

CENTRO UNIVERSITÁRIO DA FEI

ELIANA ZACCHI TENÓRIO QUIROGA

**ANÁLISE MULTILINEAR DISCRIMINANTE DE FORMAS FRONTAIS DE  
IMAGENS 2D DE FACE**

São Bernardo do Campo

2012

ELIANA ZACCHI TENÓRIO QUIROGA

**