa autoface representa ou captura. Essas variações em torno de uma determinada componente principal representam o grau de generalização da autoface e também as possíveis características capturadas e passíveis de serem representadas.

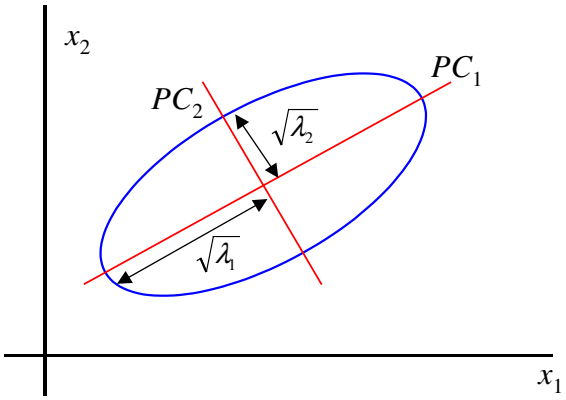


Figura 5.5 - Gráfico que representa uma densidade de distribuição constante formado pelas duas primeiras componentes principais. Adaptado de (FUKUNAGA, 1990).

Todos os experimentos foram executados tomando-se como referência a imagem da face média de cada conjunto de treinamento. A razão fundamental da escolha da face média recai sobre o fato já mencionado na seção 3.1.1 de que a face média carrega as informações que são comuns a todas faces do conjunto de treinamento. Como todas as faces do conjunto de treinamento são uma variação da face média, qualquer alteração na face média será incorporada por todas as outras faces.

Os experimentos com o PCA objetivam avaliar o poder de generalização e discriminação que as três primeiras componentes principais possuem. Para a generalização, observaremos visualmente quais características faciais cada componente principal modela e para determinar o poder de discriminação utilizaremos uma função critério. Em (JOLLIFFE, 2002), o autor discute algumas propostas de se usar as componentes principais para encontrar as direções discriminantes. A maior dificuldade, entretanto, é determinar uma métrica que indique o potencial de discriminação de cada componente principal. Chang (CHANG, 1983) propôs uma função critério denominada Δ_q^2, e que mede a distância quadrática de Mahalanobis em relação a duas populações distintas π_1 e π_2. A função Δ_q^2 pode ser calculada segundo a equação (5.58) a seguir

$$\Delta_q^2 = \left[\left\{ \sum_{k=1}^{q} \frac{\left(\varphi_k^T \delta\right)^2}{\lambda_k} \right\} - m(1-m) \right]^{-1}, \qquad (5.58)$$

onde, φ_k e λ_k são os autovetores e autovalores da matriz de covariância Σ, $\delta = (\mu_1 - \mu_2)$ que é a diferença entre as médias das duas populações π_1 e π_2 que formam cada conjunto de treinamento, $k = 1,2,...,p$, m é proporção da mistura das duas distribuições normais em relação às médias μ_1 e μ_2, e q é o número das primeiras componentes principais que se deseja avaliar. No entanto, em (JOLLIFFE et al., 1996) os autores discordam que o método proposto

por (CHANG, 1983) seja o mais adequado para a determinação das melhores componentes principais para fins de discriminação. Isto porque a função critério Δ_q^2 depende de um conhecimento a *priori* de como está composta a mistura das densidades de distribuições e também porque não determina o quanto cada componente principal contribui para encontrar a melhor direção de discriminação. Em (KSHIRSAGAR et al., 1990) foi proposto uma outra função critério que determina a contribuição de cada componente principal dentro do processo de discriminação. Os autores propõem o uso de uma função critério denominada θ_k^2 que é determinada por

$$\theta_k^2 = \frac{\left(\varphi_k^T \left(\bar{x}_1 - \bar{x}_2\right)\right)^2}{\lambda_k} \ , \tag{5.59}$$

onde φ_k e λ_k são os autovetores e autovalores da matriz de covariância Σ, $\bar{x}_1$ e $\bar{x}_2$ são as médias das duas populações π_1 e π_2 que formam cada conjunto de treinamento, $k = 1,2,...,p$, e p é o número de componentes principais que se deseja avaliar. A idéia fundamental é construir uma base vetorial, ordenando os autovetores de modo que $\theta_1^2 > \theta_2^2 > ... > \theta_k^2$. Esta dissertação utiliza então a função critério θ_k^2 para avaliar as componentes principais em relação à sua capacidade de apontar a direção de maior discriminância.

5.2.1 Experimento 1 com o PCA

Neste experimento foi utilizado o primeiro conjunto de treinamento (grupo "a") composto de 100 faces frontais femininas com expressão neutra e 100 faces frontais masculinas com expressão neutra. O conjunto foi treinado no módulo do PCA, conforme descrito na seção 4.1 e os resultados visuais podem ser acompanhados na figura 5.6. Percorrendo os eixos das três primeiras componentes principais através da equação (5.57) e limitando o parâmetro b_i dentro de $\pm 3\sqrt{\lambda_i}$, é possível observar visualmente as variações que cada componente principal capturou. Este experimento visa avaliar como o PCA captura e trata informações de amostras que têm classes bem definidas e distintas.

A primeira componente principal modelou principalmente variações na iluminação, seja em torno da face, quanto dentro da face, e essas variações podem ser compreendidas como a quantidade de cabelo em torno da face e a cor da textura da face. Observa-se que a primeira componente também modela o tamanho e formato da face, e diferentes identidades de faces no lado masculino. A segunda componente principal capturou informações que modelam mais as partes internas do rosto e a cor do cabelo, mas não é muito claro o que esta

componente está capturando. A terceira componente capturou variações no tamanho e o formato da cabeça, mas observa-se também que a determinação do gênero já não é tão nítida, apesar do conjunto de treinamento ter uma separação bem clara entre as classes.

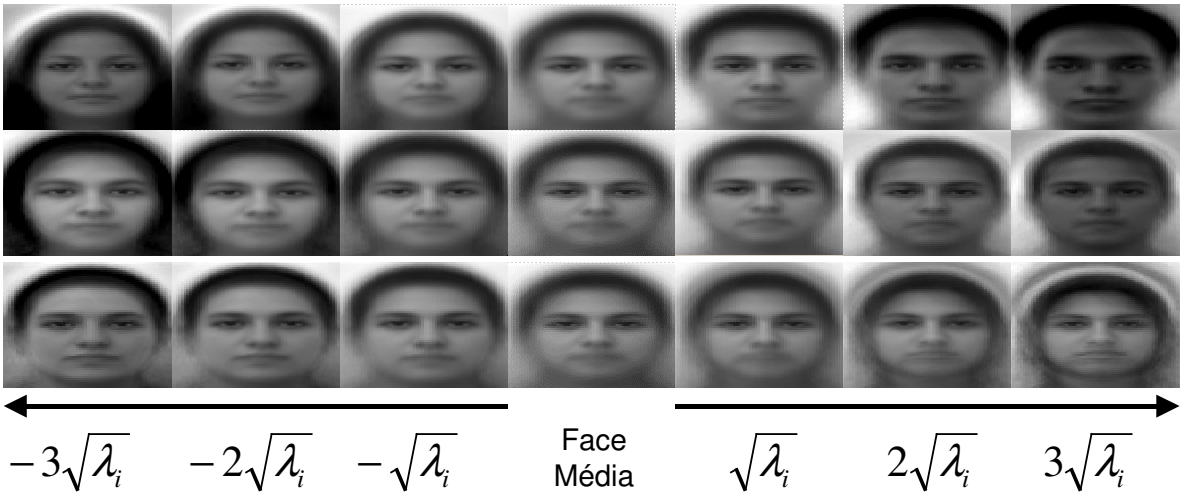


Figura 5.6 - Reconstruções visuais variando-se as três primeiras componentes principais utilizando-se o conjunto de treinamento do grupo "a". Do alto da figura para baixo e na seqüência, a primeira, segunda e terceira componentes principais.

Analisando-se visualmente as imagens sintetizadas neste experimento observa-se que a característica mais expressiva capturada foi a variação do gênero que existe entre as duas classes. Conclui-se que se as classes têm informações discriminantes na mesma direção das informações mais expressivas, o PCA aponta para a direção de maior discriminação. Em outras palavras, navegando-se ao longo da primeira componente principal, observaremos informações visuais que permitem descrever uma face como sendo definitivamente feminina ou definitivamente masculina, ou seja, o que efetivamente varia entre as duas classes do conjunto de treinamento.

O gráfico da figura 5.7 representa a distribuição dos pontos do espaço PCA em relação às duas primeiras componentes principais. Observa-se que há uma separação visível entre os pontos projetados do grupo masculino (quadrados azuis) e feminino (círculos vermelhos), indicando que a navegação ao longo da primeira componente principal permite mapear as características que descrevem cada um dos grupos. Todavia, a navegação pela segunda componente principal não permite caracterizar claramente as diferenças entre os dois grupos, assim como também foi observado durante a reconstrução visual. Como a percepção de gênero não é significativa na segunda componente principal, considera-se que o conjunto de características mapeadas pela segunda componente principal pertence aos dois grupos. O gráfico da figura 5.7 mostra que as variações de modo da segunda componente não fazem discriminação, mas mapeam características comuns.

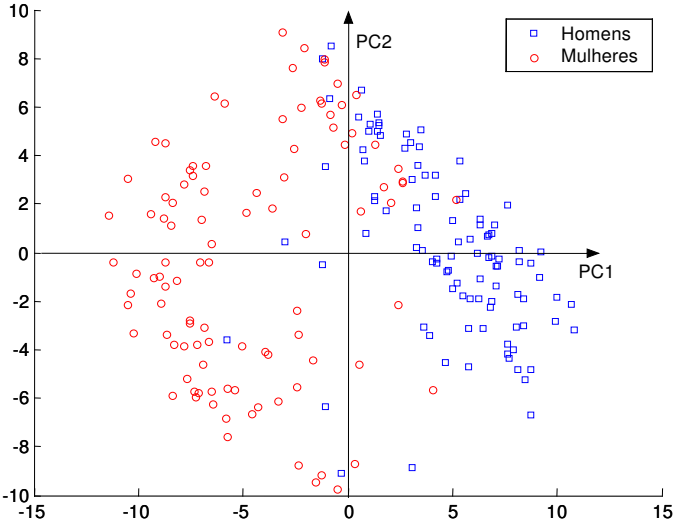


Figura 5.7 - Gráfico tipo *scatter-plot* do grupo "a" homens e mulheres, em relação a primeira e segunda componentes principais (*Principal Components* PC).

Aplicando-se a equação (5.59) para se determinar a função critério das 100 primeiras componentes principais, obtém-se o gráfico apresentado na figura 5.8. Observa-se que a função critério θ_k^2 para a primeira componente principal, ou *Principal Components* (PC), indica que essa componente tem uma alta capacidade de discriminação, assim como pôde ser observado no gráfico da distribuição dos pontos da figura 5.7.

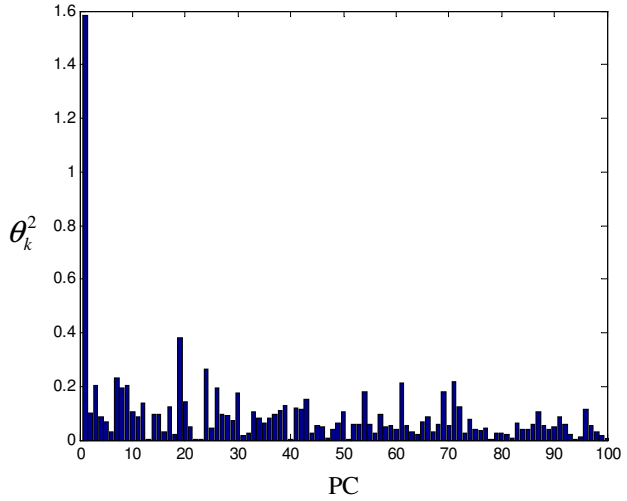


Figura 5.8 - Gráfico da contribuição das 100 primeiras componentes principais para fins de discriminação do grupo "a".

5.2.2 Experimento 2 com o PCA

O experimento 2 foi conduzido utilizando-se o conjunto de treinamento formado pelo grupo "b", onde temos 100 faces frontais masculinas com expressão neutra e 100 faces fron-

tais masculinas com expressão sorrindo. De forma semelhante ao experimento 1, foram reconstruídas as variações que ocorrem dentro das três primeiras componentes principais variando-se o parâmetro b_i. Neste experimento, as amostras já não têm classes com características tão expressivas como nas amostras do grupo "a", ou seja, a informação que distingue uma classe da outra é muito mais sutil. Deseja-se avaliar com este experimento, qual característica o PCA captura como a mais expressiva, se é a variação de identidade ou a expressão facial.

Analisando-se as imagens reconstruídas apresentadas na figura 5.9 observa-se que a primeira componente principal modela o tamanho da face, a cor e a quantidade do cabelo, a cor da textura da face e uma pequena variação na identidade da face. A segunda componente capturou variações no tamanho da testa, tamanho da face e variações na identidade da face. A terceira componente capturou variações no formato do cabelo e na cor da textura da face. De um modo geral, todas as componentes capturam informações sobre tamanhos e formatos de face, cor da pele, formato e quantidade de cabelo, variações na identidade e variações na etnia. Entretanto, a informação mais discriminante que distingue a diferença entre as duas classes do conjunto de treinamento não foi capturada. Em outras palavras, o PCA não detectou a variação sutil que existe entre as duas classes, isto é, a variação na expressão das faces, neste caso sorrindo e não sorrindo. Isto ocorre porque neste caso a direção da informação mais discriminante não está na mesma direção da informação mais expressiva ou que tem a maior variabilidade.

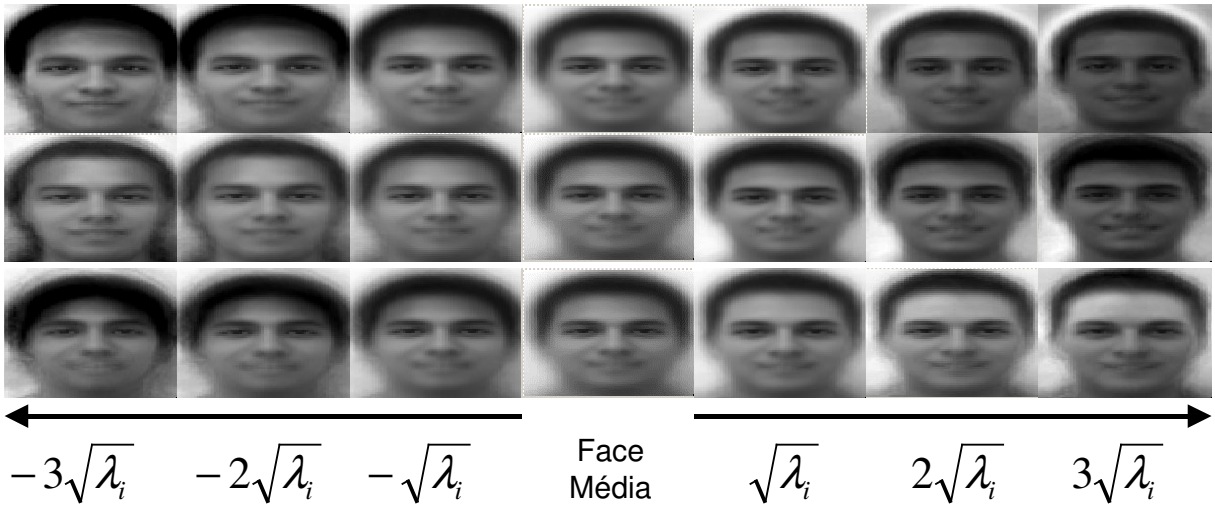


Figura 5.9 - Reconstruções visuais variando-se as três primeiras componentes principais utilizando-se o conjunto de treinamento do grupo "b". Do alto da figura para baixo e na seqüência, a primeira, segunda e terceira componentes principais.

De maneira análoga ao primeiro experimento, a figura 5.10 apresenta o gráfico da projeção do grupo "b" no espaço PCA com relação às duas primeiras componentes principais. Pelo gráfico da figura 5.10 não é possível definir através da primeira componente principal

como se discriminam os dois grupos. A primeira componente modela mais as características externas da face, como o cabelo e o tamanho da face. E essas características são observadas em todas as amostras do conjunto de treinamento. A segunda componente principal mostra um espalhamento das amostras que não permite uma separação linear entre as duas classes.

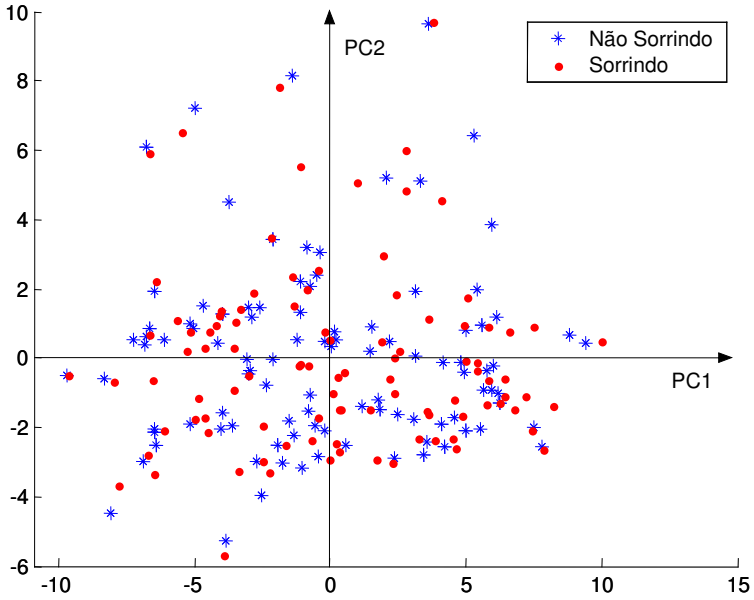


Figura 5.10 - Gráfico tipo *scatter-plot* do grupo "b" homens sorrindo e não sorrindo, em relação a primeira e segunda componentes principais.

O gráfico da figura 5.11 mostra o resultado do cálculo da função discriminante para as 100 primeiras componentes principais, onde observa-se que as três primeiras componentes principais têm pouco ou nenhum poder de discriminação. Observa-se também que as componentes principais que têm o melhor índice θ_k^2 estão entre as componentes de número 10 e 30.

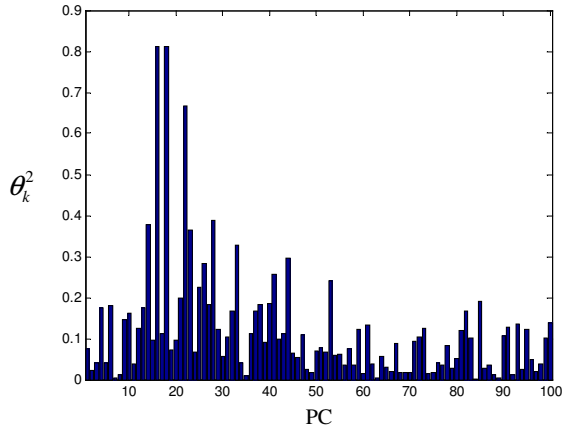


Figura 5.11 - Gráfico da contribuição das 100 primeiras componentes principais para fins de discriminação do grupo "b".

5.2.3 Experimento 3 com o PCA

O terceiro experimento trabalha com o conjunto do grupo “c”, formado por 100 faces frontais masculinas com expressão neutra e 100 faces frontais femininas com expressão sorrindo. Este experimento tem o objetivo de produzir uma mistura de duas informações discriminantes, onde o conjunto de faces femininas juntamente com as masculinas produz a informação discriminante de gênero. Por outro lado, as faces femininas têm expressão sorrindo e as faces masculinas têm expressão neutra, fornecendo uma informação de discriminação bem sutil. Deseja-se determinar qual informação o PCA considera a mais expressiva. Observa-se que este conjunto de treinamento tem misturado quatro classes distintas de informações, masculino/feminino e sorrindo/não sorrindo.

Durante a navegação ao longo das três componentes principais o PCA retornou informações visuais semelhantes àquelas do experimento 1. Entretanto observa-se que as imagens de face média também incorporaram a expressão sorrindo, mesmo nas faces que seriam do grupo masculino, e que não necessariamente apresentam essa expressão. A segunda e terceira componentes principais incorporaram a expressão sorrindo ao longo de seus eixos, fazendo com que mesmo as imagens de face média do lado masculino incorporassem essa informação. Em outras palavras, na segunda e terceira componentes principais, a informação do sorriso foi distribuída ao longo das duas componentes principais.

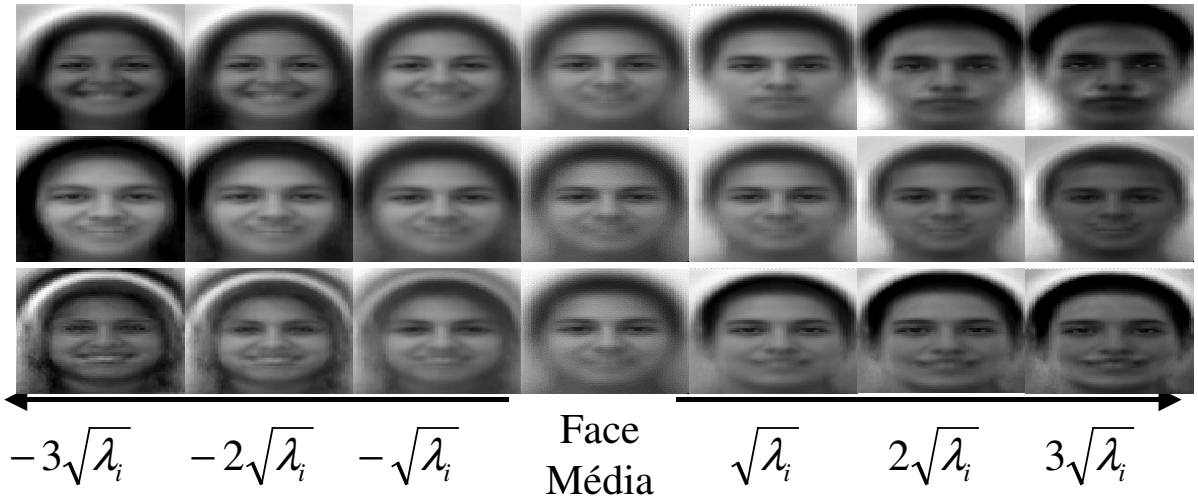


Figura 5.12 - Reconstruções visuais variando-se as três primeiras componentes principais utilizando-se o conjunto de treinamento do grupo “c”. Do alto da figura para baixo e na seqüência, a primeira, segunda e terceira componentes principais.

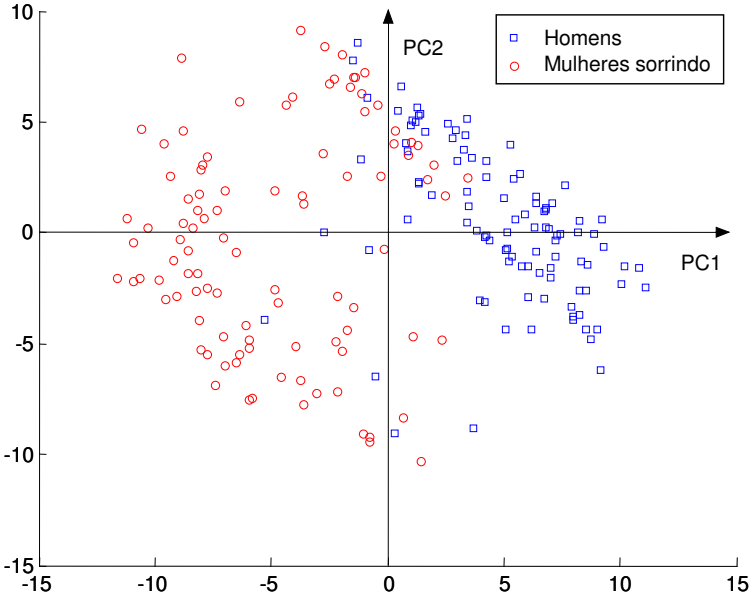


Figura 5.13 - Gráfico tipo *scatter-plot* do grupo "c" homens não sorrindo e mulheres sorrindo, em relação a primeira e segunda componentes principais.

O gráfico da figura 5.13 mostra que a distribuição das amostras no espaço PCA em relação à primeira e segunda componentes principais é muito semelhante àquela apresentada na figura 5.7 pelo grupo "a" com homens e mulheres com expressão neutra. A direção da primeira componente principal permite a separação entre as classes de uma maneira mais clara. Como já foi descrito acima, a característica das expressões sorrindo e não sorrindo também foram capturadas pela primeira componente principal, entretanto, os sorrisos sintetizados pela primeira componente principal não mostram uma transição gradual entre um sorriso leve até um mais intenso, como foi descrito na seção 3.1.2.

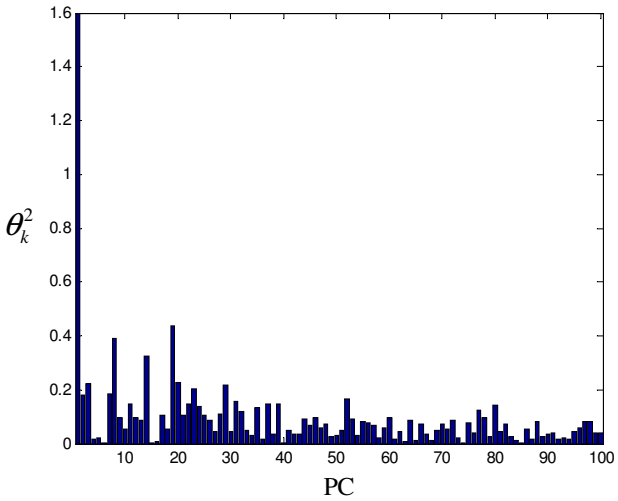


Figura 5.14 - Gráfico da contribuição das 100 primeiras componentes principais para fins de discriminação do grupo "c".

O gráfico da figura 5.14 mostra o resultado do cálculo da função discriminante para as 100 primeiras componentes principais, indicando que a primeira componente principal tem uma alta capacidade discriminatória.

5.2.4 Experimento 4 com o PCA

Neste último experimento com o PCA pretende-se avaliar o comportamento do PCA quando o conjunto de treinamento (grupo "d") não apresenta efetivamente uma densidade de distribuição Gaussiana. Os resultados podem ser vistos na figura 5.15, onde observa-se que os resultados são semelhantes aos obtidos pelo experimento 1, diferindo apenas no sentido da reconstrução das autofaces. Os autovetores retornaram as mesmas imagens do experimento, quando navegou-se ao longo de cada autovetor. Os resultados demonstram a capacidade do PCA em extrair informações significativas e independentes de como as amostras se distribuem dentro do conjunto de treinamento, como já se conhecia teoricamente esse resultado.

O gráfico da figura 5.15 representa uma distribuição de pontos muito semelhantes ao gráfico da figura 5.7 anterior, sendo quase que um espelho daquela distribuição. Analogamente ao resultado do experimento 1, a segunda componente principal também não apresenta potencial para discriminação da informação.

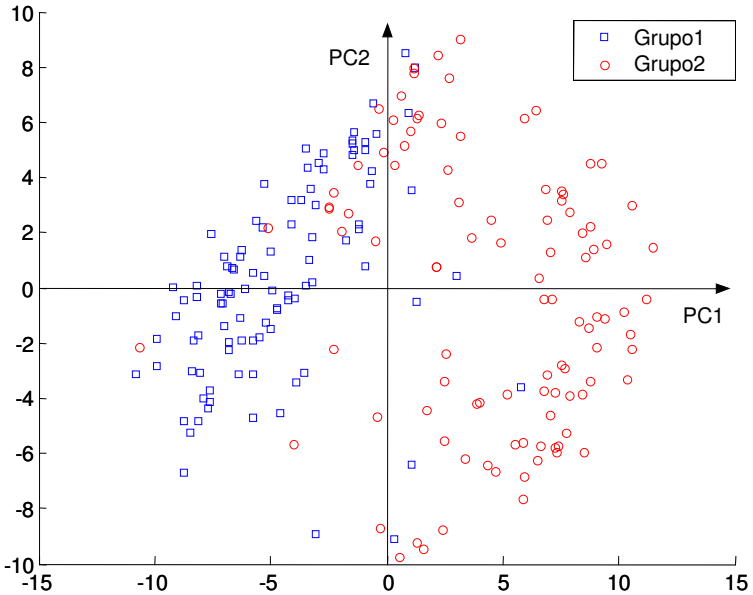


Figura 5.15 - Gráfico tipo *scatter-plot* do grupo "d" em relação a primeira e segunda componentes principais.

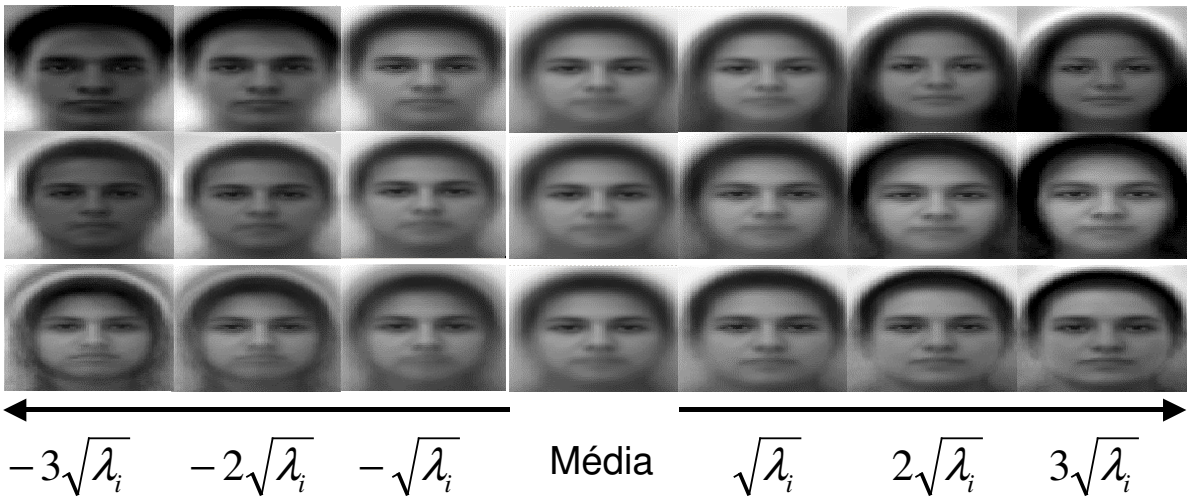


Figura 5.16 - Reconstruções visuais variando-se as três primeiras componentes principais utilizando-se o conjunto de treinamento do grupo "d". Do alto da figura para baixo e na seqüência, a primeira, segunda e terceira componentes principais.

O gráfico da figura 5.17 apresenta o cálculo da função discriminante, e nele é possível observar que o cálculo proposto por (KSHIRSAGAR et al., 1990) somente é útil quando as populações têm uma densidade de distribuição Gaussiana. Observa-se que quando produzimos uma mistura não Gaussiana no conjunto de treinamento, o cálculo da função critério é afetado. Isto ocorre porque o cálculo de θ_k^2 está diretamente relacionado com a distância entre as médias das populações, e pressupõem-se um distribuição normal.

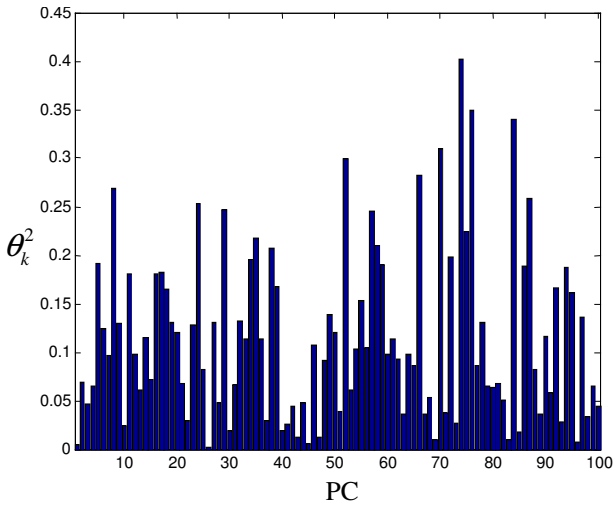


Figura 5.17 - Gráfico da contribuição das 100 primeiras componentes principais para fins de discriminação do grupo "d".

5.2.5 Considerações sobre os Experimentos com o PCA

Os experimentos com o PCA demonstram visualmente que as componentes principais que descrevem as maiores variâncias não trazem informações significativas sobre o que exa-

tamente existe de diferente entre dois conjuntos de treinamento, como demonstrado nos experimentos 2 e 3. Entretanto, os experimentos 1 e 4 demonstraram que a primeira componente principal consegue modelar a diferença entre dois grupos. Isto foi possível porque os dois grupos são muito distintos entre si (masculino e feminino), demonstrando que quando a direção da informação de maior variância coincide com a de maior discriminância o PCA tem o poder de discriminação associado à capacidade de representação. Mesmo quando adicionamos características mais sutis, como aquelas apresentadas no experimento 3, o PCA continua apontando para a direção que discrimina o gênero, porém anexando a informação da expressão do sorriso ao longo das autofaces. Entretanto, quando se apresenta ao PCA conjuntos cujas diferenças são realmente sutis, como o caso do experimento 2, observa-se que o PCA não captura informações discriminantes com muita facilidade. Nas próximas seções serão apresentados os experimentos executados com a abordagem SDM.

5.3 Experimentos com SDM

Os experimentos conduzidos com a metodologia SDM utilizaram os mesmos quatro conjuntos que foram formados para os experimentos com o PCA. Assim como foi descrito na seção 4.1, inicialmente toma-se um conjunto de treinamento e cria-se o espaço de faces PCA. Em seguida, toda a matriz $\mathbf{Y}_{PCA}$ é utilizado pelo MLDA para se executar a extração das características discriminantes, formando-se assim o espaço MLDA. E como foi descrito na seção 3.4, o MLDA formará um vetor W_{mlda} contendo as informações mais discriminantes da matriz $\mathbf{Y}_{pca}$. Os experimentos com o SDM visam principalmente analisar visualmente quais são as características mais discriminantes que um classificador captura, e verificar o grau de generalização. Similarmente aos experimentos com o PCA, percorre-se o autovetor mais discriminante e reconstroem-se visualmente as variações capturadas. Entretanto, o limite da navegação não será aquele definido pelo comprimento do autovalor, como foi o caso do PCA, isto porque o autovalor exprime a densidade de distribuição considerando-se toda a matriz do espaço PCA, e consequentemente todas as amostras. A figura 5.18 representa graficamente esta situação, onde a elipse de densidade constante cobre toda a população, porém, as distribuições de cada classe não estão na mesma orientação. Como deseja-se conhecer com a utilização do SDM o comportamento das populações de cada classe, é conveniente estudar as distribuições de cada classe utilizando-se os dados fornecidos pelas amostras.

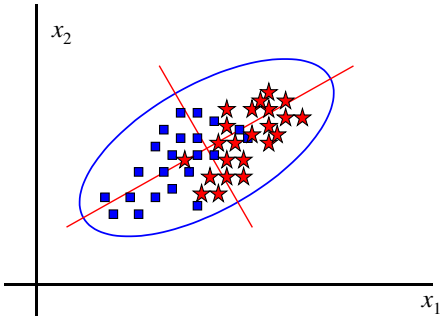


Figura 5.18 - Área de distribuição constante definida pelos autovetores PCA e a distribuição de duas classes.

A determinação dos parâmetros de distribuição de cada classe foi executada da seguinte maneira: projetou-se cada conjunto de classes no espaço MLDA e calculou-se a média e o desvio padrão de cada classe. Em seguida, a navegação ocorreu dentro dos limites de $\pm 2sd$ (*standard deviation*) para cada classe. As próximas seções apresentam os resultados da síntese da imagem de face média, para cada um dos conjuntos de treinamento testados.

5.3.1 Experimento 1 com SDM

De maneira análoga ao experimento 1 com o PCA, o conjunto de treinamento formado por imagens de faces frontais masculinas e femininas com expressão neutra foi apresentado ao SDM. Conforme descrito anteriormente, calculou-se a média e o desvio padrão de cada classe, obtendo-se as duas curvas de distribuição ilustradas na figura 5.20. A figura 5.20 apresenta a reconstrução da imagem de face média global $\bar{x}$ quando navega-se pelo hiper-plano W_{MLDA}, assim como foi descrito na seção 4.1. Observa-se que diferentemente das reconstruções obtidas no PCA, as reconstruções trouxeram variações bem sutis da face média. Não há grandes variações de identidade, mas sim alterações morfológicas que indicam uma transformação da face média para uma face definitivamente feminina (direita) ou para uma definitivamente masculina (imagem mais à esquerda). Observa-se que não há variações significativas de iluminação, ao contrário das imagens sintetizadas pela primeira componente principal.

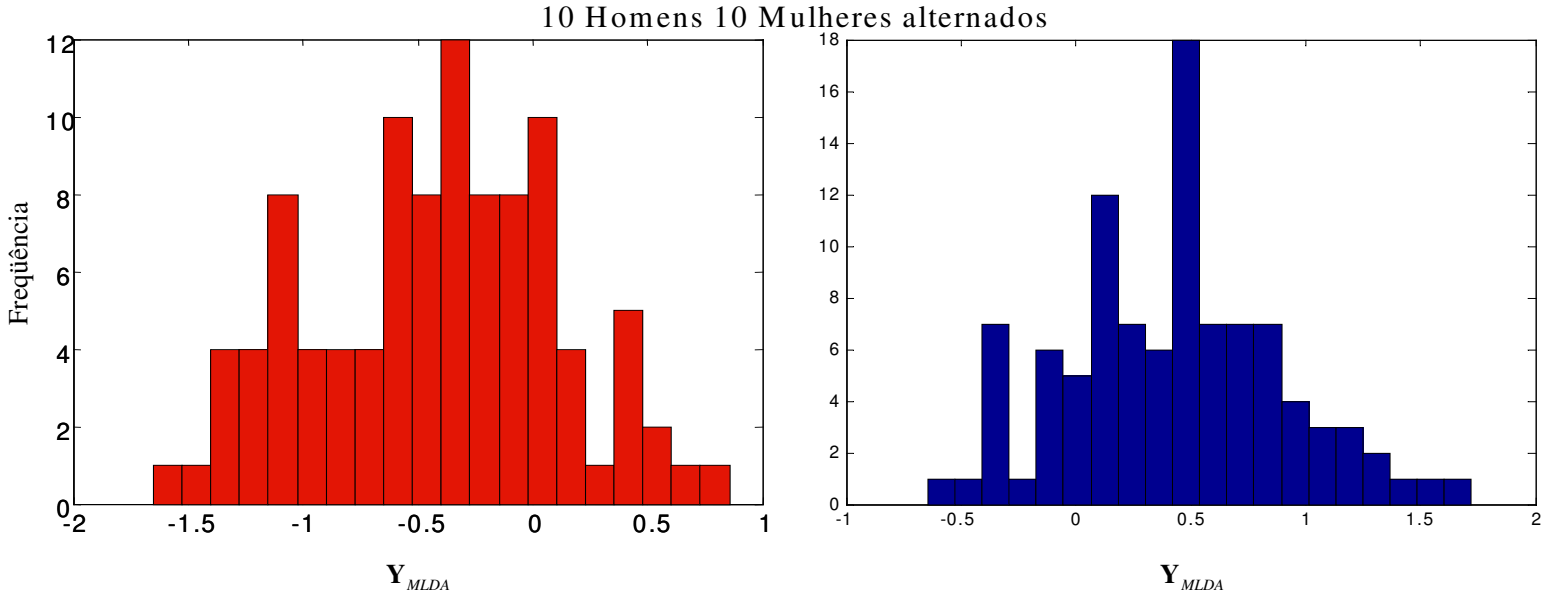


Figura 5.19 - Histograma da distribuição das projeções de 100 faces femininas e 100 masculinas no espaço MLDA.

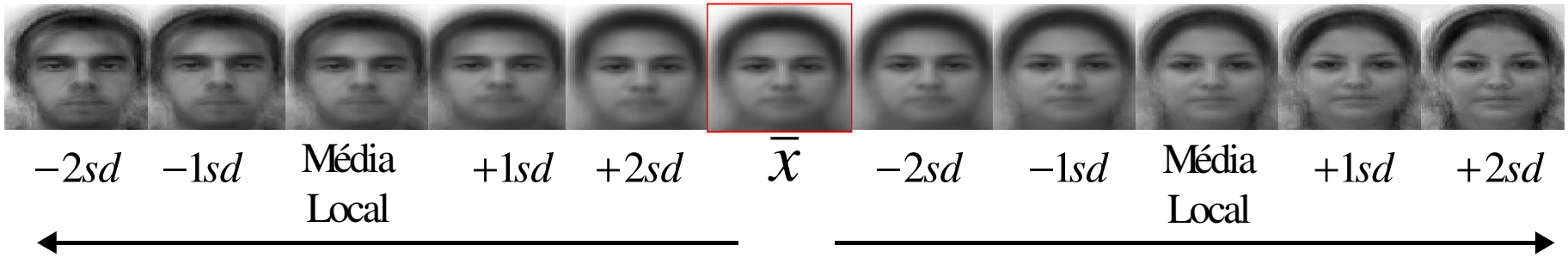


Figura 5.20 - Reconstrução visual da imagem de face média para o grupo “a”.

Entretanto, o que mais varia entre os dois grupos pode ser resumido como:

1) O formato do rosto, no feminino, tende a ser mais arredondado e no masculino, é mais fino na região do queixo;
2) A região da testa, no feminino, é mais rasa e no masculino, é mais alta;
3) A região dos olhos, no feminino, tende a ser mais amendoada e no masculino, é mais funda;
4) O nariz, no feminino, tende a ser menor e no masculino, maior;
5) A boca, no feminino, aparece com mais destaque;
6) A região do cabelo aparece com forma regular nos dois grupos, entretanto no feminino tende a aparecer um ruído abaixo da linha das orelhas, indicando a presença de cabelos ou de algum adorno. No masculino, há também um ruído, mas menor;
7) A região ao redor da boca, no feminino, é mais clara e no masculino, é mais escuro indicando a possível presença de barba.

As descrições acima são baseadas na análise visual, mas que coincidem com o conhecimento a *priori* que temos das características de cada conjunto de treinamento. Entretanto, as descrições anteriores indicam exatamente o que foi capturado pelo classificador como as características mais discriminantes entre os dois conjuntos, e mostram, consequentemente, a capacidade de generalização do método SDM.

Os resultados visuais indicam que o classificador reteve informações que descrevem as diferenças entre os grupos, e não exatamente as características individuais de cada amostra. Esta questão tem relevância caso o conjunto de treinamento tenha dois grupos distintos, mas cujas características não são previamente conhecidas.

A síntese visual pode facilitar ou modificar a forma como se interpretam os resultados de uma classificação. Como as diferenças visuais entre as duas classes são muito significativas, os resultados obtidos são muito próximos ao do PCA, entretanto, o classificador MLDA captura variações sutis que o PCA decompôs ao longo das outras componentes principais. Possivelmente a diferença significativa é que o MLDA agrupa em apenas um autovetor as diferenças entre as duas classes.

A conclusão acima expõe uma outra dúvida. Se as características que discriminam os dois grupos estão distribuídas ao longo das componentes principais, como determinar em quais componentes elas se encontram? A resposta pode estar no vetor de pesos W_{MLDA}. Portanto, cada parâmetro w_i^{MLDA} indicaria o grau de representatividade que cada vetor da base W_{MLDA} tem em termos de poder de discriminação, seja de forma aditiva ou subtrativa. Se investigarmos o que a componente principal, relacionada com o ponto mais discriminante, re-

torna como informação visual poderemos responder a questão acima. O gráfico da figura 5.21 representa em termos absolutos o valor da projeção de cada elemento do vetor W_{MLDA}.

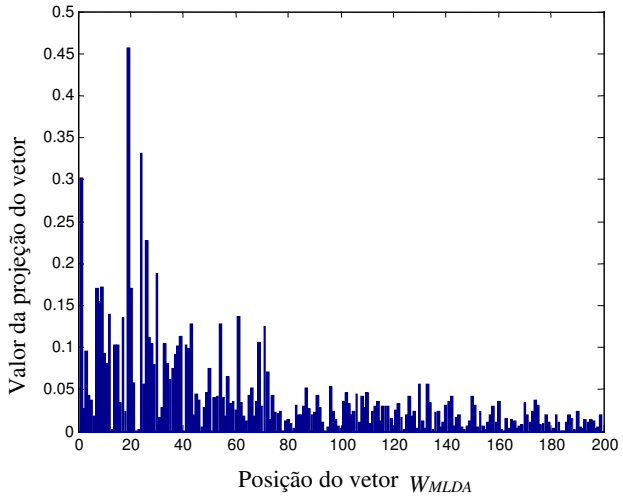


Figura 5.21 - Gráfico que representa em valores absolutos a projeção das amostras do espaço PCA no espaço MLDA para o grupo "a".

Para investigar a hipótese de que podem haver componentes principais que guardam informações discriminantes, tomou-se os índices do vetor W_{MLDA}, cujos valores são os três maiores e relacionou-os com as respectivas componentes principais, que neste caso são as componentes de número 19, 24 e 1. Repetiu-se o mesmo experimento executado no experimento 1 do PCA, porém navegando-se nos autovetores de número 19, 24 e 1, e obteve-se os resultado vistos na figura 5.22.

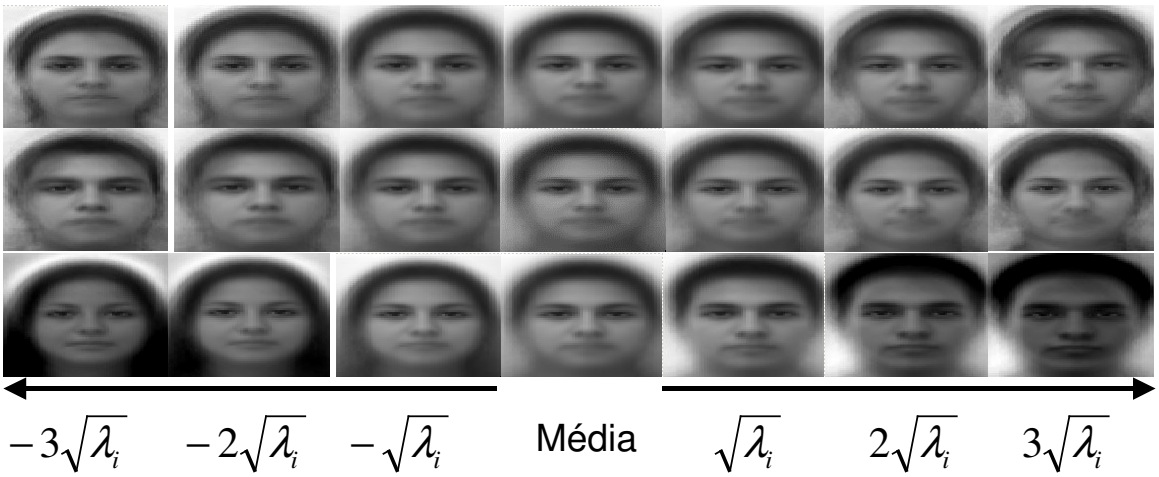


Figura 5.22 - Reconstrução visual da componentes principais de números 19, 24 e 1 do grupo "a", do alto para baixo.

É interessante observar o porquê a primeira componente principal, que no PCA é considerada a de maior variância, teve sua importância reduzida para a terceira posição. Na reali-

dade, para o PCA, a primeira componente principal capturou variações significativas sob o ponto de vista de representação, e nessa componente há uma variação muito grande de iluminação, tamanho e tipo de cabelo. Entretanto, sob o ponto de vista da discriminação, a iluminação não é tão relevante para o classificador, inclusive o trabalho apresentado por (BELHUMEUR et al., 1997) demonstrou que o LDA é muito mais robusto que o PCA, quando o conjunto de treinamento apresenta forte variação na iluminação.

5.3.2 Experimento 2 com SDM

O segundo experimento tem o objetivo de observar melhor a capacidade de síntese de informações discriminantes pelo SDM. Apresentou-se ao SDM o conjunto de treinamento formado pelas faces frontais masculinas não sorrindo e sorrindo, e de maneira análoga ao experimento anterior, a curva de distribuição foi determinada para se definir os limites da navegação no hiper-plano discriminante, e cujos gráficos de densidade podem ser observados na figura 5.23. A navegação no hiper-plano discriminante traz as imagens mostradas na figura 5.24 da página seguinte, onde pode-se observar que o sorriso foi a variação mais discriminante que o SDM capturou.

Observa-se também que a face média não sofre mudança de identidade, ao contrário das reconstruções produzidas pelo PCA. Porém a constatação mais importante é que o SDM pode sintetizar variações de modo contínuo a partir de imagens discretas. Em outras palavras, o SDM interpola linearmente imagens de face média no espaço de alta dimensão.

Como já discutido anteriormente, a interpolação no espaço de alta dimensão permite observar visualmente a capacidade de generalização do classificador MLDA e assim determinar qual informação discriminante foi considerada. Os resultados do PCA para este mesmo experimento não trouxeram informações úteis para discriminação, contudo as características discriminantes estão em algumas das componentes principais. Fez-se então a análise do índice de poder discriminatório através do peso do vetor W_{MLDA} e obteve-se o gráfico da figura 5.25. O gráfico mostra que as componentes principais de número 18, 16 e 22 são as que apresentam maior poder discriminatório, indicando que efetivamente a informação que discrimina realmente está contida no espaço PCA, no entanto não é a que apresenta a maior variância.

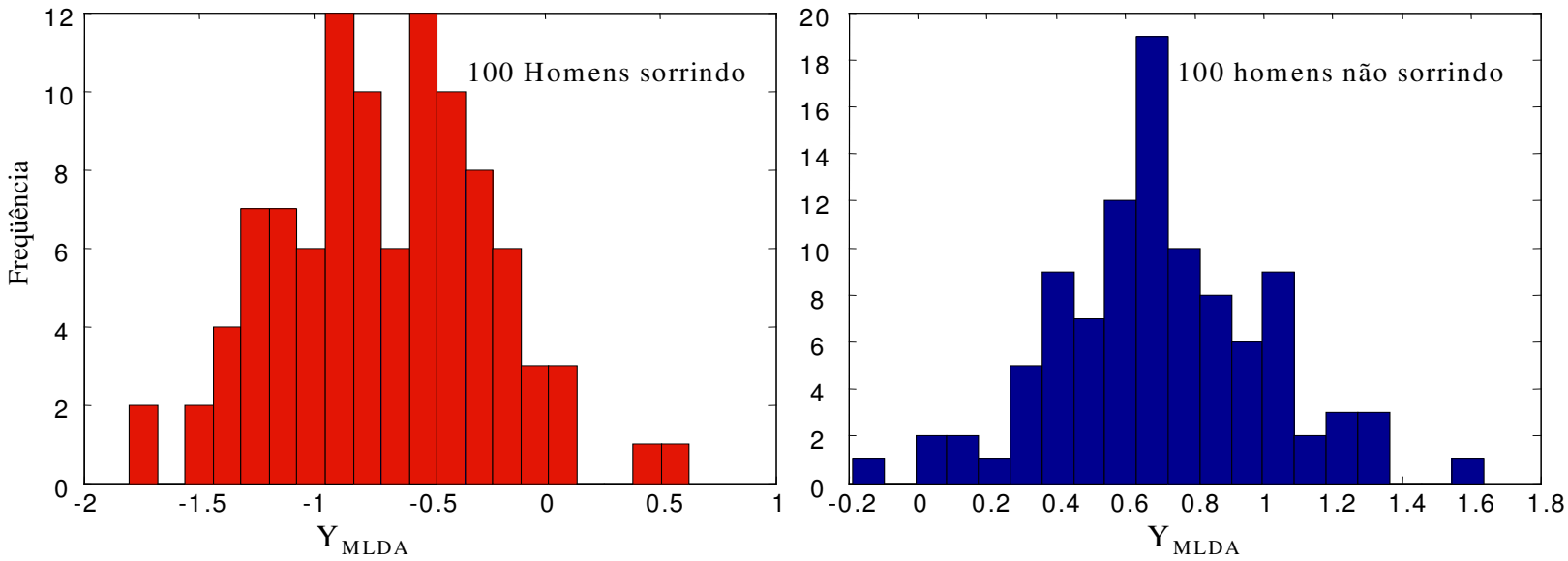


Figura 5.23 - Histograma da distribuição das projeções de 100 faces masculinas não sorrindo e 100 faces masculinas sorrindo, no espaço MLDA.

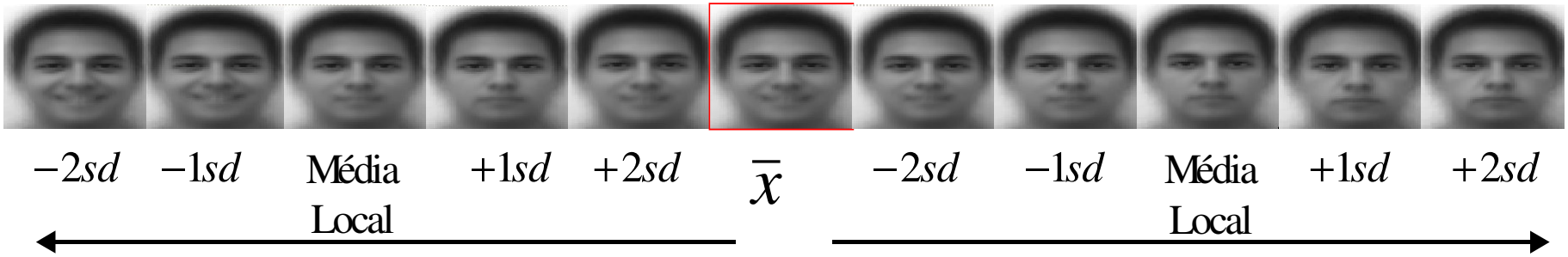


Figura 5.24 - Reconstrução visual da imagem de face média para o grupo “b”.

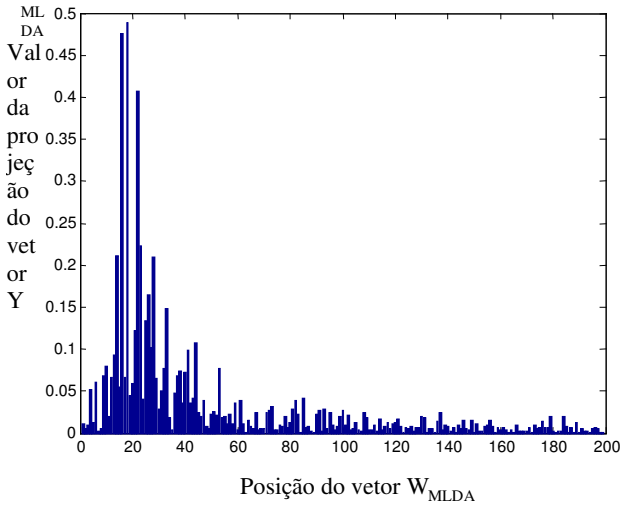


Figura 5.25 - Gráfico que representa em valores absolutos a projeção das amostras do espaço PCA no espaço MLDA para o grupo “b”.

Similarmente ao experimento 1, retorna-se ao PCA e investiga-se a síntese visual produzida quando se navega ao longo das componentes principais de número 18,16 e 22. Observa-se pela figura 5.26 a seguir que a componente principal 18 reteve efetivamente a informação que caracteriza os dois grupos. O resultado deste experimento novamente indica que a informação discriminante está contida nas componentes principais, entretanto, não necessariamente se encontra entre as que têm a maior variância.

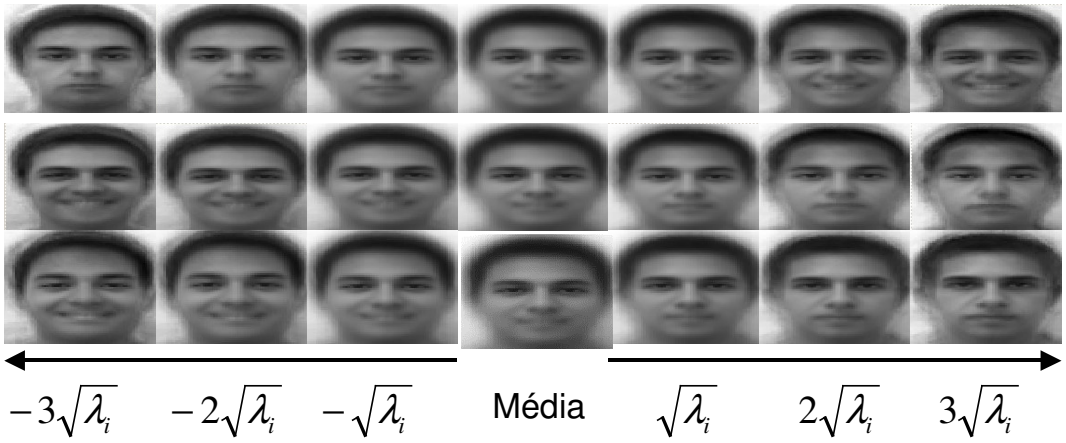


Figura 5.26 - Reconstrução visual da componentes principais de números 18, 16 e 22 do grupo “b”, do alto para baixo .

5.3.3 Experimento 3 com SDM

O terceiro experimento conduzido com o grupo “c” procura investigar como o SDM captura duas informações bem distintas entre duas classes. O experimento equivalente com o PCA mostrou que o PCA capturou o gênero como a informação mais significativa. Contudo, o grupo feminino também tinha a expressão com sorriso como uma informação distinta do grupo masculino com expressão neutra e, assim, o PCA também incorporou a informação do sorriso. No entanto os resultados e conclusões deste experimento são bem similares ao do experimento 1 com o SDM.

Na figura 5.27 a seguir, observa-se o histograma da distribuição dos grupos, determinada pelo vetor do MLDA. A navegação ao longo do hiper-plano definido pelo MLDA trouxe a síntese visual de faces, conforme ilustrado na figura 5.28. A reconstrução visual indica que o SDM considerou tanto as características de gênero quanto de expressão facial como sendo discriminante entre os dois grupos. Desta maneira a face média incorpora ambas características, indicando que o classificador encontrou um hiper-plano que separa adequadamente as duas informações discriminantes combinadas. Seguindo-se o mesmo processo dos experimentos anteriores, determinou-se qual componente principal tem alto poder de discriminação baseado no peso absoluto fornecido pelo vetor W_{MLDA}.

O gráfico da figura 5.29 mostra que as componentes principais de número 19, 14 e 1 são as que apresentam os maiores poderes discriminantes. Retornando ao PCA e navegando-se ao longo dessas três componentes principais, obtém-se as imagens ilustradas na figura 5.30. Apesar do PCA considerar o autovetor 1 como a mais discriminante, observa-se que esse autovetor modela muito mais a variação da iluminação. Os autovetores considerados mais discriminantes pelo MLDA modelam muito mais as partes internas da face.

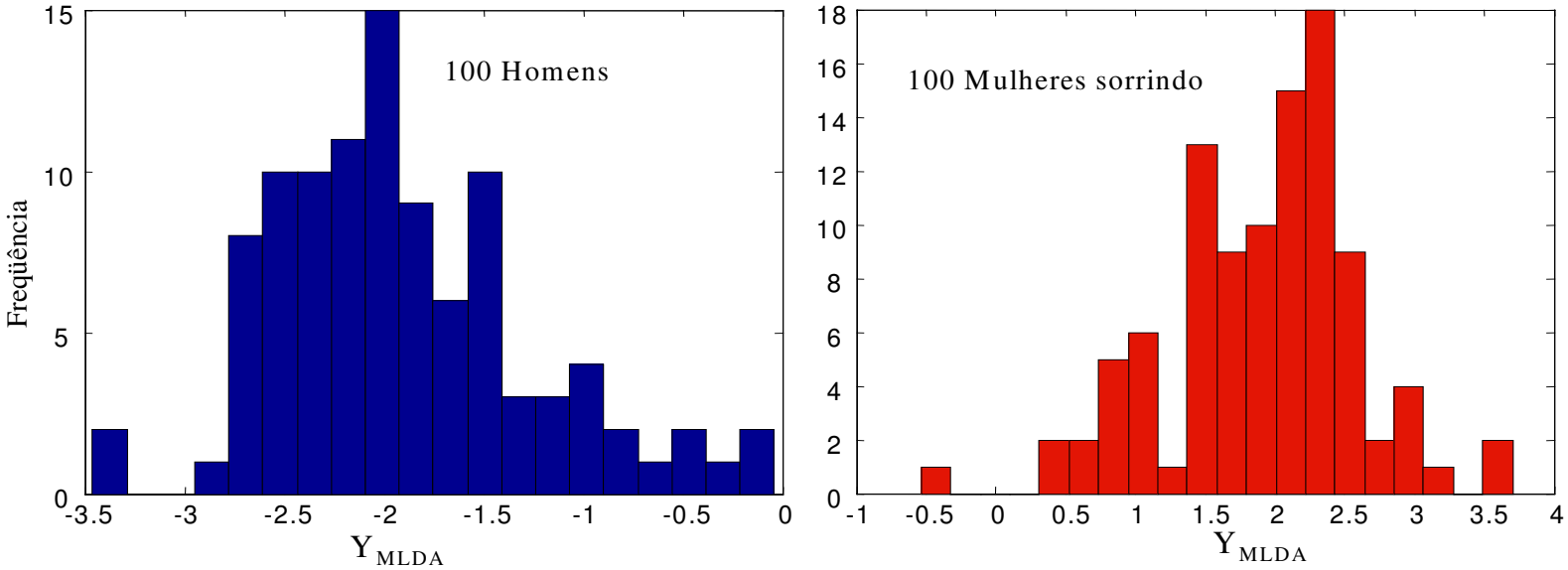


Figura 5.27 - Histograma da distribuição das projeções de 100 faces masculinas não sorrindo e 100 face femininas sorrindo, no espaço MLDA.

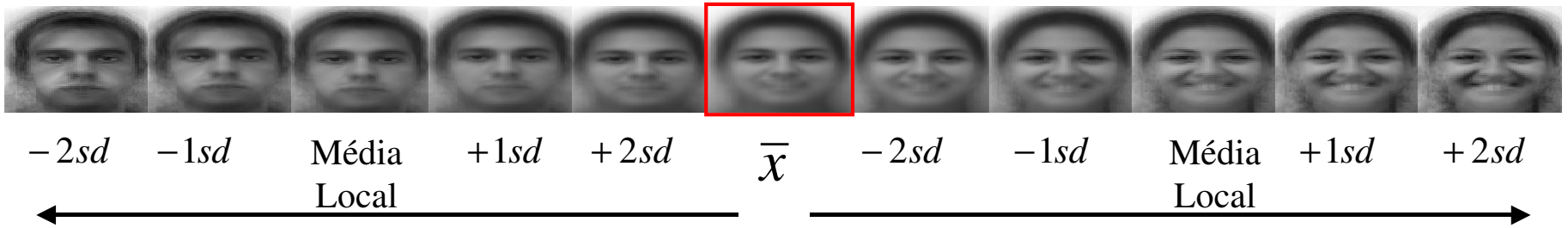


Figura 5.28 - Reconstrução visual da imagem de face média para o grupo “c”.

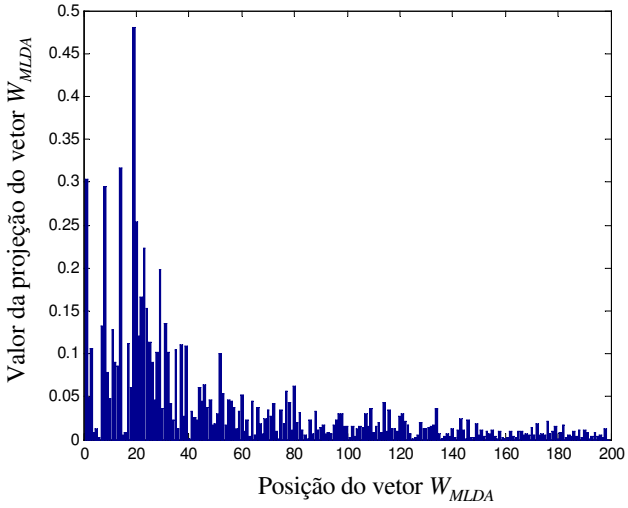


Figura 5.29 - Gráfico que representa em valores absolutos a projeção das amostras do espaço PCA no espaço MLDA para o grupo “c”.

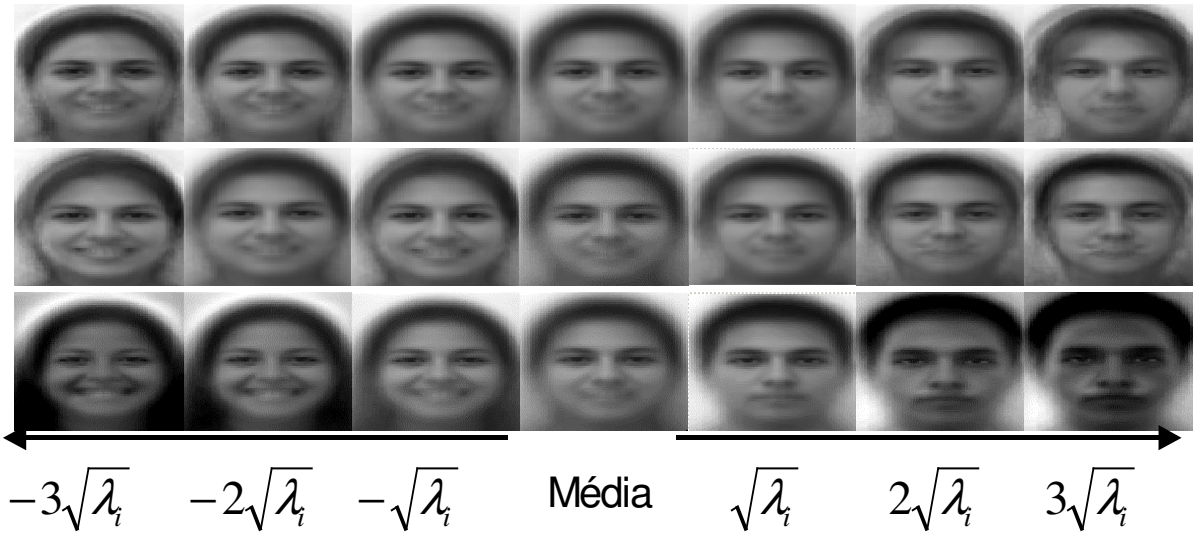


Figura 5.30 - Reconstrução visual da componentes principais de números 19, 14 e 1 do grupo “c”, do alto para baixo.

5.3.4 Experimento 4 com SDM

Neste último experimento desta série com o SDM procura-se compreender como o PCA e o MLDA se comportam com amostras que não têm uma distribuição Gaussiana. Como foi descrito anteriormente, o conjunto de treinamento do grupo “d” é formado por faces masculinas e femininas de 10 em 10 alternadamente. Isto faz com que o conjunto de treinamento seja multi-modal. Para o PCA, foi observado que este tipo de mistura não altera a capacidade de capturar a informação de maior variância, isto porque a matriz de covariância é a mesma para o conjunto, não alterando a direção de máxima variância. Como o MLDA é uma abordagem supervisionada, parte-se da hipótese que as classes são Gaussianas e distintas. Como a

determinação da melhor base para discriminação depende da maximização da distância entre as médias, se um conjunto apresenta várias distribuições Gaussianas em cada classe, ele produzirá informações errôneas sobre como as amostras se distribuem. É relevante observar que como estamos informando a existência de apenas duas classes, o classificador espera que as duas classes tenham uma distribuição unimodal. O gráfico da figura 5.31 da página seguinte ilustra a densidade de distribuição das amostras do conjunto "d" baseado no vetor $\mathbf{Y}_{MLDA}$.

A reconstrução visual do grupo "d" navegando-se no hiper-plano do MLDA trouxe as informações visuais ilustradas na figura 5.32 da próxima página. Observa-se pelas reconstruções que não é possível determinar com exatidão qual a característica mais discriminante existente entre os dois grupos. O que se observa é que a imagem de face média está quase sem nenhuma variação ao longo do hiper-plano, indicando que ela domina o espalhamento. Há uma pequena variação no tamanho e formato da cabeça e principalmente no grupo à esquerda da face média apresenta uma alternância entre essas variações. Isto é um indicador que há uma pequena sutileza que o classificador conseguiu capturar. Entretanto, ainda não é o suficiente para se fazer uma distinção visual.

Para se demonstrar o quanto o classificador foi afetado pela distribuição multi-modal das duas classes, observa-se no gráfico ilustrado na figura 5.33, como os autovetores que podem ter algum poder discriminante estão espalhados. Pelo gráfico determinou-se que as componentes principais de número 24, 29 e 74 são as que têm o maior poder de discriminação. No entanto se navegarmos nessas componentes principais, nenhuma informação muito discriminante será retornada. As reconstruções obtidas com a navegação nos três autovetores mais discriminantes trouxe as imagens ilustradas na figura 5.34. O primeiro autovetor mostra que há uma diferença entre os grupos, mas ainda assim não é possível determinar que tipo de informação esse autovetor capturou.

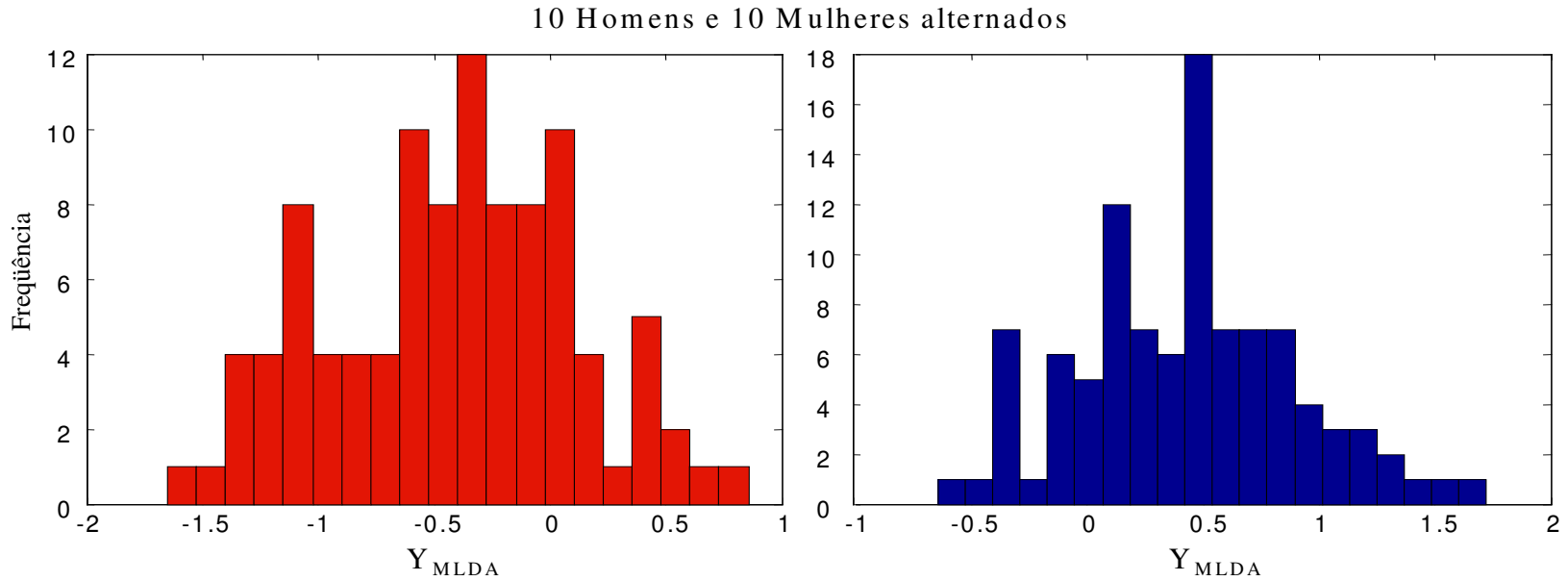


Figura 5.31 - Histograma da distribuição das projeções de 100 faces masculinas e 100 face femininas, alternadas de 10 em 10, no espaço MLDA.

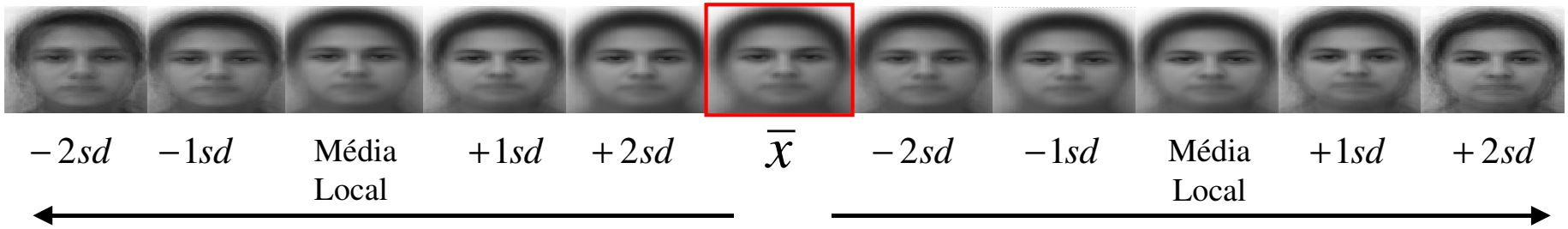


Figura 5.32 - Reconstrução visual da imagem de face média para o grupo “d”.

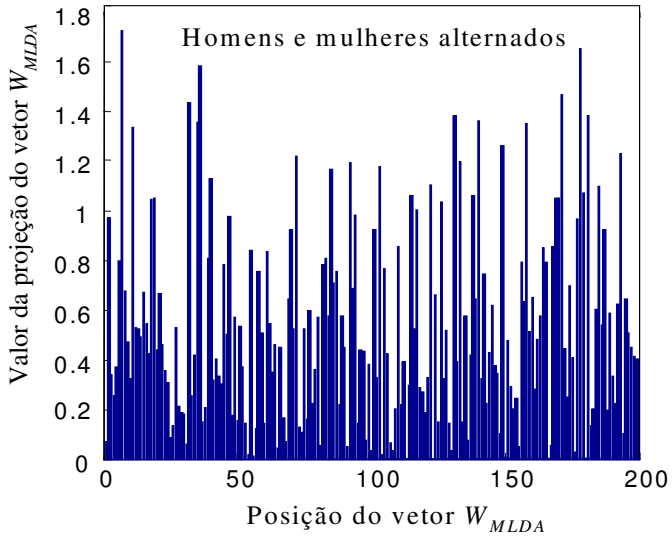


Figura 5.33 - Gráfico que representa em valores absolutos a projeção das amostra do espaço PCA no espaço MLDA para o grupo “d”.

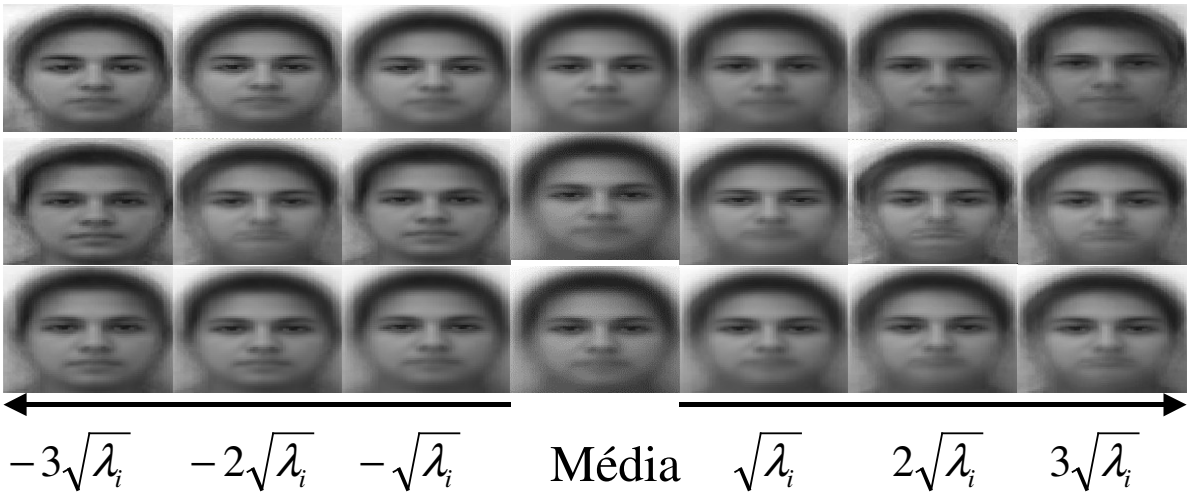


Figura 5.34 - Reconstrução visual da componentes principais de números 24, 29 e 74 do grupo “d”, do alto para baixo.

5.3.5 Determinação da Informação mais Discriminante

Os três primeiros experimentos com o SDM mostraram que é possível extrair informações discriminantes de um conjunto de treinamento, porém resta avaliar o que seria efetivamente uma informação discriminante, caso não existisse um conhecimento a *prior*i dessa informação. A investigação de uma abordagem discriminante é facilitada quando se usa um conjunto de treinamento com faces humanas, porque existe esse conhecimento prévio sobre o comportamento das amostras. Entretanto, se os conjuntos fossem formados por amostras cujo o conhecimento prévio sobre as diferenças não fosse bem estabelecido ficaria mais difícil uma interpretação dos resultados. Observou-se durante os experimentos com o SDM que quando

os conjuntos de treinamento têm uma distribuição Gaussiana, a navegação ao longo do hiper-plano discriminante indica que há uma separação clara entre as características que pertencem a cada grupo. Se navegarmos além dos limites estabelecidos de $\pm 2sd$ (*standard deviation)* e reconstruirmos essas imagens, é possível avaliar visualmente o que seria a informação mais discriminante de cada grupo, ou seja, neste caso o que seria definitivamente masculino ou feminino. A figura a seguir mostra a reconstrução visual do grupo "a" homens e mulheres, navegando-se agora até os limites de $\pm 15sd$.

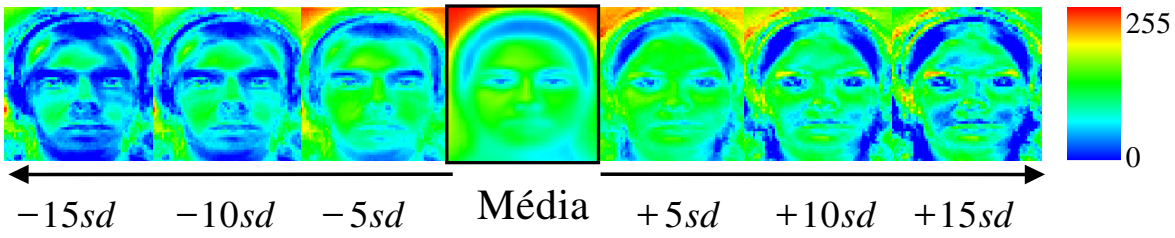


Figura 5.35 - Reconstrução da imagem de face média do grupo "a" quando se navega para além dos limites convencionais.

A figura 5.35 mostra as regiões da face que sofrem uma transformação mais intensa e destaca as características da face que produzem a diferença entre os grupos. As imagens foram reconstruídas utilizando-se uma tabela de transformação dos níveis de cinza linear para um modelo em três cores, de modo a destacar as variações. No apêndice "B" o leitor encontrará os detalhes da tabela de transformação utilizados nesta dissertação. Portanto, comparando-se as regiões que mudaram de cor em relação à face média, pode-se determinar quais características da face são afetadas. No caso das faces masculinas observa-se que a região do queixo, as sobrancelhas, o tamanho do cabelo, são as que mais discriminam, enquanto que nas faces femininas, a região em torno dos lábios, as bochechas, os olhos e o cabelo definem as características desse grupo.

Para se compreender melhor o que ocorre com a informação mais discriminante, verifica-se que o segundo termo da equação abaixo tem o seu comportamento regido pelo parâmetro y_i^{MLDA}, isto porque as bases Φ_{PCA} e Φ_{MLDA} são constantes, para um determinado conjunto de treinamento. Como o parâmetro y_i^{MLDA} faz o produto interno com as duas bases, pode-se concluir que a variação do segundo termo da equação (5.60) é linear.

$$x_i = \bar{x} + \Phi_{PCA}^T \Phi_{MLDA}^T y_i^{MLDA}. \tag{5.60}$$

Entretanto, a reconstrução visual da projeção y_i^{MLDA} sem a média indica que esse parâmetro afeta diretamente a amplitude dos níveis de cinza de cada *pixel* da imagem de face. Pode-se concluir então que as imagens reconstruídas na figura 5.35 representam a variação Δ_{img} de cada *pixel* em termos de profundidade, quando navegamos ao longo do hiper-plano

discriminante, sugerindo que as variações nas imagens são as informações discriminantes que são consideradas pelo classificador, bem como o seu grau de generalização. A equação (5.61) a seguir descreve a variação entre duas imagens.

$$\Delta_{img} = x_i - x_{i-1}. \tag{5.61}$$

Se considerarmos que Δ_{img} é a variação que ocorre entre duas imagens quando navega-se ao longo do hiper-plano discriminante, pode-se escrever essa diferença como sendo as variações impostas pela reconstrução do vetor y_i^{MLDA}, ou seja,

$$\Delta_{img} = \Phi_{PCA}^T \Phi_{MLDA}^T \left(y_i^{MLDA} - y_{i-1}^{MLDA} \right). \tag{5.62}$$

E considerando-se ainda que existe uma variação entre duas ou mais imagens, é possível determinar a taxa de variação δ_{img} da imagem em função do incremento ou decremento de y_i^{MLDA}, conforme descrito na equação abaixo

$$\delta_{img} = \frac{\Delta_{img}}{dx}. \tag{5.63}$$

Desta maneira, as imagens de faces médias reconstruídas e coloridas artificialmente indicam que há uma variação não uniforme ao longo das imagens, indicando que as áreas discriminantes sofrem alterações mais rápidas que as áreas não discriminantes (áreas em verde). Para se determinar numericamente a taxa de variação, calculou-se a variação das imagens geradas em $\pm 1sd$. Como para $y_i^{MLDA} = 0$, a taxa de variação é zero, e o que resta é a própria imagem da face média, logo, para $\pm 1sd$ a variação será a própria reconstrução dos pontos y_i^{MLDA} sem a face média, como está definido na equação (5.62). O gráfico (a) da figura 5.36 ilustra os valores do vetor δ_{img} calculados para o grupo "a". Observa-se pelo gráfico que há regiões na imagem que a taxa de variação entre os *pixels* vizinhos é muito pequena, enquanto que há outras regiões cuja a taxa de variação é muito maior. Como essas taxas representam números reais, ordenou-se crescentemente o vetor δ_{img} de modo a investigar melhor como ocorrem as variações ao longo da imagem.

O gráfico (b) da figura 5.36 representa o vetor δ_{img} ordenado crescentemente, indicando que os pesos do vetor δ_{img} podem ter uma distribuição Gaussiana. Considerando-se esta hipótese, calculou-se a média μ_{img} e o desvio padrão *sd* do vetor δ_{img}, obtendo-se a média $\mu_{img} = 2{,}299 \times 10^{-4}$ e desvio padrão $sd = 0{,}0119$. Notou-se que uma quantidade pequena de pontos do vetor δ_{img} têm taxa de variação maior que $\pm 1sd$, por exemplo. Localizando-se esses pontos com variação maior que $\pm 1sd$, observou-se que eles pertencem às regiões da

imagem que visualmente são consideradas as áreas das características mais discriminantes. Logo, conclui-se que as regiões que têm um peso significativo para fins de classificação apresentam as maiores taxas de variação quando se navega pelo hiper-plano discriminante. O histograma da figura 5.37 representa a densidade de distribuição dos pesos do vetor δ_{img}. Portanto, a investigação visual em três cores pode representar as taxas de variação das características mais discriminantes e assim facilitar a análise sobre quais regiões da face média sofrem alterações mais rapidamente.

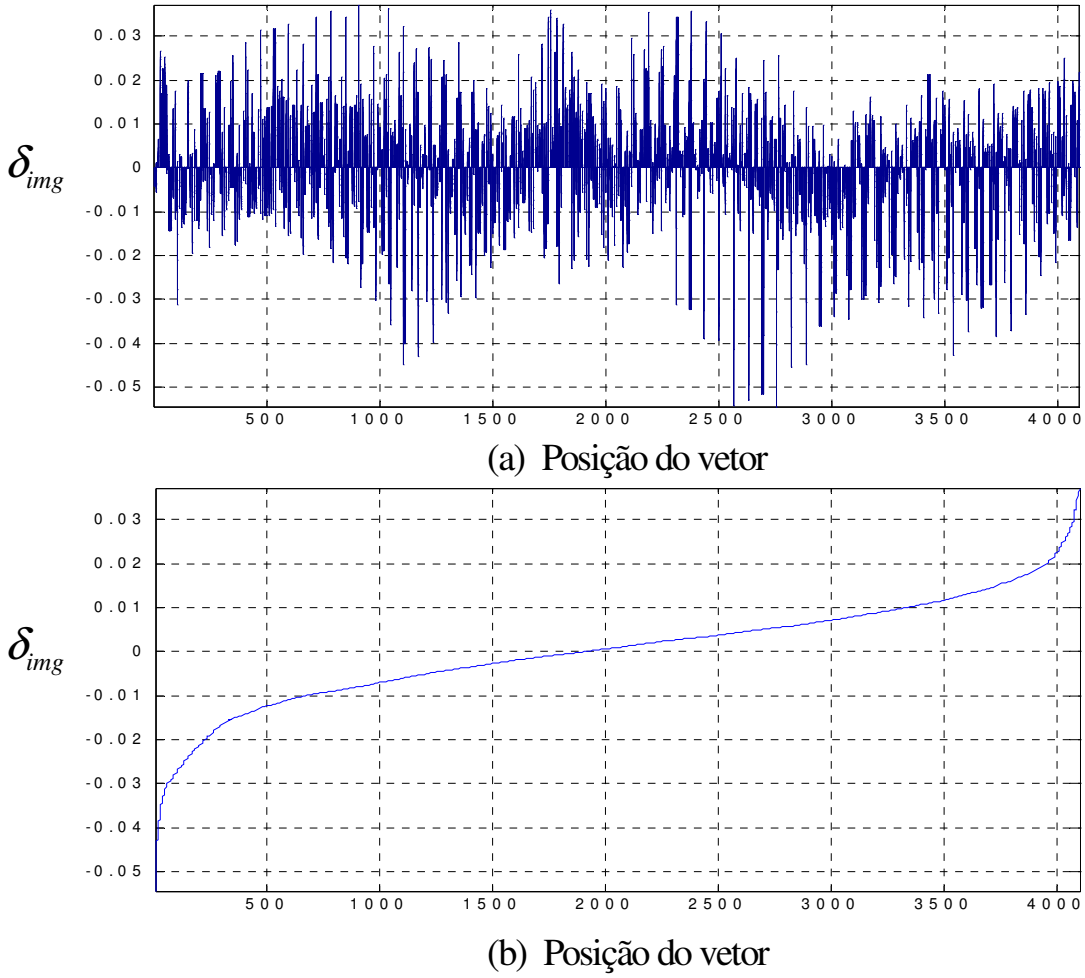


Figura 5.36 - No gráfico (a) temos o vetor δ_{img} reconstruído do ponto $y_i^{MLDA} = +1sd$ para o grupo "a" homens e mulheres. No gráfico (b) vemos as taxas de variação do gráfico ao longo do vetor δ_{img}, porém com os valores das taxas ordenadas crescentemente.

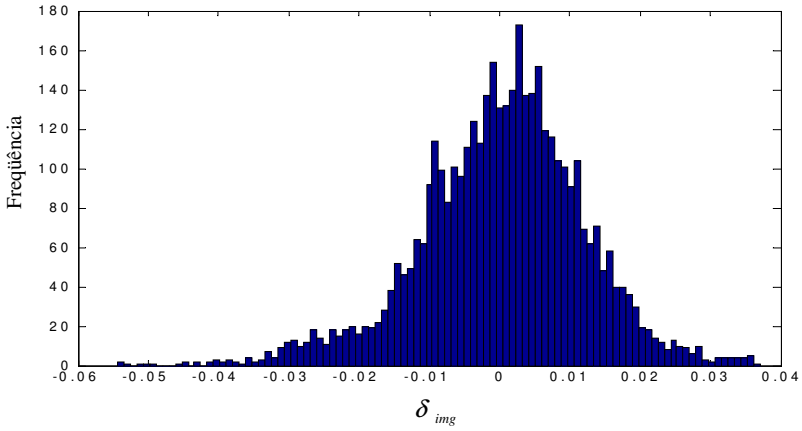


Figura 5.37 - Histograma dos pesos do vetor δ_{img} .

De maneira análoga, reconstruiu-se a imagem média do grupo “b” homens sorrindo e não sorrindo extrapolando-se a navegação para os limites de $\pm 15sd$, e obteve-se as imagens ilustradas na figura 5.38. Observa-se que a medida que o intensidade do sorriso aumenta, aumenta-se também a área facial relacionada com o sorriso. As áreas coloridas em azul e vermelho indicam as regiões da face média que sofrem alterações consideráveis quando se percorre o hiperplano discriminante.

Destaca-se que as regiões indicadas coincidem com as áreas da face que sofrem alterações quando ocorre uma mudança nas expressões faciais. A figura 5.39 faz a reconstrução utilizando o grupo “c” com homens não sorrindo e mulheres sorrindo.

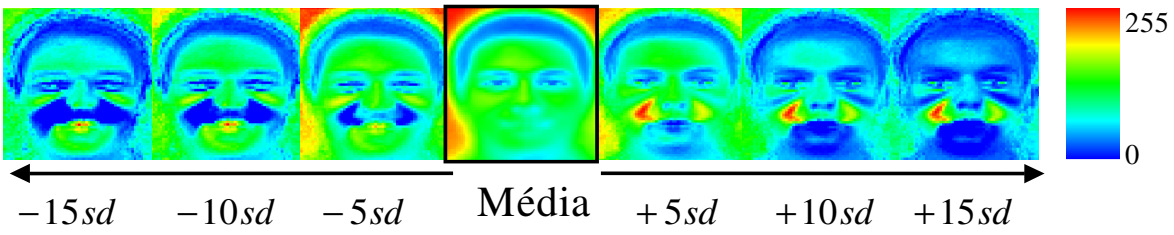


Figura 5.38 - Reconstrução da imagem de face média do grupo “b” quando se navega para além dos limites convencionais.

Observa-se que as áreas que sofrem variação são muito semelhantes àquelas produzidas pelo experimento com o grupo “b”, e percebe-se que a imagem de face feminina também incorporou variações na área relativa ao gênero, tais como no experimento com o grupo “a”.

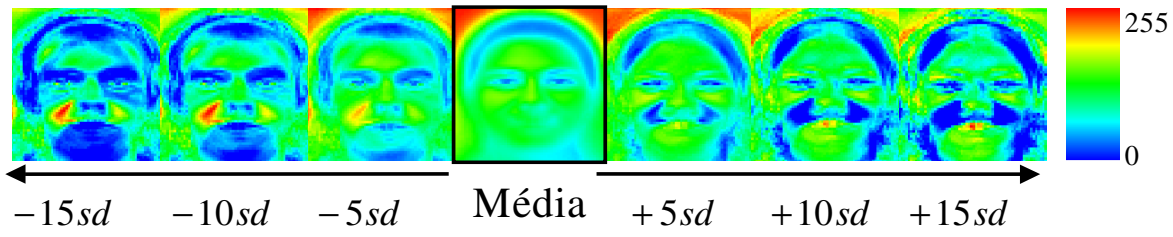


Figura 5.39 - Reconstrução da imagem de face média do grupo “c” quando se navega para além dos limites convencionais.

A última reconstrução foi executada com o conjunto de treinamento do grupo “d” homens e mulheres alternados de 10 em 10 imagens. As imagens da figura 5.40 indicam uma dificuldade em se definir as características mais discriminantes entre os dois grupos. Não é possível identificar com segurança o que as imagens capturaram.

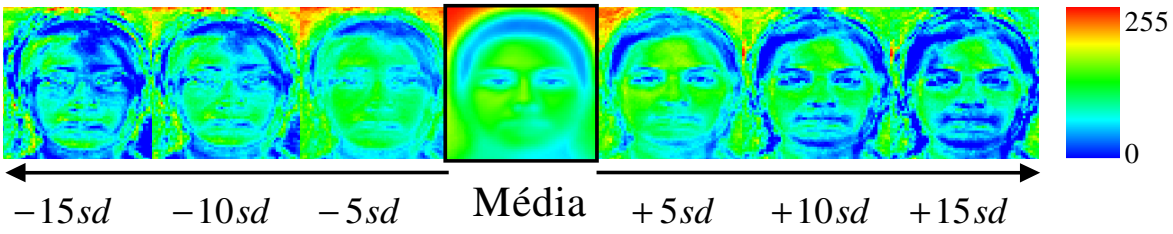


Figura 5.40 - Reconstrução da imagem de face média do grupo “d” quando se navega para além dos limites convencionais.

Os experimentos com o SDM extrapolando-se a navegação para além de $\pm 2sd$ indicam que é possível determinar quais características faciais são consideradas mais discriminante.

5.3.6 Transferência da Informação mais Discriminante

Nesta seção são descritos os dois últimos experimentos que foram conduzidos neste trabalho. Tomou-se uma imagem de face do conjunto de treinamento e executou-se a quarta fase do SDM, que é tentar incorporar as características mais discriminantes em uma imagem de face. O primeiro experimento foi realizado utilizando-se o conjunto de treinamento do grupo “b“, homens não sorrindo e sorrindo. Já o segundo experimento foi realizado utilizando-se o conjunto de homens e mulheres, ou seja, o conjunto de treinamento do grupo “a”. Para se fazer o teste da síntese ou transferência de expressão foi utilizada uma imagem de face feminina. A motivação é investigar o que ocorre quando uma imagem de face tão distinta daquelas do conjunto de treinamento recebe as informações discriminantes. O processo de síntese e transferência de expressões são abordagens cuja aplicação são estudadas para a criação de avatares. Em (TERZOPOULOS; WATERS, 1993) os autores apresentaram um complexo

modelo de síntese de expressões faciais baseados na decomposição geométrica das linhas da face e da anatomia dos músculos faciais. Em (WANG; AHUJA, 2003) foi proposto a decomposição de expressões e posterior transferência utilizando-se o modelo *Higher Order Singular Value Decomposition* (HOSVD) baseado em tensores. Esta dissertação, entretanto, utiliza uma abordagem baseada em discriminantes lineares para-se obter o efeito da síntese e transferência de expressões ou características. A figura 5.41 representa a imagem da face de teste utilizada durante as transferências de informações discriminantes.

Figura 5.41 - Imagem de uma pessoa com expressão neutra.

O resultado da primeira transferência pode ser visualizado na figura 5.42, onde observa-se que a imagem feminina incorporou as expressões vindas da face média, mas sem necessariamente agregar qualquer característica masculina. Como a face média é o modelo de representação, todas as outras serão uma variação desse modelo. Logo a imagem feminina tomou o modelo e incorporou a sua variação própria, atenuando os traços masculinos e adicionando as variações de expressão da face média. Para determinar se o resultado das transformações podem ser considerados válidos, projetou-se a imagem da transformação considerada como sendo definitivamente sorrindo baseada no julgamento visual e calculou-se a distância Euclidiana entre essa transformação e as médias locais a fim de se determinar, sob os aspectos da classificação, se a face transformada poderia pertencer ao grupo sorrindo.

Tabela 2 – Comparativo entre face média e a face transformada (grupo “b”).

Expressão da face média de cada grupo para teste	Face média de cada grupo projetada no espaço MLDA	Face transformada e projetada no espaço MLDA	Distância Euclidiana
Sorrindo	-0,6988	-2,1832	1,4844
Não sorrindo	0,6988	-2,1832	2,882

Os resultados da tabela 2 indicam que a face transformada pode ser classificada como pertencente ao grupo de expressões sorrindo porque é a que tem a menor distância Euclidiana,

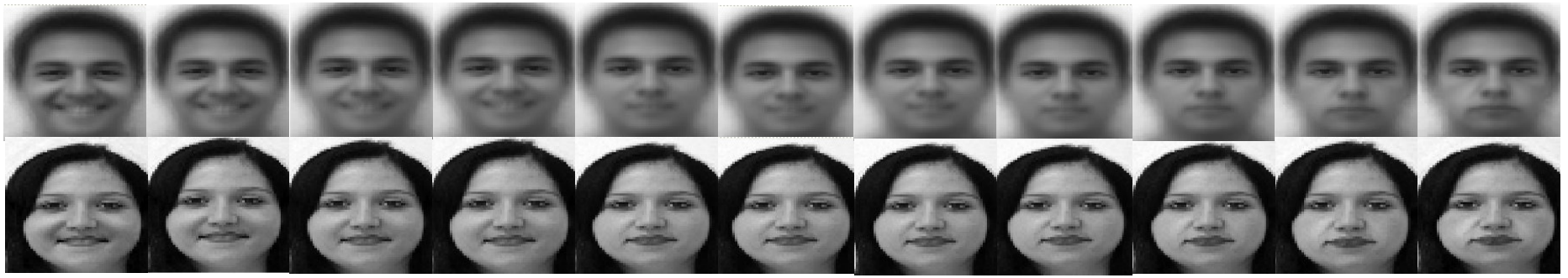


Figura 5.42 - Imagens sintetizadas de uma pessoa quando se navega ao longo do hiperplano discriminante de homens não sorrindo e sorrindo.

confirmando a análise visual. De modo semelhante, tomou-se a imagem de face transformada que foi considerada visualmente como sendo definitivamente não sorrindo e calculou-se a distância Euclidiana. Os resultados aparecem na tabela abaixo.

Tabela 3 – Comparativo entre face média e a face transformada (grupo "b").

Expressão da face média de cada grupo para teste	Face média de cada grupo projetada no espaço MLDA	Face transformada e projetada no espaço MLDA	Distância Euclidiana
Sorrindo	-0,6988	1,5256	2,2244
Não sorrindo	0,6988	1,5256	0,8268

A menor distância Euclidiana em relação à face média local das expressões não sorrindo indica que a análise visual estava correta para este caso. Portanto, essa imagem de teste transformada seria classificada como sendo não sorrindo.

A segunda transferência utilizou o conjunto de treinamento formado por homens e mulheres, e novamente trabalhando-se com a mesma imagem de face do experimento anterior reconstruiu-se visualmente o que foi incorporado. Os resultados podem ser observados na figura 5.43. Nota-se na imagem de teste feminina que a navegação no sentido masculino (à esquerda) incorpora algumas características tornando a expressão mais rude e masculinizada sem no entanto perder a identidade. A navegação e síntese para o outro sentido (à direita) mostra que há pequenas transformações, basicamente na região dos olhos. Como a face de teste já era feminina, a navegação naquele sentido não trouxe grandes alterações. Novamente, para se validar os resultados visuais, foram projetados no espaço MLDA as imagens de face consideradas definitivamente masculina e definitivamente feminina, baseadas no julgamento visual, e determinou-se a distância Euclidiana entre as faces transformadas e as médias locais. O resultado com a face considerada definitivamente masculina pode ser visto na tabela 4, confirmando a análise visual.

Tabela 4 – Comparação entre a face média e a face transformada (grupo "a").

Expressão da face média de cada grupo para teste	Face média de cada grupo projetada no espaço MLDA	Face transformada e projetada no espaço MLDA	Distância Euclidiana
Masculino	-1,7799	-0,9266	0,8533
Feminino	1,7799	-0,9266	2,7065

Por último, tomou-se a face transformada considerada definitivamente feminina, sob os aspectos visuais, e projetou-se no espaço MLDA. O valor da projeção foi testado com os valores das faces médias locais de modo a determinar qual classe produz a menor distância

Euclidiana. O resultados são apresentados na tabela 5 a seguir e indicam que os resultados numéricos coincidem com a avaliação visual.

Tabela 5 – Comparação entre a face média e a face transformada (grupo ”a”).

Expressão da face média de cada grupo para teste	Face média de cada grupo projetada no espaço MLDA	Face transformada e projetada no espaço MLDA	Distância Euclidiana
Masculino	-1,7799	5,6247	6,3235
Feminino	1,7799	5,6247	4,9259

5.4 Considerações Finais

Os experimentos conduzidos e relatados neste capítulo demonstram que é possível a um classificador linear extrair informações que não necessariamente estão presentes nos conjuntos de treinamento. Essas imagens com as informações discriminantes destacadas são equivalentes a uma interpolação linear em uma alta dimensão. Verificou-se também que o PCA é eficaz na extração de características para fins de representação, no entanto o abordagens com classificadores lineares podem trazer informações mais ricas sobre como ocorrem as variações dentro desses conjuntos, a partir de imagens estáticas. No próximo capítulo, conclui-se este trabalho apresentando todos os pontos relevantes desenvolvidos nesta dissertação.

Figura 5.43 - Imagens sintetizadas de uma pessoa quando se navega ao longo do hiperplano discriminante de homens e mulheres.

6 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho apresentou os resultados da síntese e reconstrução visual do espaço de discriminantes lineares capturado pela abordagem MLDA, demonstrando-se que características sutis podem ser capturadas e modeladas estatisticamente. A navegação ao longo das componentes discriminantes permitiu observar quais características foram consideradas discriminantes e o grau de generalização das mesmas. A abordagem SDM permitiu ainda modelar e sintetizar as transições existentes entre imagens discretas de faces, sem que necessariamente elas existissem no conjunto de treinamento. Como já foi descrito nas seções anteriores, este é um processo equivalente a uma interpolação linear em um espaço de alta dimensão.

Determinou-se a capacidade de extração de informação pelo PCA e concluiu-se que o PCA é eficiente nessa extração. Entretanto, não necessariamente o algoritmo original do PCA, que ordena os autovetores conforme os maiores autovalores, é a melhor solução para algumas aplicações. Durante os experimentos, percebeu-se que há um potencial enorme de exploração dos autovetores, uma vez que eles retém informações muito importantes sobre o conjunto de treinamento. Consequentemente, existe um campo de investigação sobre o uso das componentes principais para fins de classificação, o que poderia aumentar o poder do PCA. Destaca-se também a extensa discussão sobre a abordagem PCA para se compreender a sua aplicação no domínio de faces.

Ainda como contribuição do trabalho, mostrou-se que o método proposto baseado no modelo de Fisher não é apenas útil como um classificador, mas também pode ser utilizado para extrair e predizer informações, e principalmente destacando informações que mostram como os grupos se separam. A abordagem SDM demonstra que pode capturar, mapear, predizer e reconstruir visualmente as variações que ocorrem com dois grupos distintos de imagens, mas sem a necessidade de um auxílio humano na determinação das características discriminantes. Demonstrou-se que a face média pode ser considerada como um modelo de representação das faces que formam um determinado conjunto de treinamento, indicando que todas as faces são uma variação da face média. A partir desta representação verificou-se que é possível transferir determinadas características visuais para imagens de face que não necessariamente pertencem ao conjunto de treinamento, somente variando-se a face média.

Avaliou-se também o poder de generalização das componentes principais e qual tipo de informação elas estariam retendo. E observou-se ainda que as componentes principais ordenadas pela maior variância não necessariamente determinam a melhor base vetorial para fins de classificação. Estudou-se também como e onde ocorrem as variações das informações

mais discriminantes nas imagens de face, de modo a facilitar a compreensão do fenômeno de generalização do classificador quantificado por uma métrica baseada na taxa de variação das transformações. Entretanto, a abordagem SDM exige algumas pré-condições para que se possa determinar a informação mais discriminante. Cita-se como necessário um pré-alinhamento de todos as amostras, assim como também é exigido por todas as abordagens semelhantes ao SDM. A formação de dois ou mais conjuntos distintos de amostras uni-modais cujas densidades de distribuição possam ser aproximadas por um função Gaussiana mostrou-se desejável, de modo a permitir uma melhor caracterização da informação mais discriminante. Esta condição é também básica para todas as abordagens, que como o SDM, são lineares.

Foi discutido também que a abordagem SDM utilizaria todas as componentes principais da base PCA, isto porque partiu-se da hipótese que não se desejaria perder qualquer informação discriminante que eventualmente poderia estar contida nas componentes principais de menor variância. Como conseqüência natural desse processo, foram apresentadas e testadas duas métricas para se determinar o poder de discriminação das componentes principais. Finalmente, considera-se relevante destacar que a utilização de abordagens combinadas, tendo-se o PCA como um pré-processador de informações, torna muito mais eficiente qualquer estudo com conjuntos que apresentam alta dimensionalidade, tais como aquelas com imagens de face. Observou-se também que o poder de representação e discriminação contidos nos autovetores, tanto da base PCA quanto da base MLDA, pode ser útil para o desenvolvimento de novas abordagens, dependendo apenas de como essas bases são trabalhadas.

Como trabalhos futuros destacamos a navegação nos hiperplanos discriminantes para além de duas classes e a observação das respectivas reconstruções visuais. A determinação de uma função critério para investigar como utilizar as informações discriminantes da base vetorial do MLDA para localizar as componentes principais com o melhor poder classificador também seria uma extensão natural deste trabalho. Além disso, acredita-se que seria possível determinar como utilizar as componentes principais para construir um classificador mais robusto e preciso, uma vez que se observou que a base vetorial do PCA pode ser eficientemente manipulada para outros fins que não o da representação da informação com mínimo erro de reconstrução.

Uma outra aplicação promissora seria o uso da abordagem SDM em experimentos que não sejam necessariamente compostos de imagens de faces. Cita-se como principal exemplo a aplicação em imagens médicas, uma vez que neste tipo de aplicação é necessário o auxilio de um especialista para se determinar quais informações são relevantes para fins de classificação. Neste tipo de aplicação, o SDM poderia ser útil ao auxiliar o pesquisador apresentando visualmente como ocorrem as variações entre duas ou mais classes em termos de informação dis-

criminante. Igualmente útil para o especialista seria avaliar o grau de generalização tanto do classificador quanto das componentes principais. Isto porque possivelmente estruturas que não são percebidas visualmente, quando se trabalha com toda a imagem, poderiam ser destacadas quando se navega pelos hiperplanos discriminantes ou então pelas componentes principais. Finalmente, o último experimento demonstrou que o SDM pode ser útil também para produzir a síntese de expressões a partir de informações aprendidas de um sub-espaço discriminante. Isto seria particularmente útil nas aplicações de computação gráfica, especificamente na criação de avatares, onde o modelagem de expressões faciais do avatar poderia utilizar as informações capturadas do sub-espaço discriminante de algum conjunto de treinamento.

REFERÊNCIAS

BARTLETT, M. S.; SEJNOWSKI, T. J. Independent component of face images: A representations for face recognition. In ANNUAL JOINT SYMPOSIUM ON NEURAL COMPUTATION, v. 4, 8 p., Pasadena. California, Anais. Pasadena , May 1997.

BARTLETT, M. S.; MOVELLAN, J. R.; SEJNOWSKI, T. J. Face recognition by independent component analysis. **IEEE Transaction on Neural Networks**, USA, v. 13, n. 6, p. 1450-1464, November 2002.

BASSANEZI, R. C. **Ensino-Aprendizagem com modelagem matemática.** SP: Contexto, 2002.

BATISTA, G. E. A. P. A, **Um ambiente de avaliação de algoritmos de aprendizado de máquina utilizando exemplos**. 1997. 171 f. Dissertação (Mestrado em Ciência da Computação) – Instituto de Ciências Matemáticas de São Carlos, USP-São Carlos, São Carlos, Setembro de 1997.

BELHUMEUR, P.; HESPANHA, J. P. N.; KRIEGMAN, D. J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, USA v. 19, n. 7, p. 711-720, July 1997.

BRUNELLI, R.; POGGIO, T. Face recognition: features versus templates. **IEEE Transactions on Pattern Recognition and Machine Intelligence**, v. 15, p. 1042-1052, 1993.

BUCHALA, S.; DAVEY, N.; GALE, T. M.; FRANK, R. J. Principal components analysis of gender, ethnicity, age and identity of face images,8 pages, In **Proceeding of IEEE ICMI UK**, 2005.

BURTON, A. M.; BRUCE, V.; HANCOCK, P. J. B.; From pixels to people: A model of familiar face recognition. **Cognitive Science,** v. 23 p. 1-31, 1999.

CALLIOLI, C. R.; HYGINO, H. D.; COSTA, R. C. F. **Álgebra linear e aplicações.** 2ª ed. SP: Atual, 1978.

CAMPOS, T. E. **Técnicas de seleção de características com aplicações em reconhecimento de faces.** 2001. 138 f. Dissertação (Mestrado em Ciência da Computação) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 25 de Maio de 2001.

CESAR JR., R. M.; Olhos nos olhos. Ensinando o computador a reconhecer pessoas. **Revista Ciência Hoje**, São Paulo v. 29, n. 174, p. 24-29, 2001.

CHANG, W. C., On using principal components before separating a mixture of two multivariate normal distributions, **Applied Statistics**, v. 32, p. 267-275, 1983.

CHELLAPPA, R.; WILSON, C. L.; SIROHEY, S.; Human and machine recognition of faces: A Survey. **Proceedings of IEEE**, v. 83. n. 5, May 1995.

COOTES, T.F.; TAYLOR, C. J. COOPER, D. H. GRAHAM, J.; Active Shape Models- Their training and application. **Computer Vision and Image Understanding**, v. 61, n.1, p. 38-59, 1995.

COOTES, T. F.; EDWARDS, G. J.; TAYLOR, C. J. Active Appearance Models. H. BURKHARDT and B. NEWMANN, EUROPEAN CONFERENCE ON COMPUTER VISION, 5., 1998, Berlin, v. 2, Anais. Berlin, Springer, 1998. p. 484-498.

COOTES, T. F.; WALKER, K. N.; TAYLOR, C. J. View-Based Active Appearance Models, In: INTERNACIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, Grenoble, Anais. France, 2000. p. 227-232.

COOTES, T.F.; LANITIS, A.; Statistical Models of Appearance for Computer Vision, Technical report on Imaging Science and Biomedical Engineering Dept., University of Manchester, 125 f. 2004.

DRAPER, B. A.; BAEK, K.; BARTLETT, M. S.; BEVERIDGE, J. R.; Recognizing Faces with PCA and ICA. **Computer Vision and Image Understanding**, v. 91, p. 115-137; 2003.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern recognition.** 2nd ed. New York: John Wiley & Sons, 2001.

EDWARDS, G. J.; LANITIS, A.; TAYLOR, C. J.; COOTES, T. F.; Statistical models of face images – Improving specificity. **Image and Vision Computing**; v. 16, p. 203-211, 1998.

EZZAT, T.; POGGIO, T.; Facial Analysis and Synthesis Using Image-Based Models. In 2nd INTERNACIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, Killington, p. 116-121, Vermont, 1996.

FERIS, R. S. **Rastreamento eficiente de faces em um subespaço wavelet.** 2001. 74 f. Dissertação (Mestrado em Ciência da Computação) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 11 de Maio de 2001.

FISHER, R. A., The use of multiple measures in taxonomic problems, **Ann. Eugenics**, v. 7, p. 179-188, 1936.

FUKUNAGA, K. **Introduction to statistical pattern recognition.** 2nd ed. Boston: Academic Press, 1990.

GOODALL, G., Procustes methods in the statistical analysis of shape. **Journal of royal statistical society**, series B 53:285-339, 1991.

GONZALEZ, R. C.; WOODS, R. E., **Digital Image Processing**, Addison-Wesley Publishing Company, 1992.

HAYKIN, S. **Redes neurais princípios e práticas.** 2ª ed. Porto Alegre: Bookmann, 2001.

HIREMATH, P. S.; PRABHAKAR, C. J., Face recognition technique using symbolic PCA method, **PATTERN RECOGNITION AND MACHINE INTELLIGENCE, PROCEEDINGS,** 3776: p. 266-271, **LECTURE NOTES IN COMPUTER SCIENCE**, 2005.

HOTTELING, H., Analysis of a complex of statistical variables into principal components, **Journal of educational psychology, v.** 24, p. 498-520, 1933.

JAIN, A.; DUIN, R.P.W.; MAO, J.; Statistical pattern recognition: A review, **IEEE-PAMI**, v. 22, n. 1, January 2000.

JIN, Z.; DAVOINE, F.; LOU, Z.; YANG, J.; A novel PCA-based Bayes classifier and face analysis, **ADVANCES IN BIOMETRICS**, Proceedings, 3832: p. 144-150, 2005, **LECTURE NOTES IN COMPUTER SCIENCE**.

JOLLIFFE, I. T. **Principal components analysis**, UK: 2ª ed., Springer, 2002.

JOLLIFFE, I. T., MORGAN, B. J. T., YOUNG, P. J., A simulation study of the use of principal components in linear discriminant analysis, **Journal of statistical computing**, v. 55, p. 353-366, 1996.

JOHNSON, R. A.; WICHERN, D. W., **Applied multivariate statistical analysis**, USA 5th ed., Prentice Hall, 2002.

KITANI, E. C.; THOMAZ, C. E.; Um tutorial sobre análise de componentes principais para o reconhecimento automático de faces; Relatório Técnico 01/2006, Departamento de Engenharia Elétrica da FEI, São Bernardo do Campo, SP; 23f. 2006a, disponível em www.fei.edu.br/~cet.

KITANI, E. C.; THOMAZ, C. E.; GILLIES, D. F. A statistical discriminant model for face interpretation and reconstruction. In: proceedings of SIBGRAPI' 06, **IEEE CS Press**, Manaus, Amazonas, Brazil, October 2006b.

KOBER, R.; SCHIFFERS, J.; SCHIMIDT, K.; Model-based versus knowledge-guided representation of non-rigid objects: A case study. **Proceeding of ICIP 94, IEEE**, 5 f., 1994.

KSHIRSAGAR, A. M.;KOCHERLAKOTA, S.; KOCHERLAKOTA, K., Classification procedures using principal components analysis and stepwise discriminant function. **Communication on Statistical Theory**, v. 19, p. 92-109, 1990.

LANITIS, A.; TAYLOR, C. J.; COOTES, C. F. An automatic face identification system using flexible appearance models. **Image and vision computing**, v. 13, p. 393-401, 1995.

LEME, R. J. A.; SHU, E. B. S.; CESCATO, V. A. S.; MACHADO, J. J.;SARKIS, P. S.; LEFREVE, B. Prosopagnosia após ferimento com arma de fogo. Arquivo Brasileiro de Neurocirurgia; p. 104-108; 1999.

LI, S.; JAIN, A. K. **Handbook of face recognition.** New York: Springer, 2005.

LIU, C.; WECHSLER, H. Comparative assessment of independent component analysis (ICA) for face recognition. In: **2nd International Conference on Audio and Video-based Biometric Person Authentication**, USA, March 1999.

MACHADO, A. M. C. **Metodologia para reconhecimento de padrões em visão computacional.** 1994. 99 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Minas Gerais, Minas Gerais, 16 de Dezembro de 1994.

MARR, D.; NISHIHARA, H. K. Representation and recognition of the spatial organization of three dimensional shapes. MIT AIM 416; May 1977.

MAURO, R.; KUBOVY, M. Caricature and Face Recognition. **Memory and cognition**, v. 20, p. 433-440, 1992.

MOGHADDAM, B.; PENTLAND, A. Probabilistic visual learning for object representation. In: **Proceedings of IEEE International Conference on Computer Vision**, p. 786-793, Cambridge, USA, June 1995.

MOGHADDAM, B.; WAHID, W; PENTLAND, A. Beyond eigenfaces: Probabilistic matching for face recognition. **IEEE International Conference on Automatic Face and Gesture Recognition**, Nara, Japan, April 1998.

MOGHADDAM, B.; JEBARA, T.; PENTLAND, A. Bayesian face recognition. PATTERN RECOGNITION, v. 33, p. 1771-1782, November 2000.

MOON, H.; PHILLIPS, P. J.; Computational and performance aspects of PCA-based face-recognition algorithms, **Perception** v. 30, p. 303-321, 2001.

OLIVEIRA JR., L. L.; THOMAZ, C. E.; Captura e alinhamento de imagens: Um banco de faces brasileiro. Relatório de iniciação científico, Departamento de Engenharia Elétrica da FEI, São Bernardo do Campo, SP, 10 f., 2006, disponível em www.fei.edu.br/~cet.

OSUMA, R. G, Principal components analysis. Lecture Notes 9. Texas A&M University, Texas, 2004a, disponível em www.couses-cs.tamu.edu/rgutier/cs790-w02, acessado em 20/12/2005.

OSUMA, R. G., Linear discriminant analysis. Lecture Notes 10. Texas A&M University, Texas, 2004b, disponível em www.couses-cs.tamu.edu/rgutier/cs790-w02, acessado em 20/12/2005.

O´TOOLE, A. J.; DEFFENBACHER, K. A.; VALENTIN, D.; ABDI, H. Structural aspects of face recognition and the other race effects, **Memory and cognition**, v. 26, p. 146-160, 1994.

PANTIC, M.; ROTHKRANTZ, J. M.; Automatic analysis of facial expressions: The state of art. **IEEE-PAMI**, v. 22, n. 12, p. 1424-1445, December 2000.

PEARSON, K. On lines and planes of closest fit to system of point in space**. Philosophical Magazine**, 2: p. 550-572, 1901, disponível em http://pbil.univ-lyon1.fr/R/pearson190.pdf, acessado em 06/05/2006.

PHILLIPS, P. J.; WECHSLER, H.; HUANG, J.; RAUSS, P. The FERET database and evaluation procedure for face recognition algorithms. **Image and Vision Computing Journal**, v. 16, n. 5, p. 295-306; 1998.

POOLE, D., **Álgebra linear**, São Paulo, Thomson, 1995.

RAUDYS, S. J.; JAIN, A. K. Small sample size effects in statistical pattern recognition: Recommendations for practitioners, **IEEE PAMI**, v. 13, n. 3, March 1991.

SAKAI, T.; NAGAO, M.; FUJIBAYASHI, S., Line extraction and pattern recognition in a photograph, **Pattern recognition**, v. 1, p. 233-248, 1969.

SERGENT, J., Micro genesis of face perception, In Aspects of face processing, H. D. ELLIS, M.A, 1996.

SCHILTZ, C.; ROSSION, B., Faces are represented holistically in the human occipital-temporal cortex, Neuroimage Elsevier v. 32, p. 1385-1396, May 2006.

SHEWCHUK, J. R., Triangle: engineering a 2D quality mesh generator and Delaunay triangulator. In Applied Computational Geometry. FCRC' 96 Workshop, p. 203-222, Springer Verlag, 1996.

SHLENS, J.; A tutorial on principal components analysis, Technical Report, Dec 2005, disponível em http://www.snl.salk.edu/nshlens/notes.html, acessado em 12/11/2006.

SIROVICJH, L.; KIRBY, M. Low-dimensional procedure for the characterization of human faces. **Journal of Optical Society of America**, v. 4, p. 519-524, March 1987.

STEGMANN, M. B., FISKER, R.; ERSBOLL, B. K.; THODBERG, H. H.; HYLDSTRUP, L. Active appearance models: Theory and cases. Department of Mathematical Modeling Technical University of Denmark, 2000.

STEWART, I. **Será que Deus Joga Dados? A nova matemática do caos**. Rio de Janeiro: Jorge Zahar Editor, 1991.

SWETS, D.; WENG, J. Using discriminant eigenfeatures for image retrieval. **IEEE PAMI** v. 18, n. 8, p. 831-836, August 1996.

TERZOPOULOS, D.; WATERS, K., Analysis and synthesis of facial image sequences using physical and anatomical models, **IEEE PAMI** v. 15, n. 6, p 569-579, June 1993.

TIPPING, M. E.; BISHOP, C. M.; Probabilistic Principal Components Analysis, **Journal of the Royal Statistical Society**, series B, 61, Part 3, p. 611-622, 1997.

THOMAZ, C. E. **Estudo de classificadores para o reconhecimento automático de faces.** 1999. 101 f. Dissertação (Mestrado em Engenharia Elétrica) – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica, Rio de Janeiro, 19 de Janeiro de 1999.

THOMAZ, C. E.; GILLIES, D. F. Small sample size: A methodological problem in Bayes plug-in classifier for image recognition. Technical report 6-2001, DOC, Imperial College of London, 2001.

THOMAZ, C. E. **Maximum entropy covariance estimate for statistical pattern recognition.** 2004. 152 f. These (Doctor of Philosophy Ph.D.) – Department of Computing, Imperial College London, University of London, London.

THOMAZ, C. E.; GILLIES, D. F.; A Maximum uncertainty LDA-based approach for limited sample size problems with application to face recognition. **In Proceedings of SIBGRAPI' 05, IEEE CS** Press, p. 89-96, 2005.

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. **In Proceedings of SIBGRAPI' 05, Journal of Brazilian Computer Society (JBCS) IEEE CS Press**, v. 12, n. 2, p. 89-96, June 2005.

THOMAZ, C. E.; AGUIAR, N. O.; OLIVEIRA, S. H. A.; DURAM, F. L. S; BUSATTO, G. S.; GILLIES D. F. ; RUECKERT D. Extraction discriminant information from medical images: A multivariate linear approach. In: **Proceedings of SIBGRAPI' 06, IEEE CS Press**, 2006.

TURK, M.; PENTLAND, A. Eigenfaces for recognition. **Journal of Cognitive Neuroscience, MIT**, v. 3, n. 1, p. 71-86, March 1991.

TURK, M. Computer Vision in the interface. COMMUNICATIONS OF THE ACM; v. 47, n. 1, p. 61-67, January 2004.

YANG, M. H.; KRIEGMAN, D. J.; AHUJA, N. Detecting faces in images: A survey; **IEEE-PAMI**, v. 24, n. 1, p. 34-58, January. 2002.

YU, J.; TIAN, Q., Constructing descriptive and discriminant features for face classification, **IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)**, p. 14-19, Toulouse France, May 2006.

WANG, H.; AHUJA, N. Facial expression decomposition. **In: ICCV**, p. 958-965, 2003.

WEN-YI, Z., Face recognition: A tutorial, European Conference on Computer Vision, 2004

ZANA, Y.; CESAR JR., Face recognition based on polar frequency features, ACM Transaction on Applied Features, v. 3, issue 1, p. 62-82, January 2006.

ZHAO, W.; CHELLAPPA, R.; KRISHNASWAMY, SWETS, D. L., WENG, J. A. Discriminant analysis of principal components for face recognition. IN: PROC. INTERNACIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, p. 336-341, 1998.

ZHAO, W; CHELLAPPA, R. ROSENFELD, A.; PHILLIPS, P. J. Face recognition: A literature survey. Technical Report; University of Maryland, USA, 2000.

ZHUANG, X. S.; DAI, DQ.; YUEN, P. C., Face recognition by inverse Fisher discriminant features, **ADVANCES IN BIOMETRICS, PROCEEDINGS,** 3832: p. 92-98, **LECTURE NOTES IN COMPUTER SCIENCE,** 2006.

APÊNDICE A – Aplicativo FACES.EXE

APÊNDICE A

O aplicativo denominado FACES2.EXE inicialmente foi desenvolvido como um trabalho da disciplina de visão artificial do curso de mestrado da FEI. Posteriormente o aplicativo evoluiu para a versão atual, cujos recursos permitiram a execução de todos os experimentos desta dissertação. O aplicativo FACES2.EXE foi desenvolvido em uma linguagem gráfica denominada LabView® 7.1 juntamente com o módulo IMAQ Vision® 7.1.1 do fabricante National Instruments™. A *interface* com o usuário é executada em LabView® 7.1, entretanto, os cálculos matriciais são feito pelo software Matlab® 5.0 do fabricante Mathworks™ e toda a comunicação entre os dois aplicativos para a troca de variáveis reais e complexas são executados em Activex® da Microsoft™. A escolha desta configuração se deve ao fato de que o LabView® mostrou-se lento nos cálculos matriciais, principalmente com matrizes tão grande quanto aquelas que se trabalha no domínio de faces. No entanto, as interfaces produzidas pelo LabView® são mais "amigáveis" e permitem expansões rápidas e flexíveis. Todo o programa foi compilado e executado em um computador padrão PC com processador Athlon 64 2.2 GHz com 1Gb de memória RAM e sistema operacional Windows XP.

O aplicativo FACES2.EXE permite a leitura de imagens no formato JPG em qualquer dimensão, coloridas ou monocromáticas, e as ajusta internamente na escala e as torna monocromáticas. A leitura dos arquivos pode ser seqüencial ou não, e ainda a matriz de faces lidas pode ser salva no formato TXT permitindo o uso posterior. O aplicativo é formado por módulos denominados VI's (*Virtual Instruments*), e cada uma dessas VI's pode executar uma ou mais funções. Algumas dessas VI's são funções bibliotecas pré-definidas ou então são blocos de funções construídas pelo programador. De um modo geral, a programação em LabView® segue todo os conceitos e protocolos de uma linguagem procedural e textual. O leitor interessado em maiores detalhes sobre o LabView® deve inicialmente consultar (REGAZZI et al., 2005) e (JOHNSON, 1997) e para o Matlab® deve consultar (HANSELMAN; LITTLEFIELD, 2004).

Dentro do programa FACES2.EXE encontraremos módulos dedicados aos cálculos do PCA, LDA e MLDA, bem como sub-rotinas que calculam e apresentam visualmente as transformações da face média e as transformações de uma face com identidade, a partir da face média transformada. Através das abas no topo da *interface* com o usuário, pode-se escolher que tipo de resultado numérico e visual desejamos acompanhar. Tocando-se virtualmente com o apontador do *mouse* numa das abas, seleciona-se a função desejada. A primeira aba é a aba inicial, por onde se lêem as imagens de faces individualmente ou então o arquivo com as imagens de face previamente salvas. Pode-se ainda percorrer toda matriz das imagens arquivadas

na matriz de faces, ou então localizar uma imagem específica a partir da posição dela na matriz de faces. Esta aba fornece também informações sobre o total de imagens arquivadas na matriz de faces bem como o último arquivo de imagem lido. Na figura A-1 a seguir observa-se a primeira aba de entrada de dados do programa.

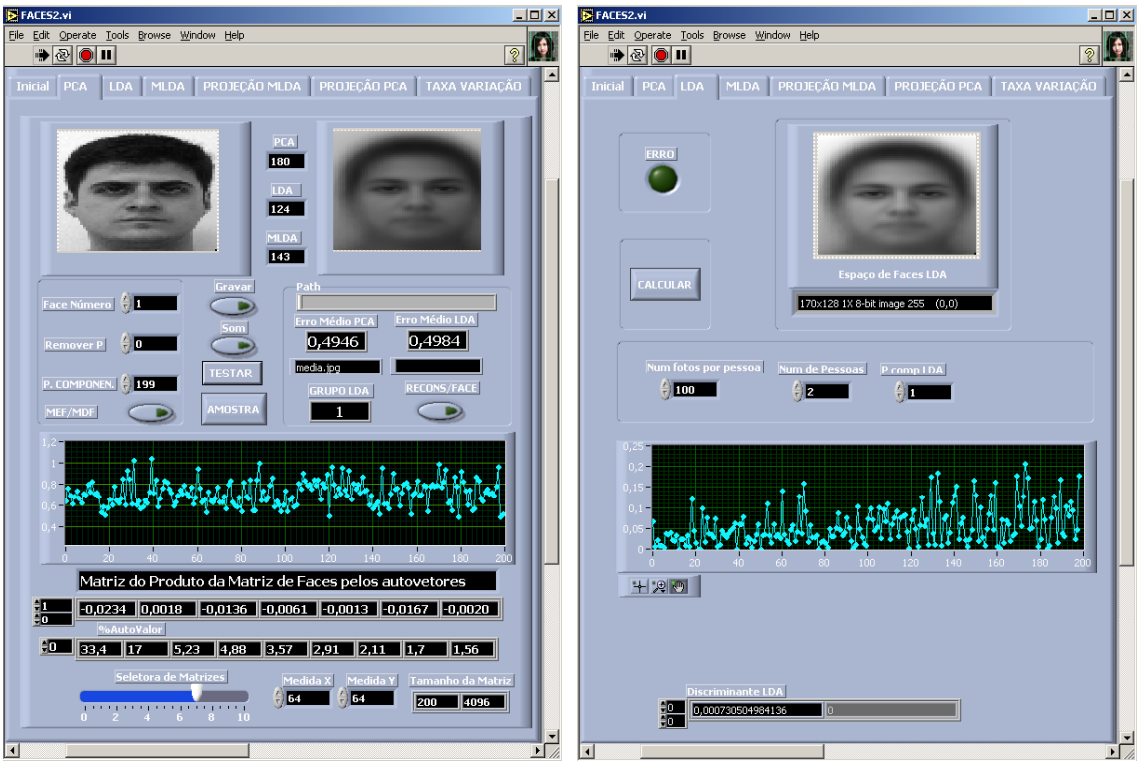
Figura A-1 – Imagem da *interface* de entrada de dados do programa FACES2.EXE.

A aba seguinte é denominada PCA, onde pode-se acompanhar o resultado de todos os cálculos da fase PCA. Através dessa é possível observar todas as imagens de faces contidas na matriz de faces, bem como os autovetores, a reconstrução de qualquer imagem de face com um número p de componentes principais, ou mesmo removendo-se as p primeiras componentes principais. Pode-se também visualizar o conteúdo numérico das várias matrizes geradas durante o cálculo do PCA.

A figura A-2 (a) a seguir representa a aba relativa aos resultados numéricos e visuais do cálculo com o PCA. Além de apresentar os resultados dos cálculos com o PCA, esta aba também possui uma função de comparação de qualquer imagem de teste com todas as imagens do conjunto de treinamento, utilizando a função critério de menor distância vetorial, com as matrizes dos três espaço gerados, PCA, LDA e MLDA. O gráfico dentro da *interface* representa a menor distância entre a imagem de teste com todas as imagens do espaço PCA.

De maneira similar à aba PCA, a aba LDA apresenta os resultados das matrizes do LDA de Fisher. A figura A-2 (b) apresenta essa *interface*. No entanto, o gráfico dentro *da interface* da figura A-2 (b) não representa a menor distância entre a imagem de teste projetada

no espaço LDA com todas as imagens do espaço LDA. Esse gráfico representa na realidade os pesos do vetor W_{LDA} em relação a cada componente principal.



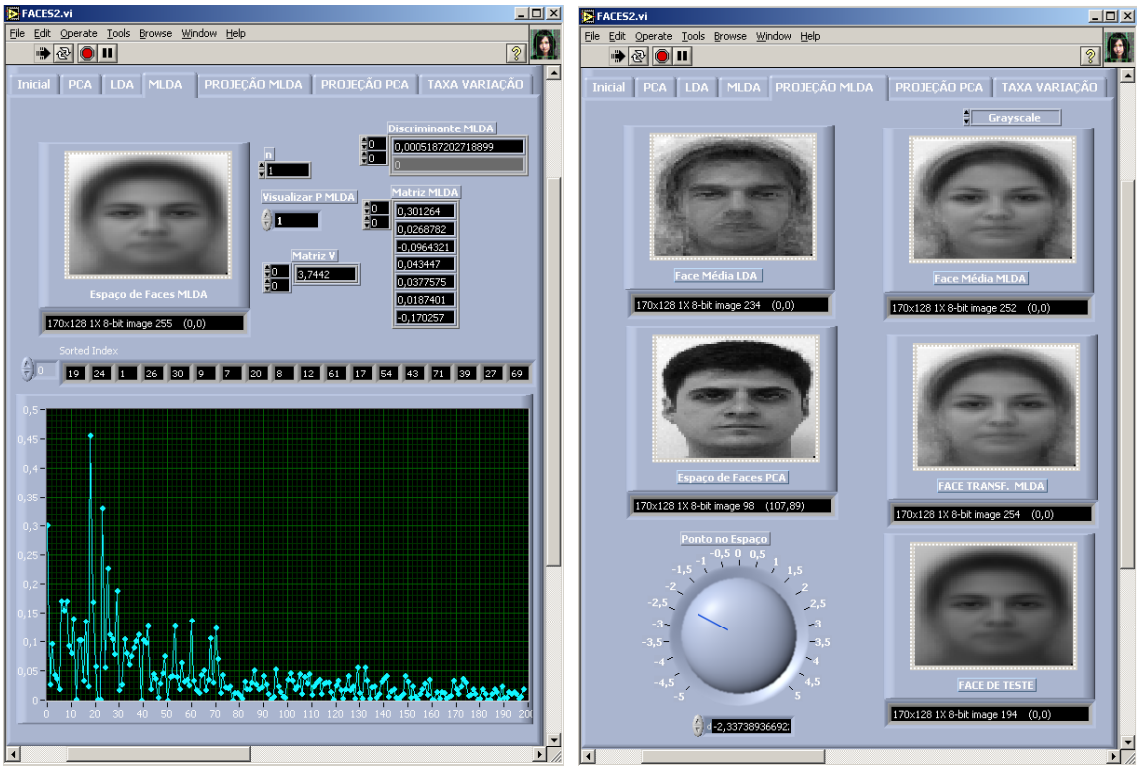
(a) (b)

Figura A-2 - Imagem da *interface* PCA, figura (a), e da *interface* LDA, figura (b).

As duas abas seguintes são relativas ao módulo do MLDA. A aba denominada MLDA faz todo o cálculo das projeções das imagens de face do espaço PCA para o espaço MLDA, e apresenta o resultado das matrizes do MLDA. O gráfico contido dentro dessa *interface* representa o vetor de pesos W_{MLDA} relativos a cada componente principal. Acima do gráfico vê-se quais são os números das componentes principais que são ordenadas conforme os maiores pesos do vetor W_{MLDA}. A aba seguinte denominada "PROJEÇÃO MLDA", faz a navegação e reconstrução das imagens de face no espaço MLDA. O botão chamado "Ponto no espaço" representa exatamente o ponto y_i^{MLDA}, de maneira que girando-se o botão no sentido horário ou anti-horário, variamos o valor de y_i^{MLDA} dentro dos limites de ± 5. Entretanto, caso se deseje navegar além desse limite, basta digitar no campo abaixo do botão o valor para y_i^{MLDA}.

Esse procedimento faz com que se observe visualmente a variação na face média quando se navega ao longo do hiperplano discriminante. Se eventualmente uma determinada face com identidade for lida previamente, poderá se observar as transformações que ocorrerão também na face com identidade, a medida que se varia o valor de y_i^{MLDA}. É também na aba

"Projeção MLDA" que observamos visualmente as transformações em três cores de modo a destacar as regiões com taxa de variações maiores, assim como está descrito no apêndice "B".

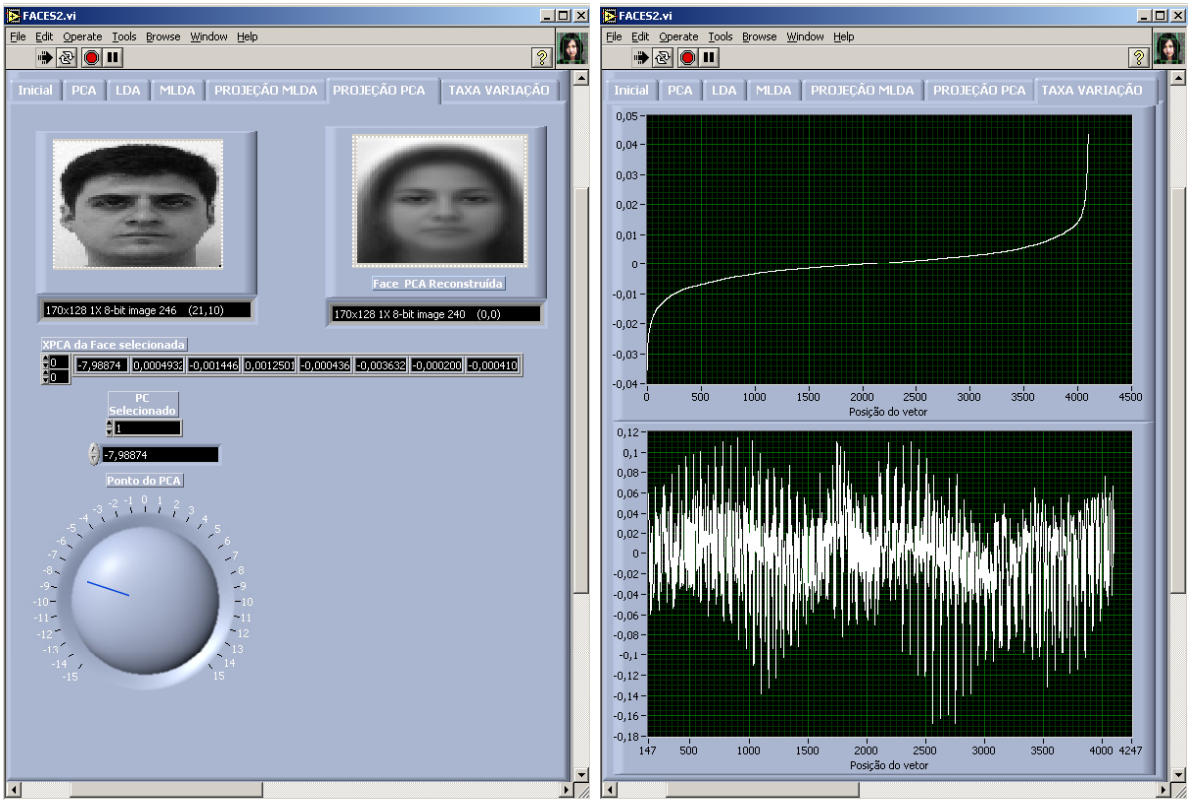


(a) (b)

Figura A-3 - Imagem da *interface* MLDA, figura (a), e da *interface* que faz a navegação e transformação quando se percorre o hiperplano discriminante, figura (b).

A aba denominada PCA possui o recursos para se navegar por todas as componentes principais, uma de cada vez. Basta selecionar a componente principal que se deseja avaliar e mover o botão denominado "Ponto do PCA". Variando-se esse botão no sentido horário ou anti-horário, observaremos as transformações que ocorrem na autoface selecionada, assim como foi descrito no capitulo 5. A ilustração da figura A-5 (a) apresenta a *interface* com o usuário para a navegação nas componentes principais.

A aba seguinte denominada "Taxa Variação" apresenta a curva da taxa de variação quando navegamos ao longo do hiperplano discriminante e também o vetor da taxa de variação ordenada crescentemente. A ilustração da figura A-4 (b) apresenta a *interfaces* da aba da "Taxa Variação". Estes gráficos são gerados quando movimenta-se o botão de navegação da aba MLDA. Observa-se pela figura A-4 (b) que temos apenas os gráficos, sendo assim, é necessário um movimentação entre as duas abas para se observar o resultado.



(a) (b)

Figura A-4 - Imagem da *interface* PCA figura (a), e da *interface* "Taxa Variação" que mostra as curvas da taxa de variação e do vetor da taxa ordenado crescentemente, figura (b).

As figuras a seguir ilustram as partes principais do código em LabView® e Matlab® do programa FACES2.EXE. Algumas partes menos relevantes do código foram suprimidas apenas para se racionalizar o espaço desta dissertação. Entretanto, o leitor interessado pode requisitar a cópia completa do código do programa para cet@fei.edu.br.

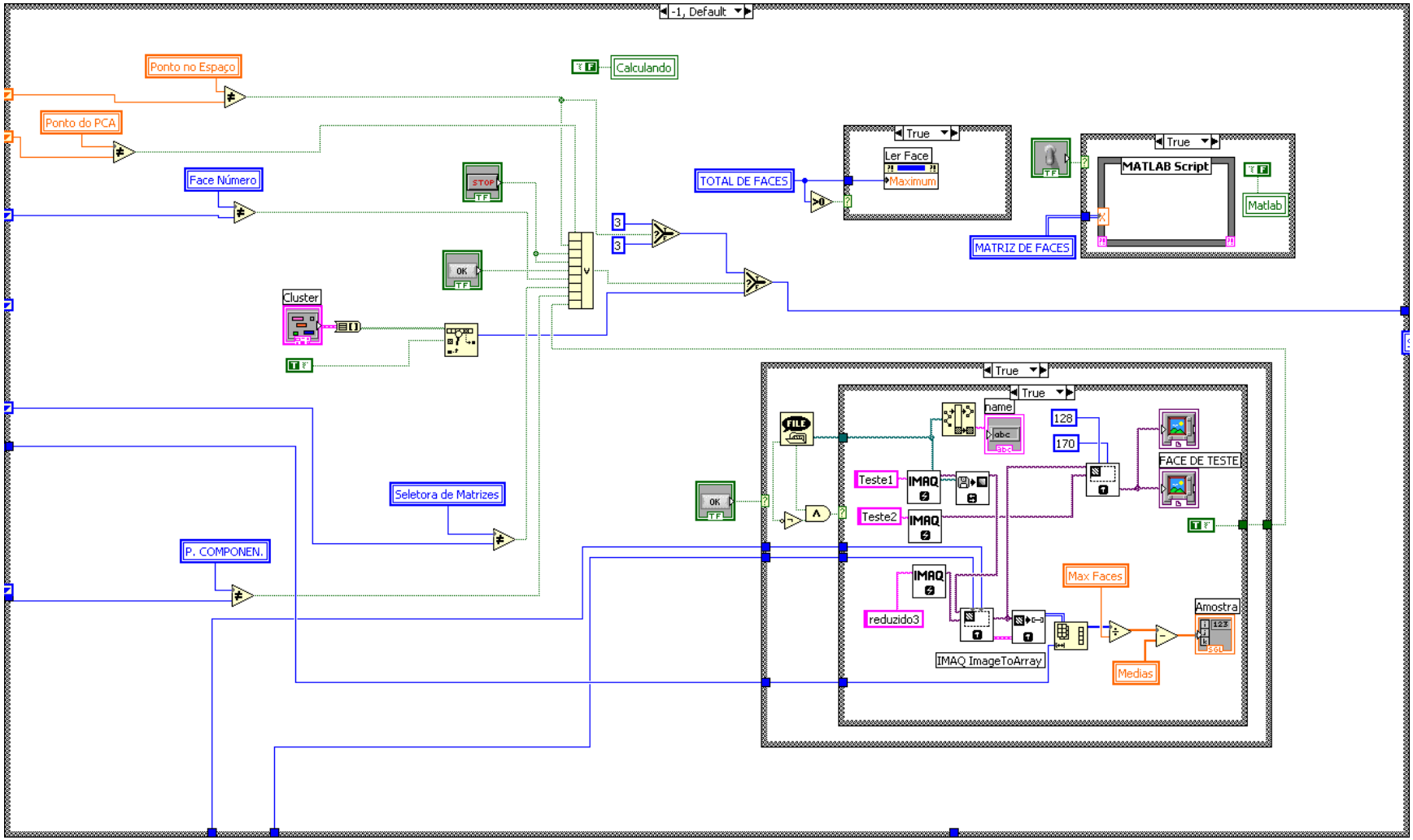


Figura A-5 - Laço principal que faz a seleção das abas de operação e também lê o arquivo solicitado.

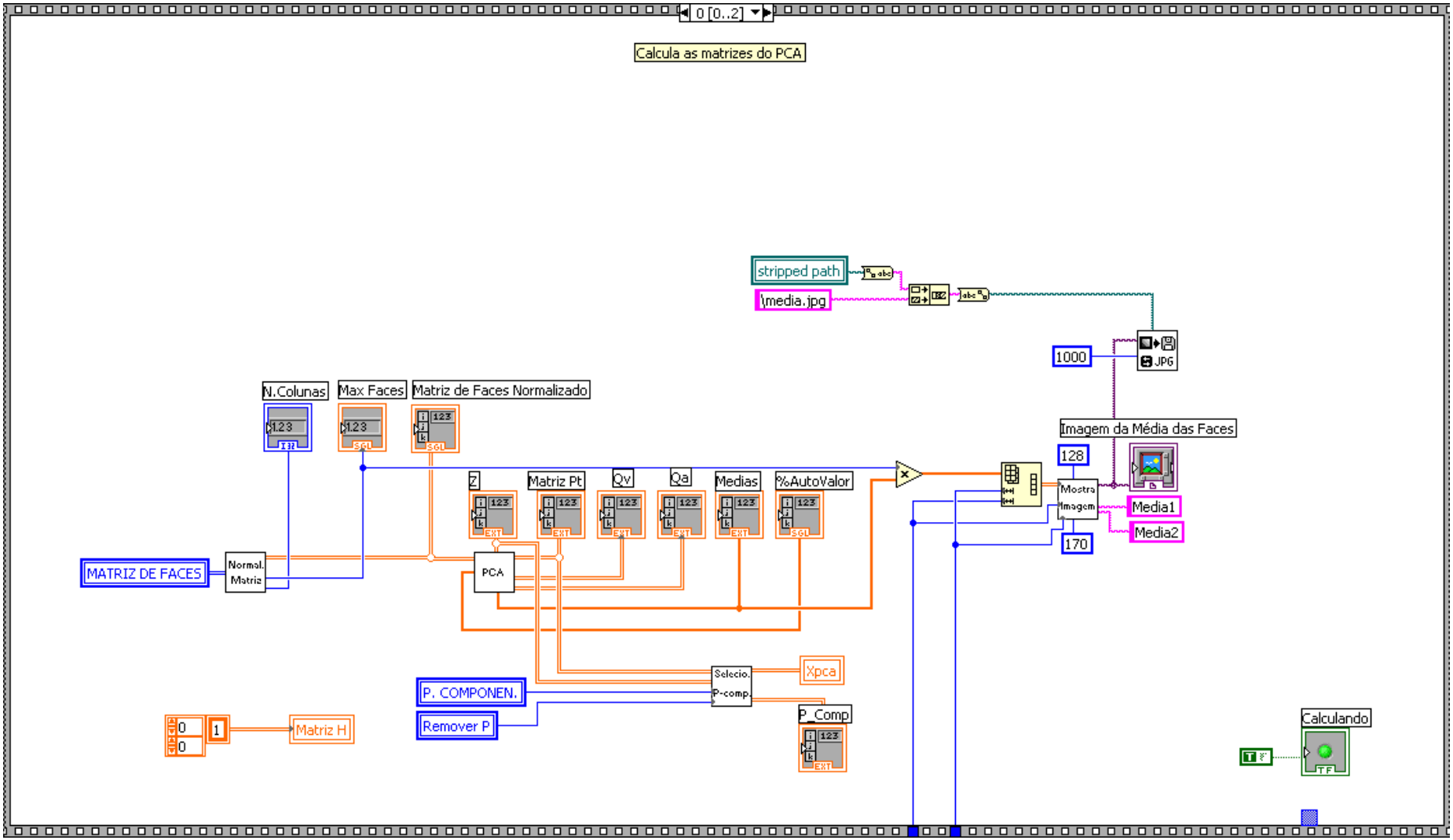


Figura A-6 - Rotina que envia a matriz de face para a rotina que calcula o PCA.

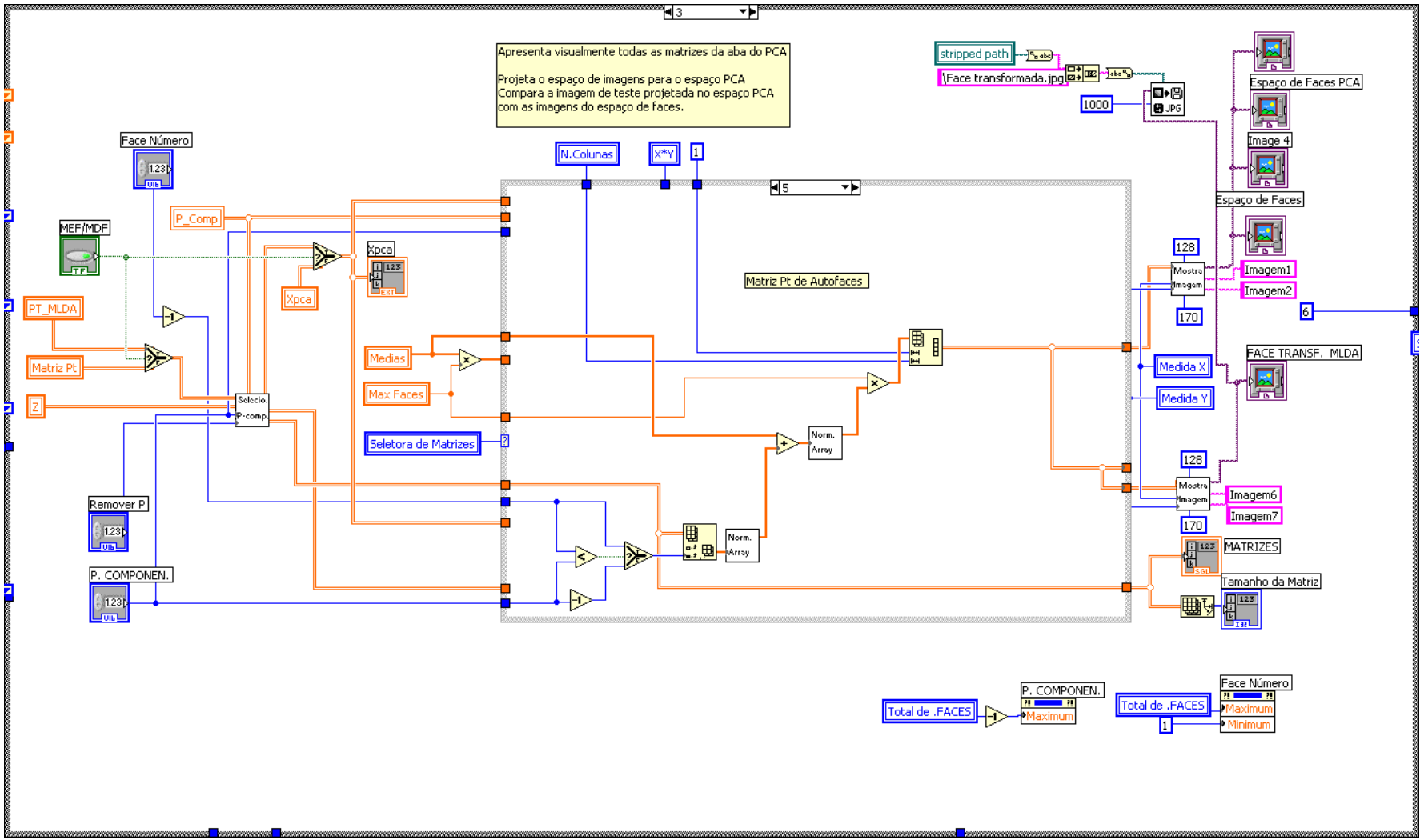


Figura A-7 - Rotina do laço principal que permite a visualização, na aba PCA, de todas as matrizes calculadas no PCA, por exemplo as autofaces.

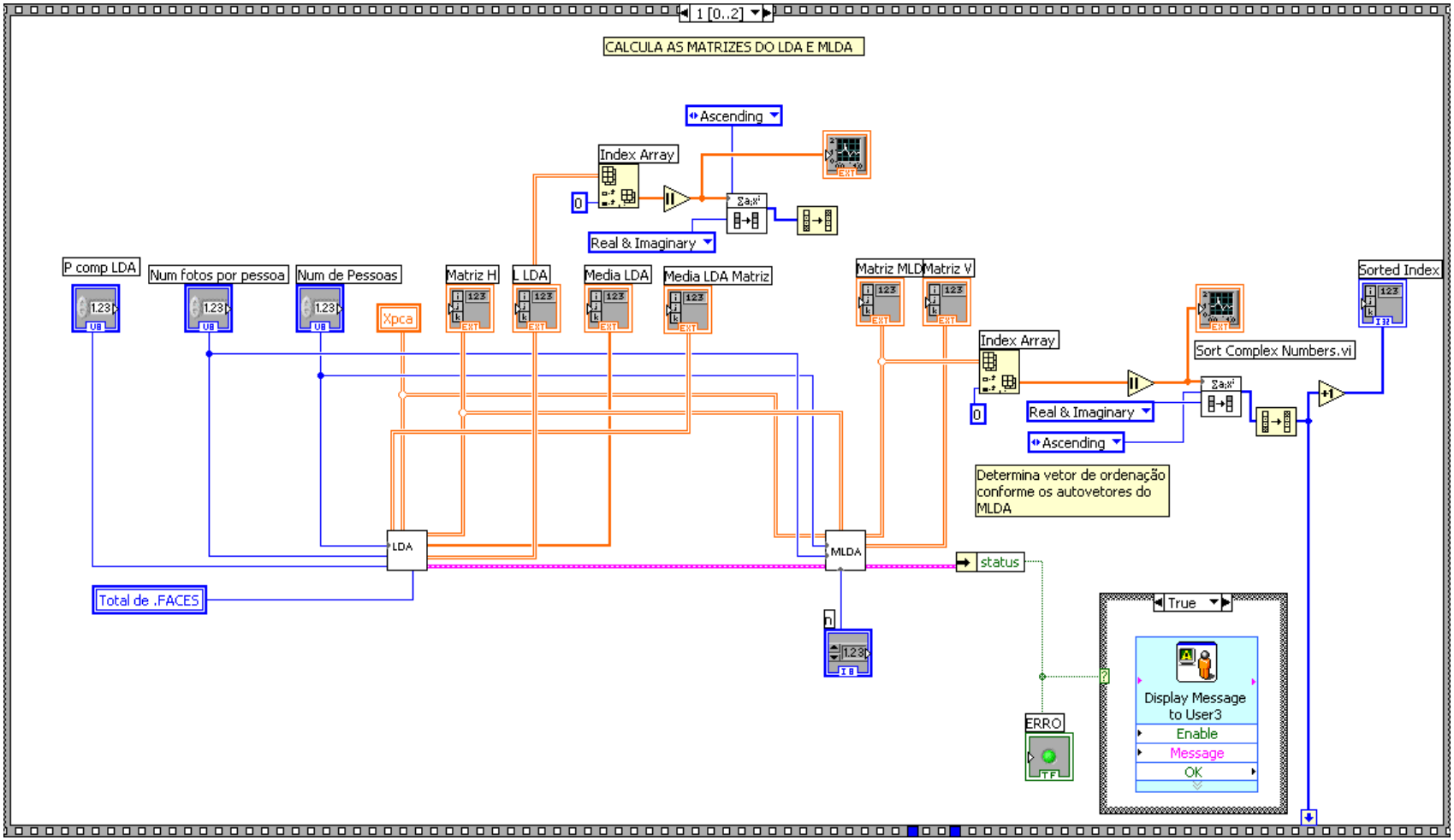


Figura A-8 - Esta rotina envia os dados para se calcular as matrizes do espaço LDA e MLDA a partir do espaço PCA. A rotina também determina qual seria a nova ordenação dos autovetores conforme o peso da matriz *W* calculada na rotina MLDA.

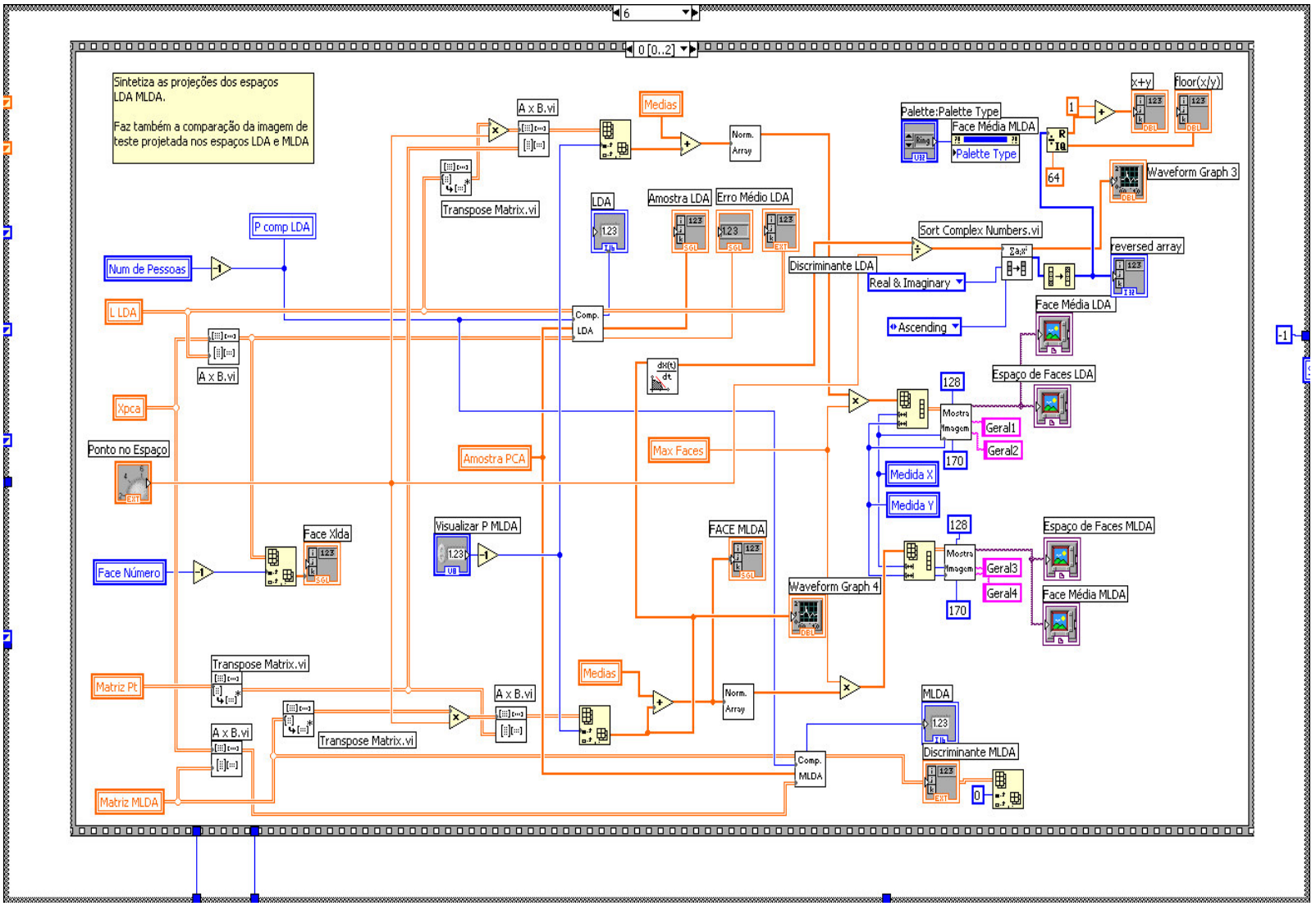


Figura A-9 - Rotina que calcula e sintetiza a transformação da face média e da face com identidade e também calcula a taxa de variação.

As figuras abaixo representam as VI's (*Virtual Instruments*) de três rotinas principais que efetuam os cálculos das matrizes dos espaço PCA, LDA e MLDA. As figuras seguintes representam os códigos em LabView® e Matlbab® de cada uma dessa subrotinas. Cada caixa indica as variáveis que são lidas e escritas em cada rotina. Os detalhes sobre o que faz cada variável encontram-se nas próximas figuras. As próximas figuras ilustram em detalhes o conteúdo de cada VI representado na figura A-10

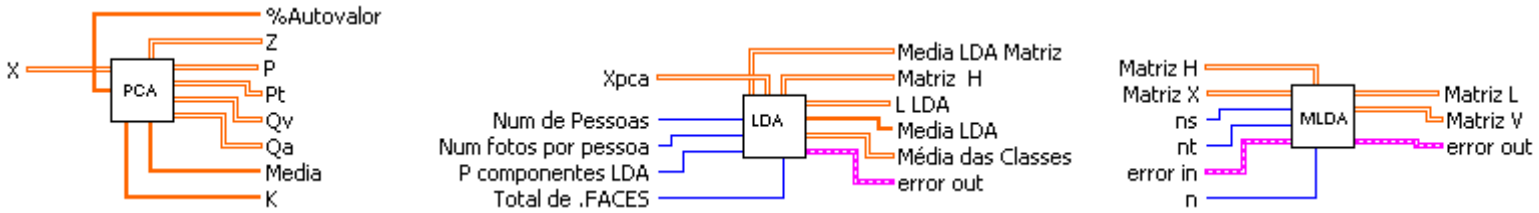


Figura A-10 – VI's relativas a cada uma das operações estatísticas que calculadas no programa FACES.EXE.

.

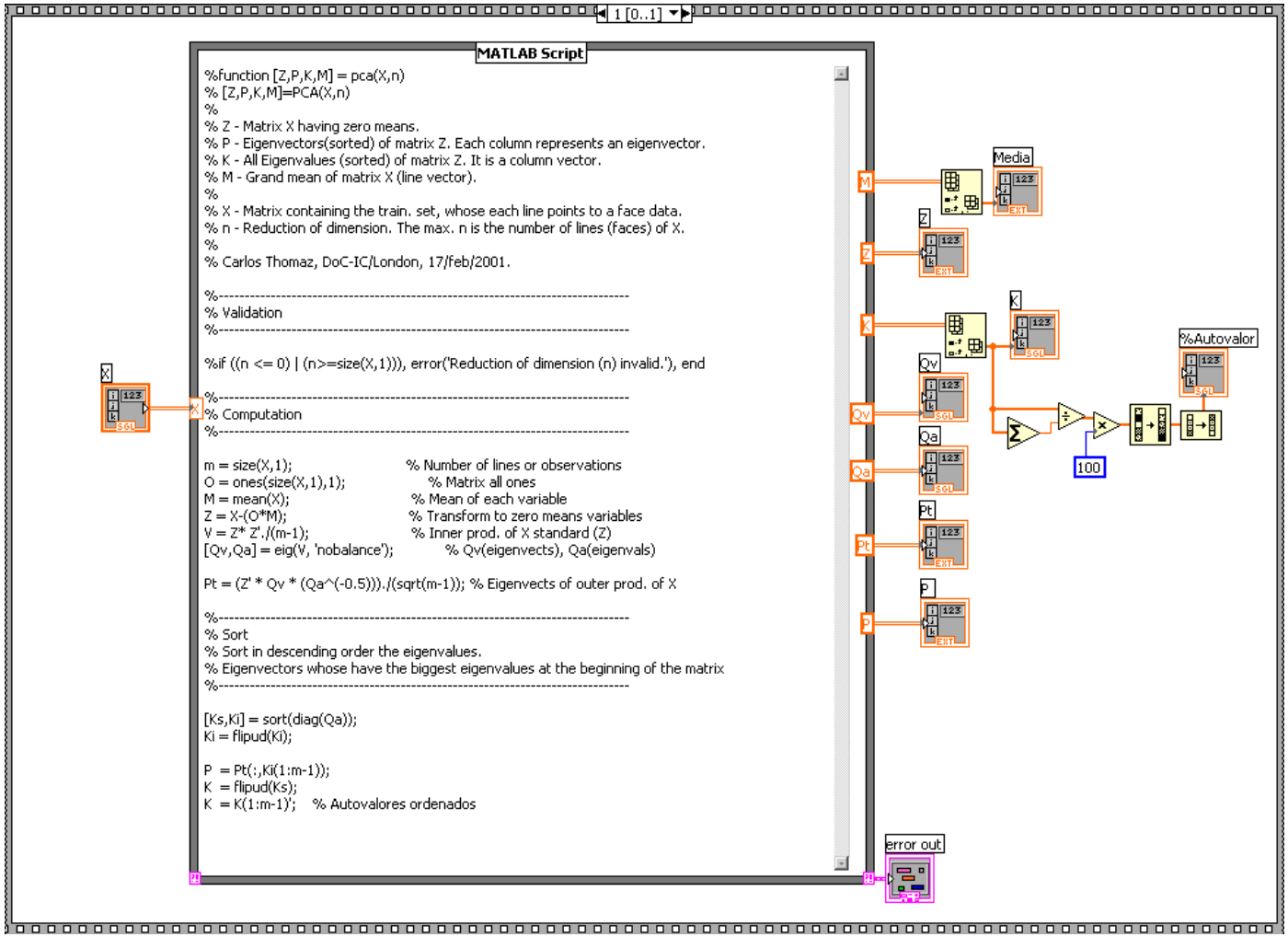


Figura A-11 – Código Matlab® para a determinação das matrizes do PCA (1).

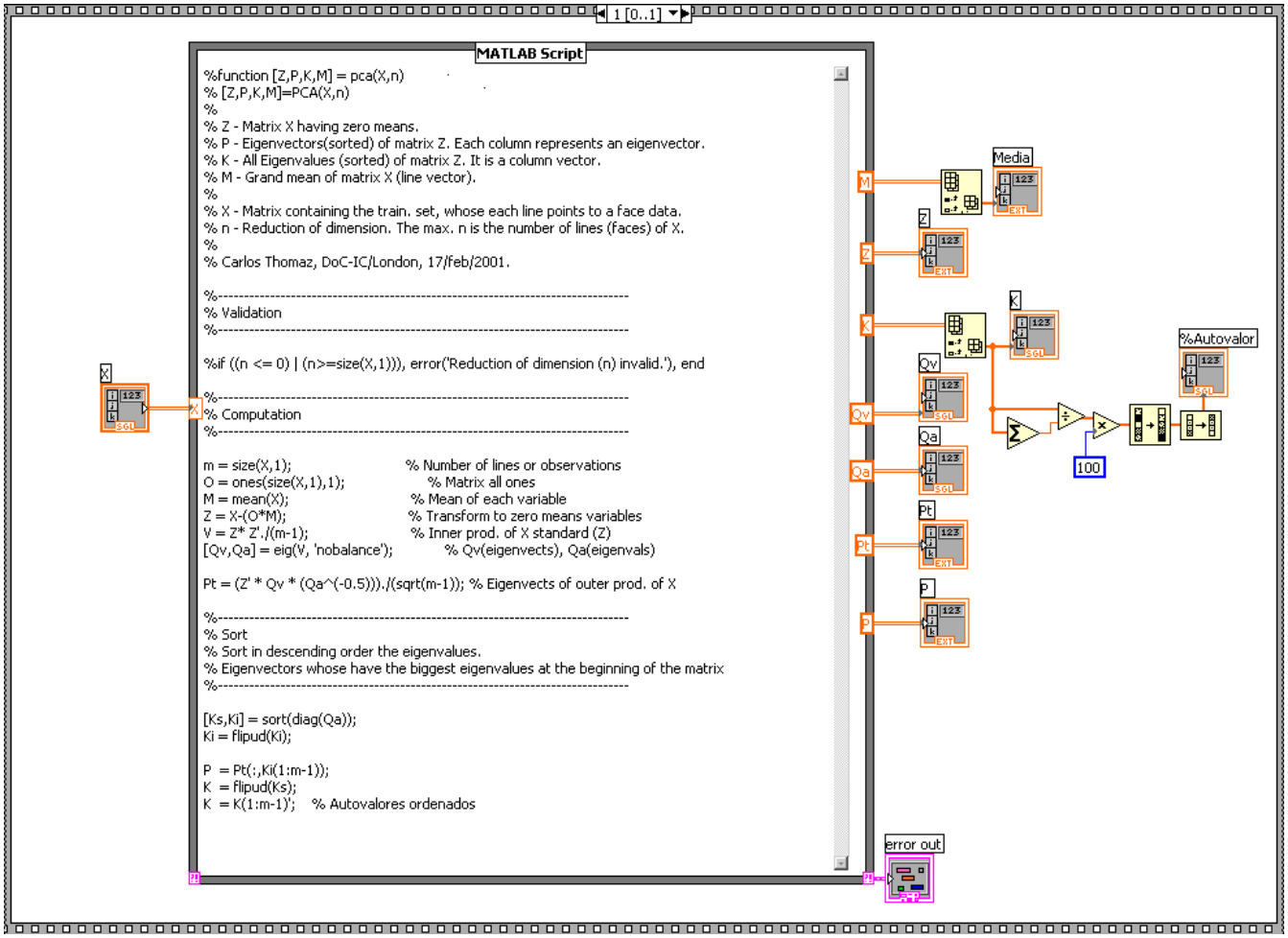


Figura A-12 - Código Matlab® para a determinação das matrizes do PCA (2).

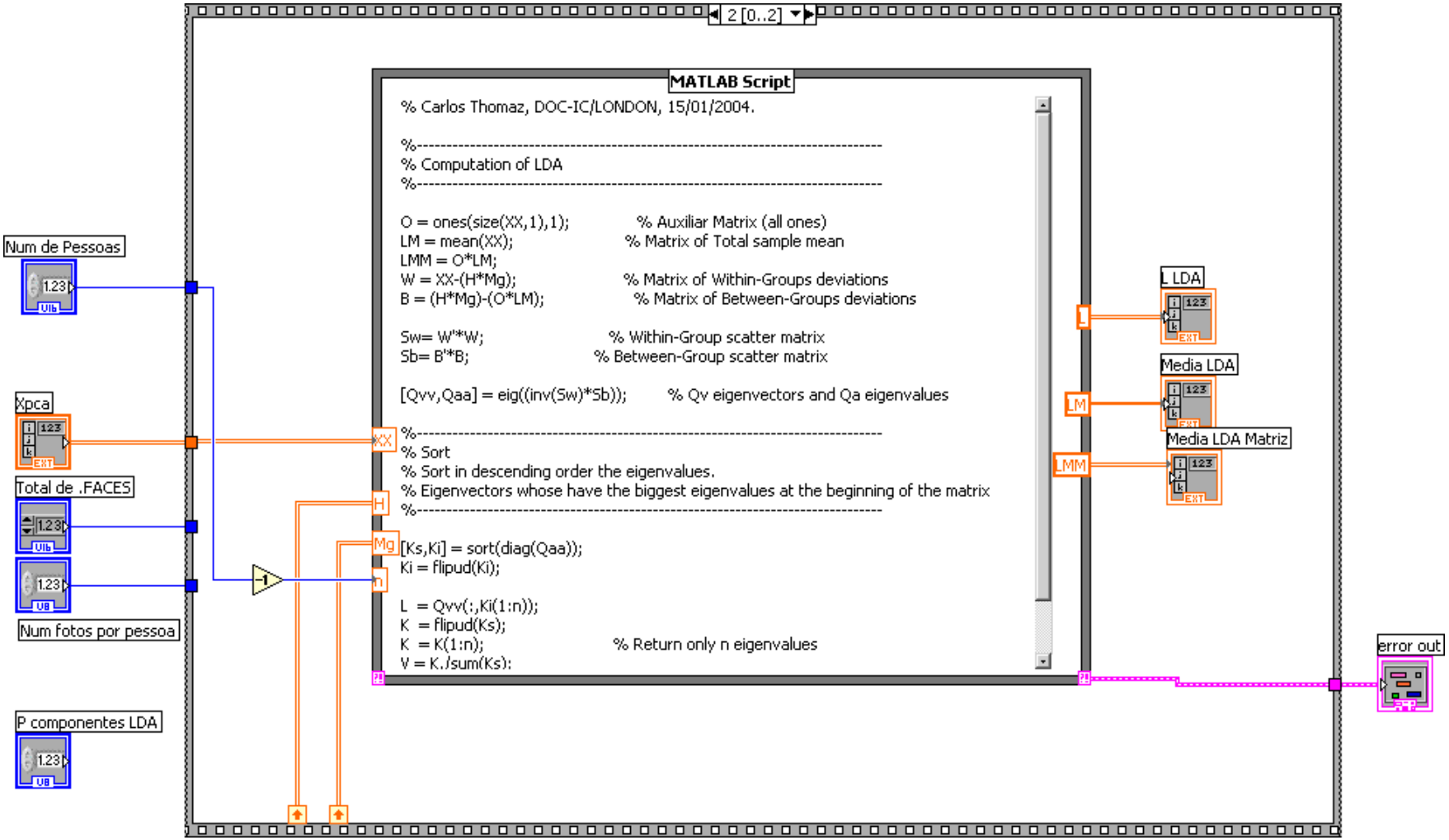


Figura A-13 - Código LabView® e Matlab® para a determinação das matrizes do LDA (1).

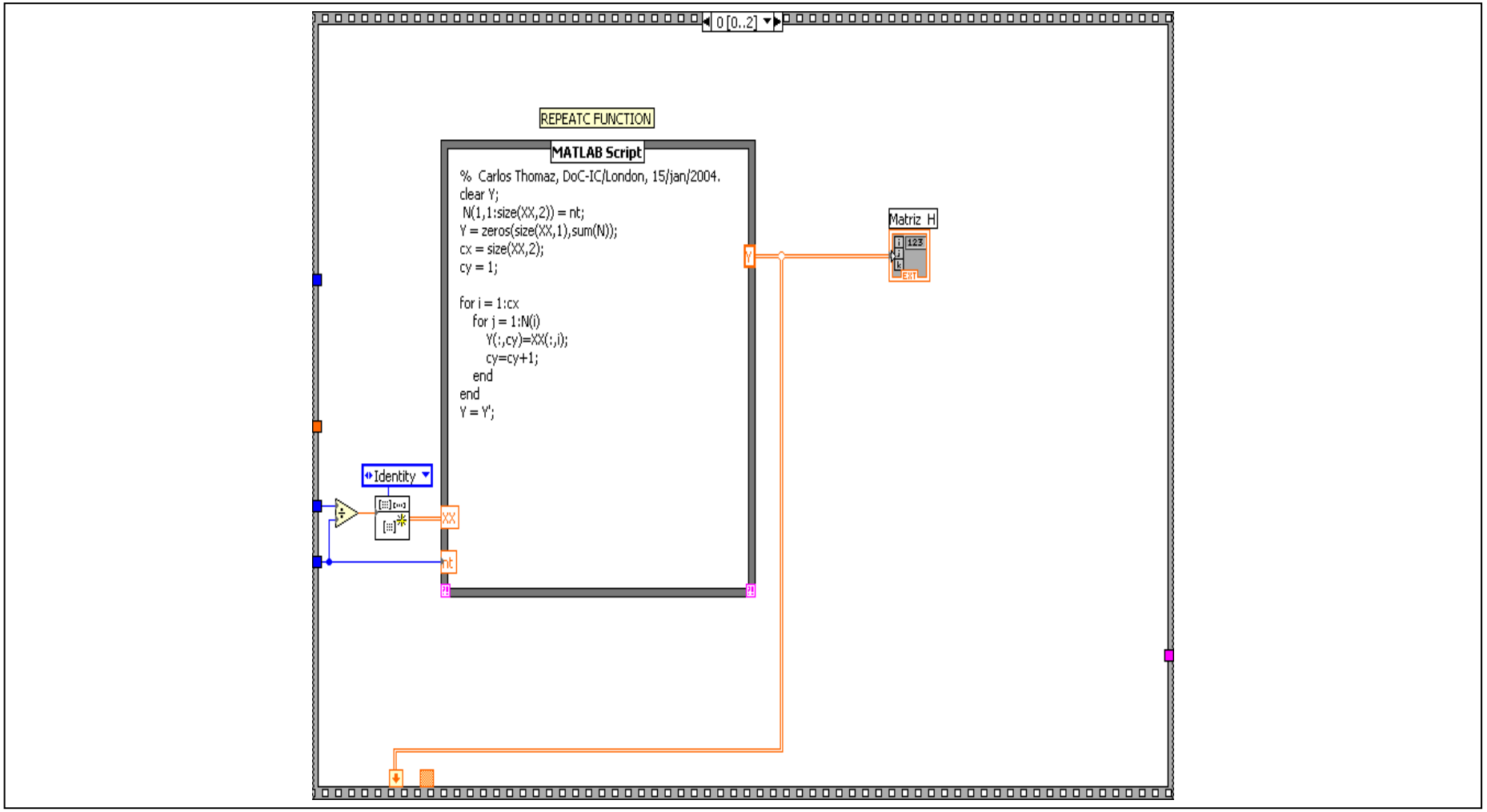


Figura A-14 - Código LabView® e Matlab® para a determinação das matrizes do LDA (2).

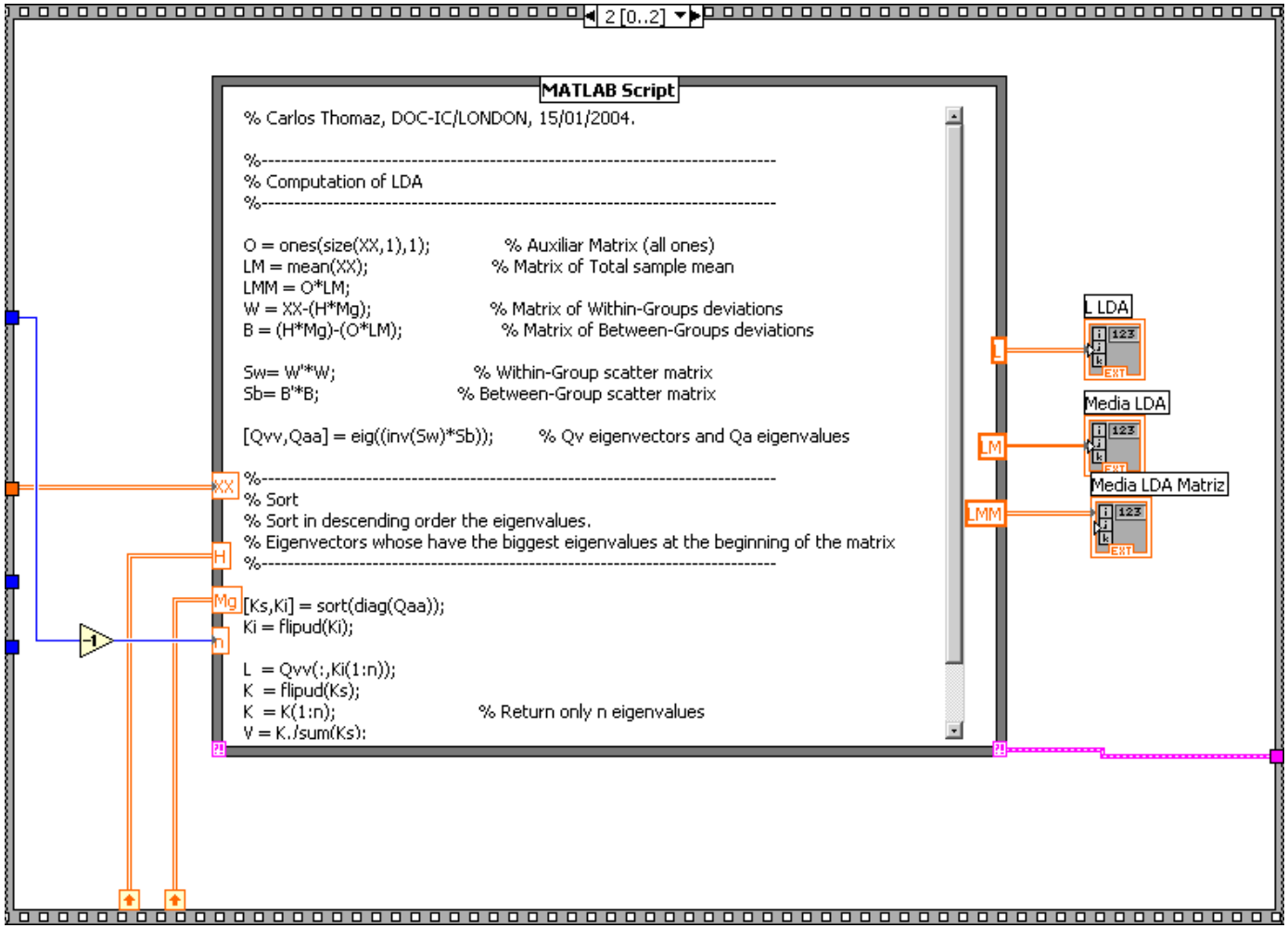


Figura A-15 - Código LabView® e Matlab® para a determinação das matrizes do LDA (3).

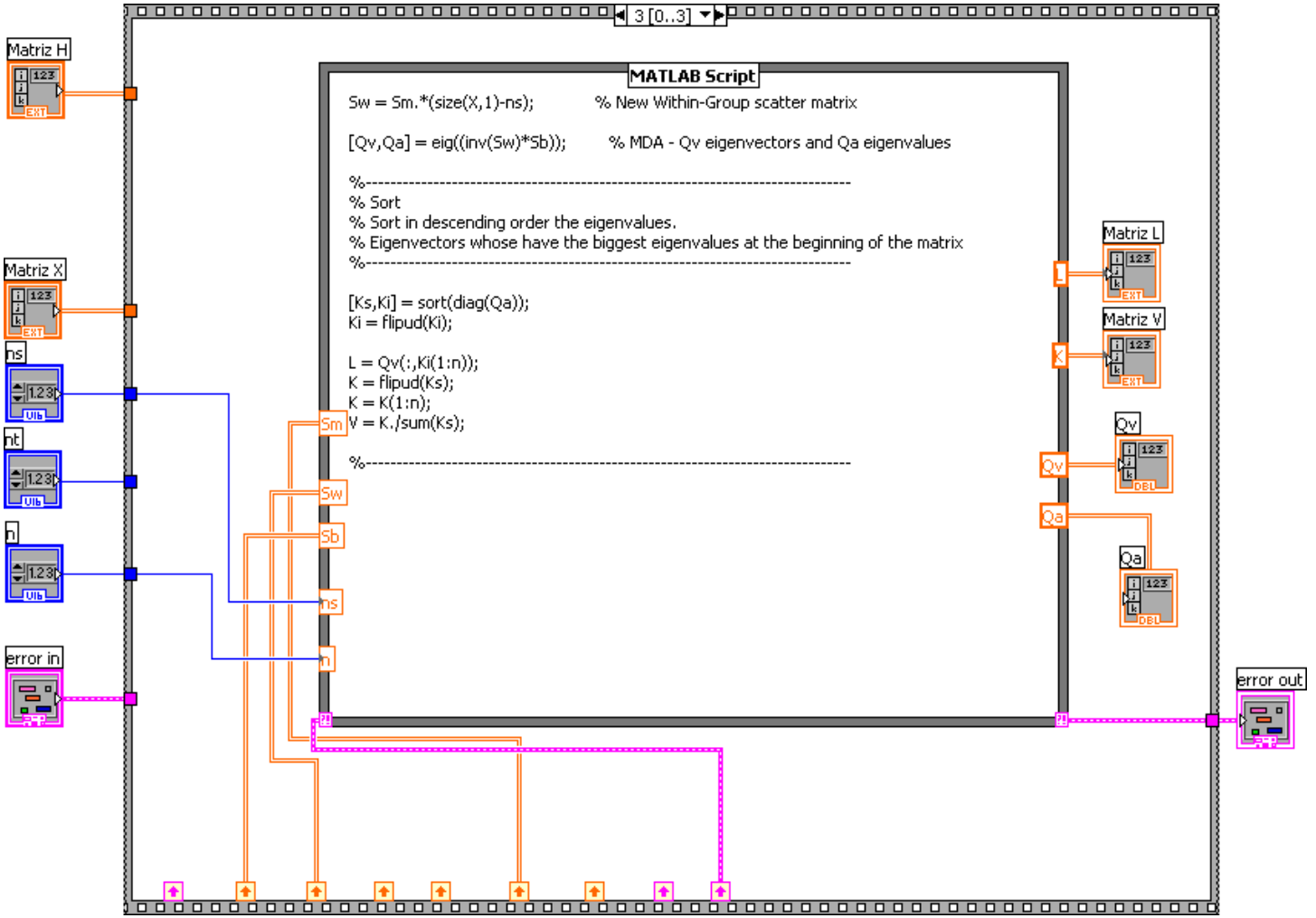


Figura A-16 - Código LabView® e Matlab® para a determinação das matrizes do LDA (4).

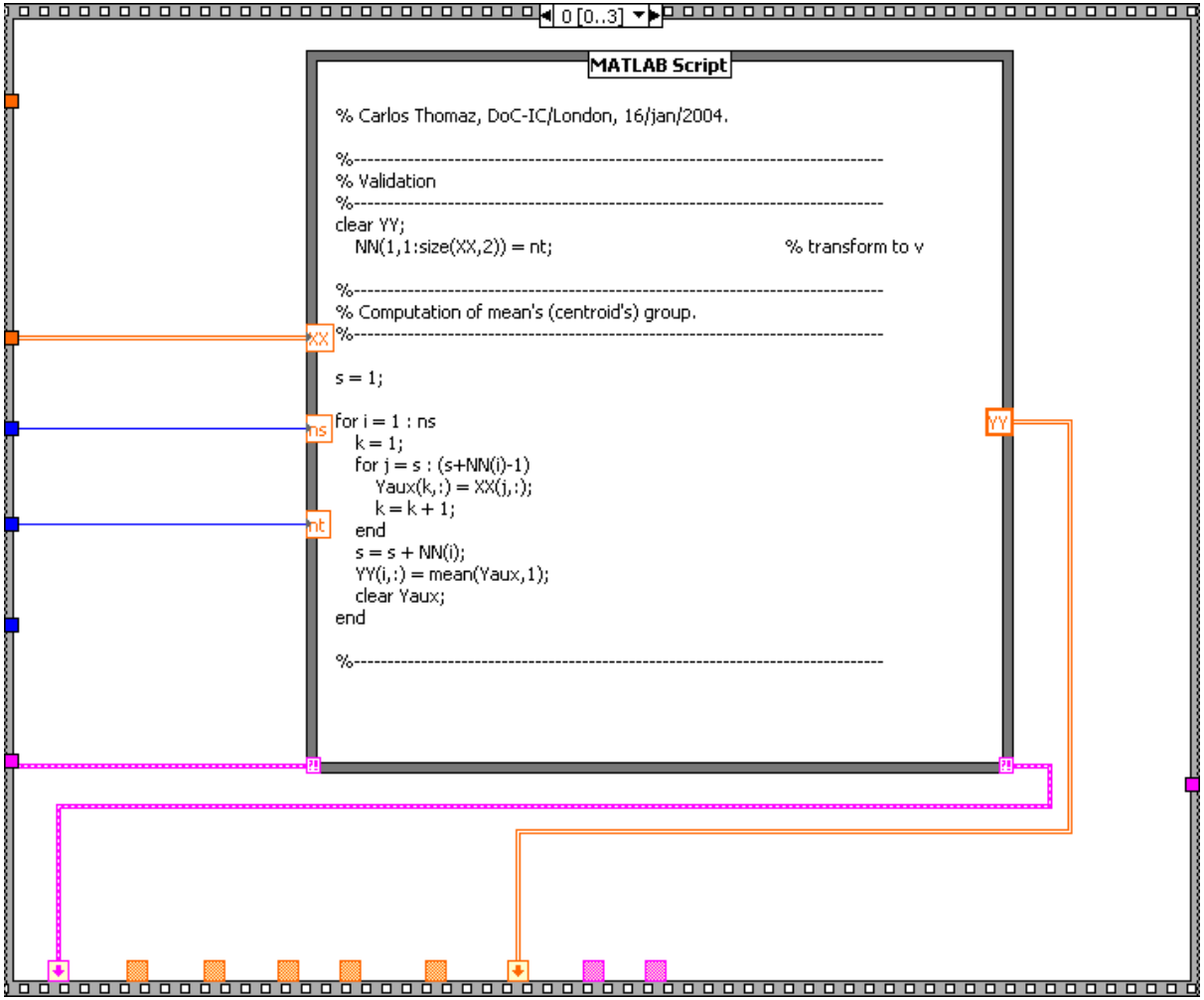


Figura A-17 - Código LabView® e Matlab® para a determinação das matrizes do MLDA (1).

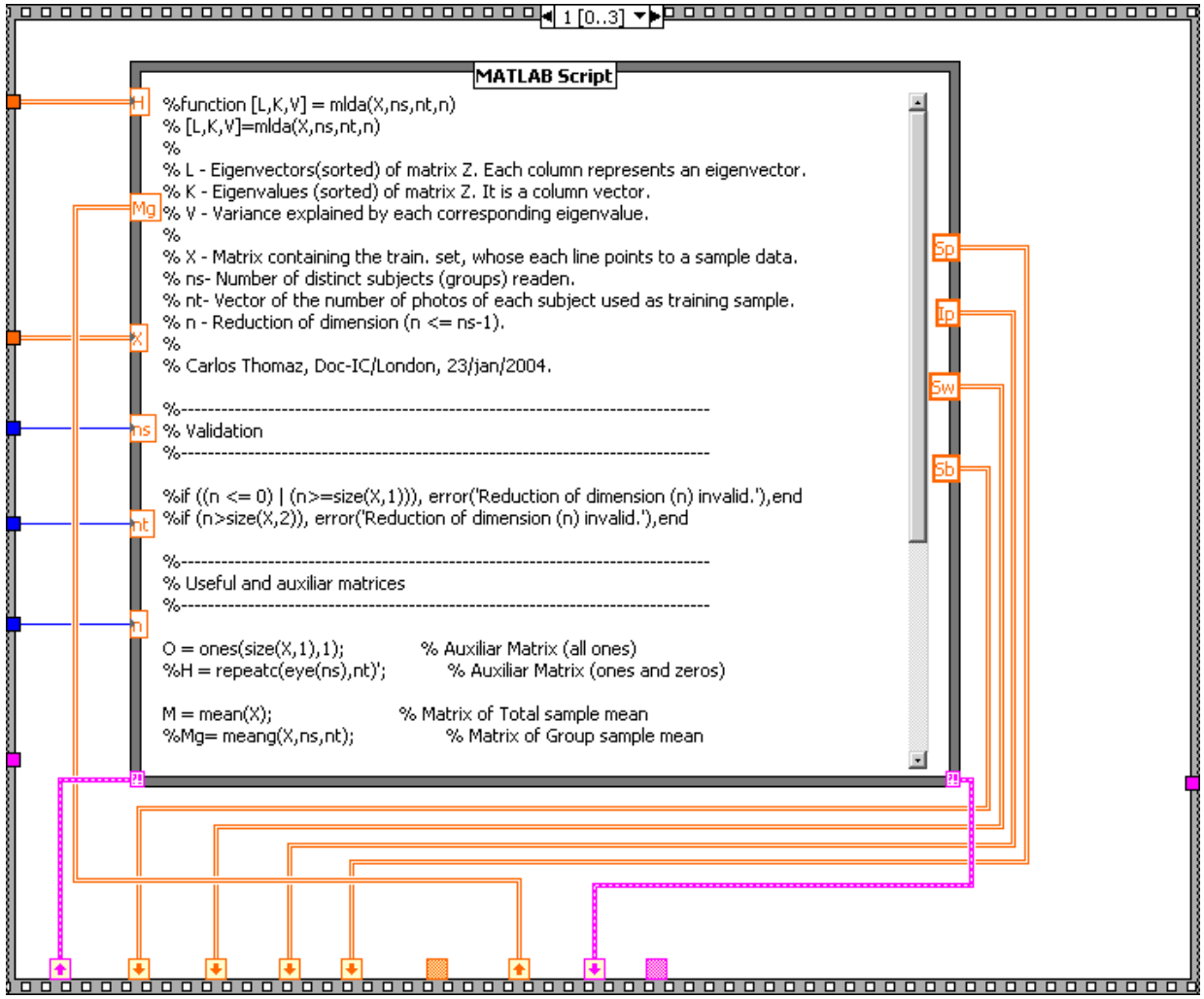


Figura A-18 - Código LabView® e Matlab® para a determinação das matrizes do MLDA (2).

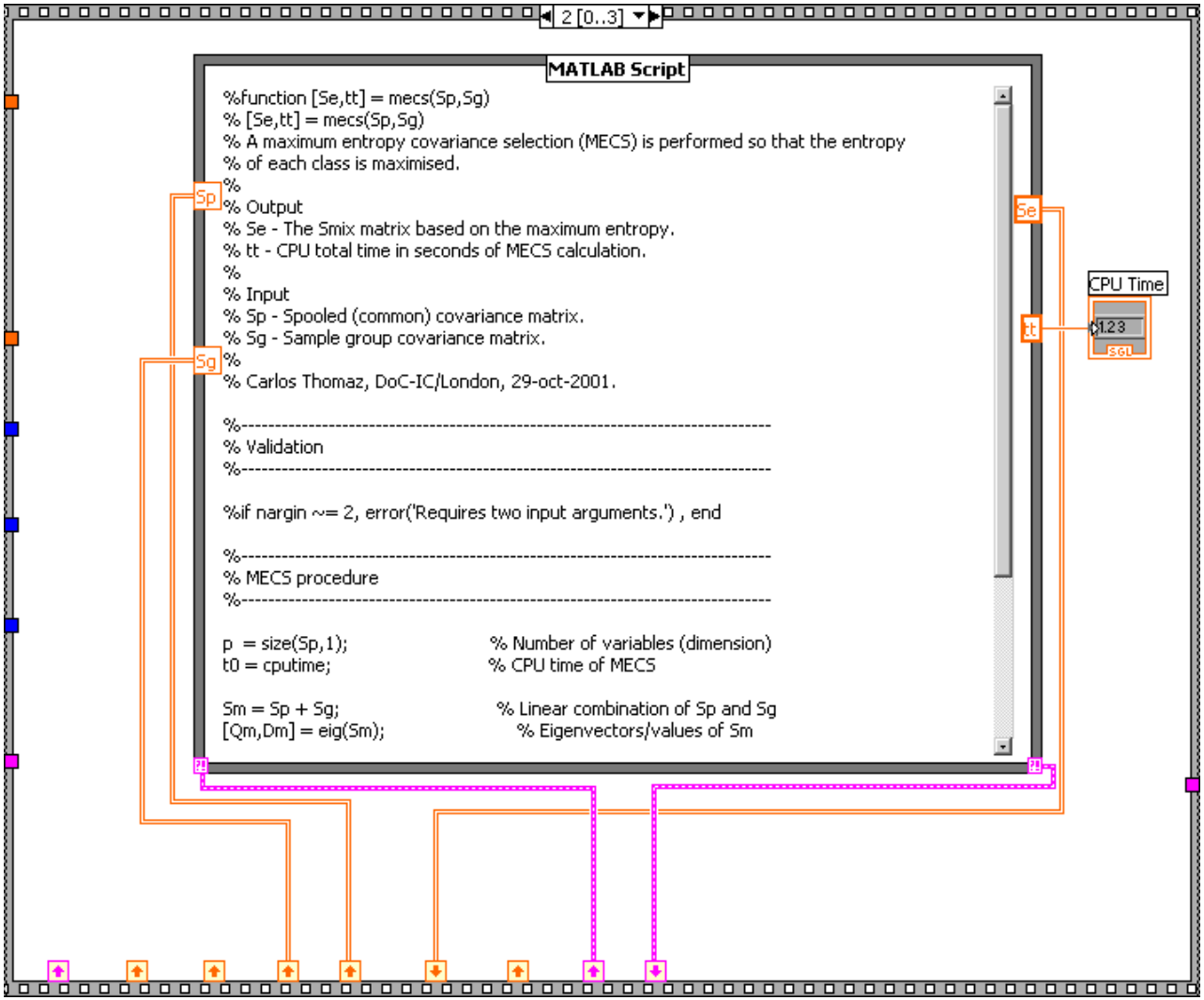


Figura A-19 - Código LabView® e Matlab® para a determinação das matrizes do MLDA (3).

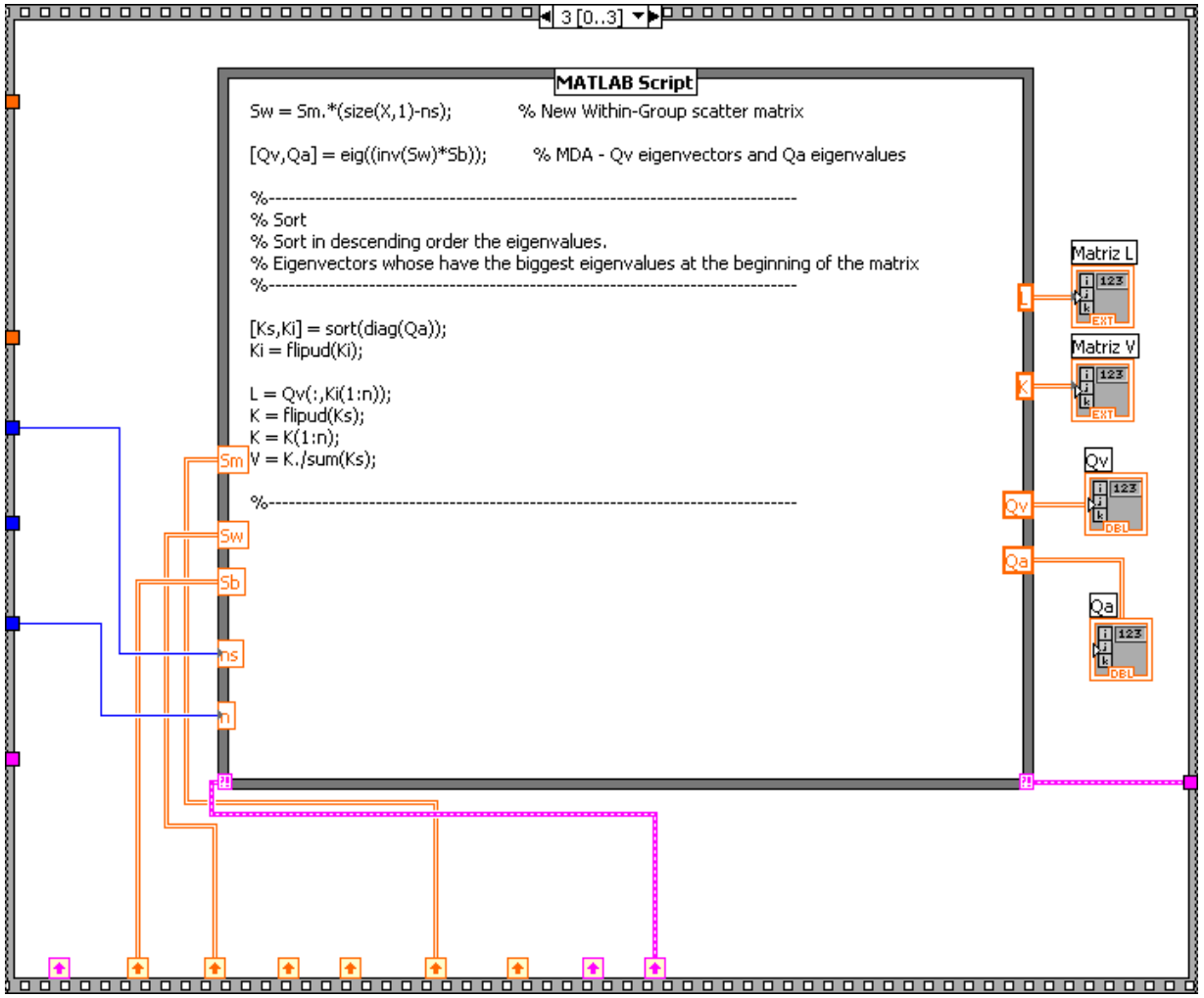


Figura A-20 - Código LabView® e Matlab® para a determinação das matrizes do MLDA (4).

APÊNDICE B – Tabela de transformação

APÊNDICE B

Neste apêndice é descrito a formação da tabela de transformação dos níveis de cinza de uma imagem monocromática, para um modelo de representação em três cores. O objetivo é destacar as variações que ocorrem na imagem que quando vistas em tons de cinza não trazem muita informação.

A representação discreta de uma imagem de face monocromática é uma função $f(x, y) = i(x, y)r(x, y)$, onde (x, y) são as coordenadas espaciais de cada ponto da imagem limitadas a $x \leq num\_colunas$ e $y \leq num\_linhas$, i é o valor da intensidade luminosa ou luminância, limitado a $0 \leq i \leq \infty$, e r é a taxa de refratância de cada ponto limitado a $0 \leq r \leq 1$. Entretanto, para imagens digitais monocromáticas convencionou-se que a função $f(x, y)$ seria limitada a $0 \leq f(x, y) \leq 255$. Desta maneira, as variações de luminosidade capturadas em uma imagem discreta seguem um modelo linear de representação (GONZALEZ; WOODS, 1992). A curva ilustrada na figura B-1 representa a função $f(x, y)$ em relação a intensidade luminosa *i*.

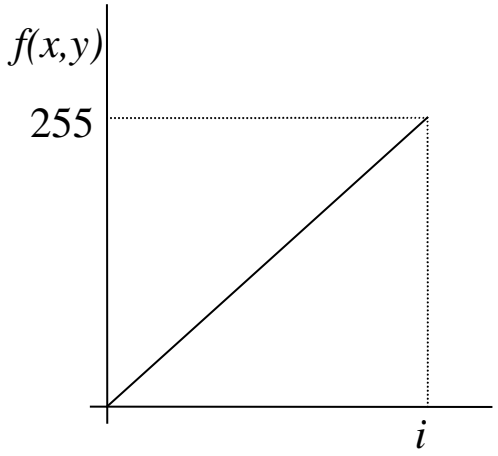


Figura B-1 – Curva da função $f(x, y)$ em relação a intensidade luminosa *i*.

Entretanto, este tipo de representação pode dificultar a análise visual de informações que têm variações bruscas ou sutis, que não necessariamente tem uma variação linear nessas transições. Portanto, para permitir uma melhor interpretação visual desses fenômenos utiliza-se outras funções de transformação que possam representar e enfatizar essas transições. Um modelo de representação em três cores utilizado nesta dissertação é denominado *Rainbow Pallete* (NATIONAL INSTRUMENTS, 1999), que faz a representação das imagens monocromáticas combinando-se as três cores básicas RGB (*Red, Green, Blue*) conforme um conjunto de novas funções de transformação baseadas no índice de iluminação *i*. Como cada componente de cor básica comporta-se como uma função $f(x, y)$, considera-se que existe uma relação de *i* para cada componente RGB. A equação (B.1) representa então a primeira

função de transformação para a componente R (*Red)*, e de maneira similar as equações (B.2) e (B.3) representam as funções para as componentes G e B respectivamente.

$$f(R) = \begin{cases} 0 \Leftrightarrow i \le 128 \\ (i-128)4 \Leftrightarrow 128 < i \le 192 \\ 255 \Leftrightarrow i \ge 192 \end{cases} \tag{B.1}$$

$$f(G) = \begin{cases} 4i \Leftrightarrow i < 64 \\ 255 \Leftrightarrow 64 \le i \le 192 \\ (256-i)4 \Leftrightarrow i > 192 \end{cases} \tag{B.2}$$

$$f(B) = \begin{cases} 255 \Leftrightarrow i < 64 \\ (128-i)4 \Leftrightarrow 64 < i < 128 \\ 0 \Leftrightarrow i \ge 128 \end{cases} \tag{B.3}$$

O resultado gráfico da combinação das três equações (B.1), (B.2) e (B.3) pode ser observado na figura B-2 a seguir.

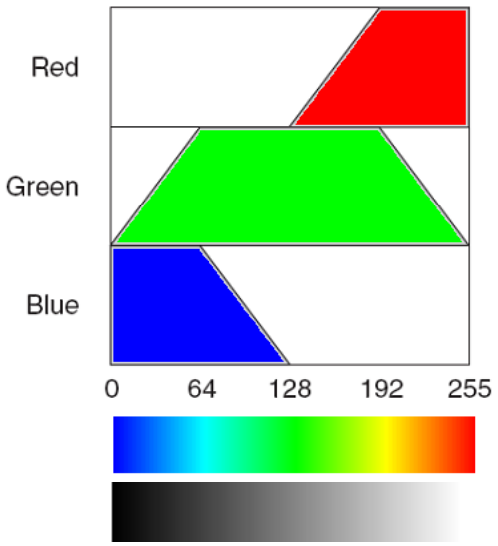


Figura B- 2 – Gráfico da curva de transformação de monocromático para três cores RGB. Adaptado de (IMAQ Vision Concept Manual, 2004).

REFERÊNCIAS DOS APÊNDICES

GONZALEZ, R. C.; WOODS, R. E., **Digital Image Processing**, Addison-Wesley Publishing Company, 1992.

HANSELMAN, D.; LITTLEFIELD, B. Matlab 6 Curso Completo, São Paulo, 2004, Prentice Hall.

JOHNSON, G. W. LabView Graphical Programming, 2nd ed. New York, Mc Graw-Hill, 1999.

NATIONAL INSTRUMENTS – Imaq Vision User Manual Part Number 322320B – May 1999.

REGAZZI, R.G.; PEREIRA, P. S.; SILVA Jr.; M. F., Soluções práticas de instrumentação e automação, Rio de Janeiro, Ed. GRAFICAKWG, 2005.

APÊNDICE C – Artigo do SIBGRAPI 2006

A Statistical Discriminant Model for Face Interpretation and Reconstruction

Edson C. Kitani[1], Carlos E. Thomaz[1], and Duncan F. Gillies[2]
[1]*Department of Electrical Engineering, Centro Universitário da FEI, São Paulo, Brazil*
[2]*Department of Computing, Imperial College, London, UK*
[1]{ekitani,cet}@fei.edu.br, [2]d.gillies@imperial.ac.uk

Abstract

Multivariate statistical approaches have played an important role of recognising face images and characterizing their differences. In this paper, we introduce the idea of using a two-stage separating hyper-plane, here called Statistical Discriminant Model (SDM), to interpret and reconstruct face images. Analogously to the well-known Active Appearance Model proposed by Cootes et. al, SDM requires a previous alignment of all the images to a common template to minimise variations that are not necessarily related to differences between the faces. However, instead of using landmarks or annotations on the images, SDM is based on the idea of using PCA to reduce the dimensionality of the original images and a maximum uncertainty linear classifier (MLDA) to characterise the most discriminant changes between the groups of images. The experimental results based on frontal face images indicate that the SDM approach provides an intuitive interpretation of the differences between groups, reconstructing characteristics that are very subjective in human beings, such as beauty and happiness.

1. Introduction

The most successful statistical models for visual interpretation of face images have been based on Principal Component Analysis (PCA) [1, 3, 10]. These approaches have used as features either shapes [3] or textures [10] alone, or a combination of both [1]. Unfortunately, however, even in the PCA approach based on a combination of features, the sources of the shapes' and textures' variations have to be isolated in order to extract and interpret the most expressive differences in the training samples. For instance, in the well-known Active Appearance Model proposed by Cootes et. al. [1] the shape model is dissociate from the texture model and a manual annotation of landmarks is necessary to perform the statistical analysis.

In this paper, we introduce the idea of using a two-stage separating hyper-plane, here called Statistical Discriminant Model (SDM), to interpret and reconstruct face images. Analogously to the Cootes et al. approaches [1 – 4], SDM requires a previous alignment of all the images to a common template to minimise variations that are not necessarily related to differences between the faces. However, instead of using landmarks or annotations on the images, SDM is based on the idea of using PCA to reduce the dimensionality of the original images and a maximum uncertainty linear classifier (MLDA) [8] to characterise the most discriminant differences between the samples of images.

The remainder of this paper is divided as follows. In section 2, we briefly review PCA and highlight its importance on reducing the high dimensionality of face images. Section 3 describes the standard linear discriminant analysis (LDA) and states the reasons for using a maximum uncertainty version of this approach to perform the face experiments required. The estimation of the separating hyper-plane and the implementation of the Statistical Discriminant Model are described in Section 4. In section 5, we present experimental results of the PCA and SDM approaches on a face database maintained by the Department of Electrical Engineering at FEI. This section includes reconstruction experiments of face images using the SDM approach proposed. In the last section, section 6, the paper concludes with a short summary of the findings of this study and future directions.

2. Principal Component Analysis (PCA)

PCA is a feature extraction procedure concerned with explaining the covariance structure of a set of variables through a small number of linear combinations of these variables. It is a well-known statistical technique that has been used in several image recognition problems, especially for dimensionality reduction. A comprehensive description of this multivariate statistical analysis method can be found in [6].

Let us consider the face recognition problem as an example to illustrate the main idea of the PCA. In any

image recognition, and particularly in face recognition, an input image with n pixels can be treated as a point in an n-dimensional space called the image space. The coordinates of this point represent the values of each pixel of the image and form a vector $x^T = [x_1, x_2, \ldots, x_n]$ obtained by concatenating the rows (or columns) of the image matrix. It is well-known that well-framed face images are highly redundant not only owing to the fact that the image intensities of adjacent pixels are often correlated but also because every individual has one mouth, one nose, two eyes, etc. As a consequence, an input image with n pixels can be projected onto a lower dimensional space without significant loss of information.

Let an N x n training set matrix X be composed of N input face images with n pixels. This means that each column of matrix X represents the values of a particular pixel observed all over the N images. Let this data matrix X have covariance matrix S with respectively Φ and Λ eigenvector and eigenvalue matrices, that is,

$$P^T S P = \Lambda . \quad \textbf{(1)}$$

It is a proven result that the set of m ($m \leq n$) eigenvectors of S, which corresponds to the m largest eigenvalues, minimises the mean square reconstruction error over all choices of m orthonormal basis vectors [6]. Such a set of eigenvectors that defines a new uncorrelated coordinate system for the training set matrix X is known as the principal components. In the context of face recognition, those P_{pca} components are frequently called eigenfaces [10].

Therefore, although n variables are required to reproduce the total variability (or information) of the sample X, much of this variability can be accounted for by a smaller number m of principal components. That is, the m principal components can then replace the initial n variables and the original data set, consisting of N measurements on n variables, is reduced to a data set consisting of N measurements on m principal components.

3. Maximum Uncertainty LDA (MLDA)

The primary purpose of the Linear Discriminant Analysis, or simply LDA, is to separate samples of distinct groups by maximising their between-class separability while minimising their within-class variability. Although LDA does not assume that the populations of the distinct groups are normally distributed, it assumes implicitly that the true covariance matrices of each class are equal because the same within-class scatter matrix is used for all the classes considered.

Let the between-class scatter matrix S_b be defined as

$$S_b = \sum_{i=1}^{g} N_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \quad \textbf{(2)}$$

and the within-class scatter matrix S_w be defined as

$$S_w = \sum_{i=1}^{g} (N_i - 1) S_i = \sum_{i=1}^{g} \sum_{j=1}^{N_i} (x_{i,j} - \bar{x}_i)(x_{i,j} - \bar{x}_i)^T \quad \textbf{(3)}$$

where $x_{i,j}$ is the n-dimensional pattern j from class π_i, N_i is the number of training patterns from class π_i, and g is the total number of classes or groups. The vector $\bar{x}_i$ and matrix S_i are respectively the unbiased sample mean and sample covariance matrix of class π_i [6]. The grand mean vector $\bar{x}$ is given by

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{g} N_i \bar{x}_i = \frac{1}{N} \sum_{i=1}^{g} \sum_{j=1}^{N_i} x_{i,j} , \quad \textbf{(4)}$$

where N is the total number of samples, that is, $N = N_1 + N_2 + \cdots + N_g$. It is important to note that the within-class scatter matrix S_w defined in equation (3) is essentially the standard pooled covariance matrix S_p multiplied by the scalar $(N - g)$, where S_p can be written as

$$S_p = \frac{1}{N-g} \sum_{i=1}^{g} (N_i - 1) S_i = \frac{(N_1 - 1)S_1 + (N_2 - 1)S_2 + \cdots + (N_g - 1)S_g}{N - g} . \quad \textbf{(5)}$$

The main objective of LDA is to find a projection matrix P_{lda} that maximizes the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix (Fisher's criterion), that is,

$$P_{lda} = \arg\max_{P} \frac{\left| P^T S_b P \right|}{\left| P^T S_w P \right|} . \quad \textbf{(6)}$$

The Fisher's criterion described in equation (6) is maximised when the projection matrix P_{lda} is composed of the eigenvectors of $S_w^{-1} S_b$ with at most $(g - 1)$ nonzero corresponding eigenvalues [5]. This is the standard LDA procedure.

However, the performance of the standard LDA can

be seriously degraded if there is only a limited number of total training observations N compared to the dimension of the feature space n. Since the within-class scatter matrix S_w is a function of $(N-g)$ or less linearly independent vectors, its rank is $(N-g)$ or less. Therefore, S_w is a singular matrix if N is less than $(n+g)$, or, analogously, might be unstable if N is not at least five to ten times $(n+g)$ [7].

To avoid the aforementioned critical issues of the standard LDA in limited sample and high dimensional problems, we have calculated P_{lda} by using a maximum uncertainty LDA-based approach (MLDA) that considers the issue of stabilising the S_w estimate with a multiple of the identity matrix [8, 9]. In a previous study [8] with application to the face recognition problem, Thomaz and Gillies showed that the MLDA approach improved the LDA classification performance with or without a PCA intermediate step and using less linear discriminant features [8].

The MLDA algorithm can be described as follows:

i. Find the Φ eigenvectors and Λ eigenvalues of S_p, where $S_p = S_w / [N-g]$;

ii. Calculate the S_p average eigenvalue $\overline{\lambda}$, that is,

$$\overline{\lambda} = \frac{1}{n}\sum_{j=1}^{n}\lambda_j = \frac{trace(S_p)}{n}; \qquad \textbf{(7a)}$$

iii. Form a new matrix of eigenvalues based on the following largest dispersion values

$$\Lambda^* = diag[\max(\lambda_1, \overline{\lambda}), ..., \max(\lambda_n, \overline{\lambda})]; \qquad \textbf{(7b)}$$

iv. Form the modified within-class scatter matrix

$$S_w^* = S_p^*(N-g) = (\Phi \Lambda^* \Phi^T)(N-g). \qquad \textbf{(7c)}$$

The maximum uncertainty LDA (MLDA) is constructed by replacing S_w with S_w^* in the Fisher's criterion formula described in equation (6). It is based on the idea [8] that in limited sample size and high dimensional problems where the within-class scatter matrix is singular or poorly estimated, the Fisher's linear basis found by minimising a more difficult but appropriate "inflated" within-class scatter matrix would also minimise a less reliable "shrivelled" within-class estimate.

4. Statistical Discriminant Model (SDM)

The Statistical Discriminant Model proposed in this work is essentially a two-stage PCA+MLDA linear classifier that reduces the dimensionality of the original images and extracts discriminant information from images.

In order to estimate the SDM separating hyperplane, we use training examples and their corresponding labels to construct the classifier. First a training set is selected and the average image vector of all the training images is calculated and subtracted from each n-dimensional vector. Then the training matrix composed of zero mean image vectors is used as input to compute the PCA transformation matrix. The columns of this n x m transformation matrix are eigenvectors, not necessarily in eigenvalues descending order. We have retained all the PCA eigenvectors with non-zero eigenvalues, that is, $m = N-1$, to reproduce the total variability of the samples with no loss of information. The zero mean image vectors are projected on the principal components and reduced to m-dimensional vectors representing the most expressive features of each one of the n-dimensional image vector. Afterwards, this N x m data matrix is used as input to calculate the MLDA discriminant eigenvector, as described in the previous section. Since in this work we have limited ourselves to two-group classification problems, there is only one MLDA discriminant eigenvector. The most discriminant feature of each one of the m-dimensional vectors is obtained by multiplying the N x m most expressive features matrix by the m x 1 MLDA linear discriminant eigenvector. Thus, the initial training set of face images consisting of N measurements on n variables, is reduced to a data set consisting of N measurements on only 1 most discriminant feature.

Once the two-stage SDM classifier has been constructed, we can move along its corresponding projection vector and extract the discriminant differences captured by the classifier. Any point on the discriminant feature space can be converted to its corresponding n-dimensional image vector by simply: (1) multiplying that particular point by the transpose of the corresponding linear discriminant vector previously computed; (2) multiplying its m most expressive features by the transpose of the principal components matrix; and (3) adding the average image calculated in the training stage to the n-dimensional image vector. Therefore, assuming that the spreads of the classes follow a Gaussian distribution and applying limits to the variance of each group, such as ± 2 sd, where sd is the standard deviation of each group, we can move along the SDM most discriminant features and map the results back into the image domain.

(a)

(b)

Figure 1. Samples of the female versus male (a) and non-smiling versus smiling training sets (b).

5. Experimental Results

We have used frontal images of a face database maintained by the Department of Electrical Engineering of FEI to carry out the experiments. This database contains a set of face images taken between June 2005 and March 2006 at the Artificial Intelligence Laboratory in São Bernardo do Campo, with 14 images for each of 118 individuals – a total of 1652 images*. All images are colourful and taken against a white homogenous background in an upright frontal position with profile rotation of up to about 180 degrees. Scale might vary about 10% and the original size of each image is 640x480 pixels.

To minimise image variations that are not necessarily related to differences between the faces, we aligned first all the frontal face images to a common template so that the pixel-wise features extracted from the images correspond roughly to the same location across all subjects. In this manual alignment, we have randomly chosen the frontal image of a subject as template and the directions of the eyes and nose as a location reference. For implementation convenience, all the frontal images were then cropped to the size of 64x64 pixels and converted to 8-bit grey scale.

We have carried the following two-group statistical analyses: female versus male experiments, and non-smiling versus smiling experiments. The idea of the first discriminant experiment is to evaluate the statistical approaches on a discriminant task where the differences between the groups are evident. In contrast, the second experiment, i.e. non-smiling versus smiling samples, poses an alternative analysis where there are subtle differences between the groups. Since the number of female images is limited and equal to 49, we have composed the female/male training set of 49 frontal female images and 49 frontal male images. For the smiling/non-smiling experiments, we have used the 49 frontal male images previously selected and their corresponding frontal smiling images. All faces are mainly represented by subjects between 19 and 30 years old with distinct appearance, hairstyle, and adorns. Figure 1 shows some examples of these two training sets selected.

5.1. PCA Results

In this section, we describe the most expressive features captured by PCA. As the average face image is an n-dimensional point ($n = 4096$) that retains all common features from the training sets, we could use this point to understand what happens statistically when we move along the principal components and reconstruct the respective coordinates on the image space. Analogously to the works by Cootes et al. [1 – 4], we have reconstructed the new average face images by changing each principal component separately using the limits of $\pm\sqrt{\lambda_i}$, where λ_i are the corresponding largest eigenvalues.

Figure 2 illustrates these transformations on the first three most expressive principal components using the female/male training set. As can be seen, the first principal component (on the top) captures essentially the variations in the illumination and gender of the training samples. The second principal component (middle), in turn, models variations related to the grey-level of the faces and hair, but it is not clear which specific variation this component is actually capturing. The last principal component considered, the third component (bottom), models mainly the size of the head of the training samples. It is important to note that as the female/male training set has a very clear separation be-

* All these images are available upon request (cet@fei.edu.br).

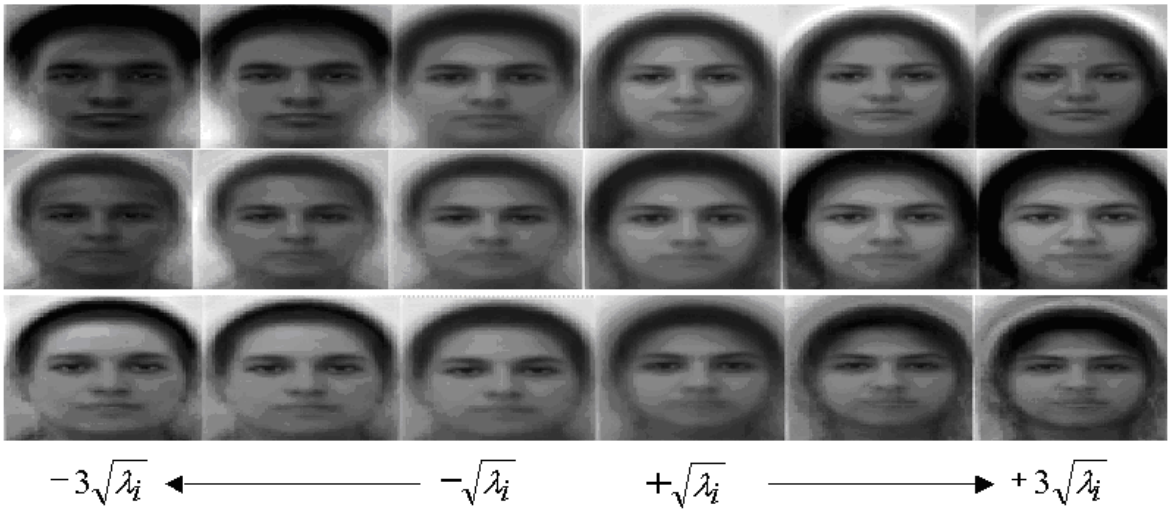


Figure 2. PCA results using the female/male training set.

tween the groups, the principal components have kept this separation and when we move along each principal component axis we can see this major difference between the samples, even though subtly, such as in the third principal component illustrated.

Figure 3 presents the three most expressive variations captured by PCA using the non-smiling/smiling training set, which is composed of male images only. Analogously to the female/male experiments, the first principal component (on the top) captures essentially the changes in illumination, the second principal component (middle) models variations particularly in the head shape, and the third component (bottom) captures variations in the facial expression among others.

As we should expect, these experimental results show that PCA captures features that have a considerable variation between all training samples, like changes in illumination, gender, and head shape. However, if we need to identify specific changes such as the variation in facial expression solely, PCA has not proved to be a useful solution for this problem. As can be seen in Figure 3, although the third principal component (bottom) models some facial expression variation, this specific variation has been captured by other principal components as well including other image artefacts. Likewise, as Figure 2 illustrates, although the first principal component (top) models gender variation, other changes have been modelled concurrently,

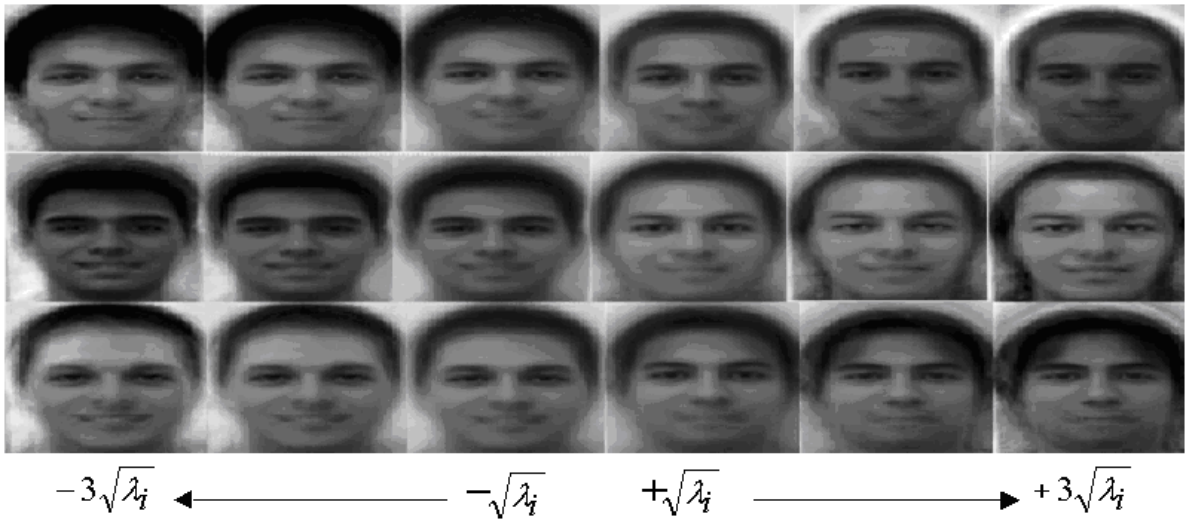


Figure 3. PCA results using the non-smiling/smiling training set (male images only).

-2sd -1sd Mean +1sd +2sd

Figure 4. SDM results using the female/male training set.

such as the variation in illumination. In fact, when we consider a whole grey-level model without landmarks to perform the PCA analysis, there is no guarantee that a single principal component will capture a specific variation alone, no matter how discriminant that variation might be.

5.2. SDM Results

As described earlier, in order to estimate the SDM separating hyperplane, we have used the female/male and non-smiling/smiling training sets previously selected and their corresponding labels to construct the classifier. Since in these experiments we have limited ourselves to two-group classification problems, there is only one SDM discriminant eigenvector. Therefore, assuming that the spreads of the classes follow a Gaussian distribution and applying limits to the variance of each group, such as ± 2 sd, where 'sd' is the standard deviation of each group, we can move along the SDM most discriminant features and map the results back into the image domain for visual analysis.

Figure 6 presents the SDM most discriminant features for the gender experiments. It displays the image regions captured by the SDM approach that change when we move from one side (left, male) of the dividing hyper-plane to the other (right, female), following limits to the standard deviation (± 2 sd) of each sample group. As can be seen, the SDM hyper-plane effectively extracts the group differences, showing clearly the features that mainly distinct the female samples from the male ones, such as the size of the eyebrows, nose and mouth, without enhancing other image artefacts.

Figure 5 shows the SDM most discriminant features for the facial expression experiments. Analogously to the gender experiments, Figure 5 displays the image regions captured by the SDM classifier that change when we move from one side (left, smiling) of the dividing hyper-plane to the other (right, non-smiling), following limits to the standard deviation (± 2 sd) of each sample group. As can be seen, the SDM hyper-plane effectively extracts the group differences, showing exactly what we should expect intuitively from a face image when someone changes their expression from smiling to non-smiling. In fact, it is possible to note that the SDM most discriminant direction has predicted a facial expression not necessarily present in our corresponding smiling/non-smiling training set, that is, the "definitely non-smiling" or may be "anger" status represented by the image +2sd in Figure 5.

Analogously to the PCA experiments, all SDM reconstructions have been made using the average face image of the corresponding training sets. However, it is possible to project any face image on the SDM feature space, move along its corresponding most discriminant features, and map the changes back to the original image space. Figure 6 shows these experimental results when we move an example image along the male/female (Figure 6a) and smiling/non-smiling (Figure 6b) hyper-planes previously calculated. As can be seen in Figure 6a, the most discriminant features be-

-2sd -1sd Mean +1sd +2sd

Figure 5. SDM results using the smiling/non-smiling training set.

(a)

(b)

Figure 6. SDM results when we move an example image along the male/female (a) and smiling/non-smiling (b) hyper-planes.

tween a male and female face images have been incorporated on the example image when we move it to the male side of the dividing hyper-plane, such as the thickening of the lips, nose, and eyebrows. In contrast, since the example chosen is from a woman, almost no facial changes occurs when we move the same example to the other side of the hyper-plane, that is, to the female side. Also, according to Figure 6b, it is possible to see that the SDM linear classifier has incorporated all the most discriminant facial changes that we intuitively expect when we change our facial expression from smiling to non-smiling status. It is important to note in this case where most of the facial changes are localised around the mouth that only the differences related to the facial expression differences have changed on the image with no impact on other face features, such as hair-style, forehead, eyebrows, and chin.

6. Conclusion

In this work, we introduced the idea of using the PCA+MLDA two-stage linear classifier to interpret and reconstruct frontal face images rather than recognising subjects. Differently from other statistical approaches, our method is based on a supervised separation between the whole images and not on the use of landmarks and isolated models for the shapes' and textures' variations. The experiments carried out in this work showed that subjective information such as beauty and happiness can be efficiently captured by a linear classifier when we pre-process the face images using a simple affine transformation. The results presented in this paper suggested that the statistical discriminant model proposed could be useful to reconstruct not only frontal face images but also face images with different profiles. Further work is being undertaken to investigate this possibility.

Acknowledgments

The authors would like to thank Leo Leonel de Oliveira Junior for acquiring and normalizing the FEI database under the grant FEI-PBIC 32-05.

References

[1] T. F. Cootes, G. J. Edwards, C. J. Taylor, "Active Appearance Models", H.Burkhardt and B. Newmann editors, in Proceedings of ECCV'98, vol. 2, pp. 484-498, 1998.

[2] T.F Cootes, A. Lanitis, "Statistical Models of Appearance for Computer Vision", Technical report, University of Manchester, 125 pages, 2004.

[3] T. F. Cootes, C. J. Taylor, D. H. Cooper, J. Graham, "Active Shape Models- Their Training and Application", *Computer Vision and Image Understanding*, vol. 61, no.1, pp. 38-59, 1995.

[4] T.F. Cootes, K.N. Walker, C.J. Taylor, "View-Based Active Appearance Models", In 4th International Conference on Automatic Face and Gesture Recognition, Grenoble, France, pp. 227-232, 2000.

[5] P.A. Devijver and J. Kittler, *Pattern Classification: A Statistical Approach*. Prentice-Hall, 1982.

[6] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, second edition. Boston: Academic Press, 1990.

[7] A. K. Jain and B. Chandrasekaran, "Dimensionality and Sample Size Considerations in Pattern Recognition Practice", *Handbook of Statistics*, 2, pp. 835-855, 1982.

[8] C. E. Thomaz and D. F. Gillies, "A Maximum Uncertainty LDA-based approach for Limited Sample Size problems - with application to Face Recognition", in Proceedings of SIBGRAPI'05, IEEE CS Press, pp. 89-96, 2005.

[9] C. E. Thomaz, D. F. Gillies and R. Q. Feitosa, “A New Covariance Estimate for Bayesian Classifiers in Biometric Recognition”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 2, pp. 214-223, 2004.

[10] M. Turk, A. Pentland, Eigenfaces for Recognition, *Journal of Cognitive Neuroscience*, MIT, vol. 73, pp. 71-86, 1991.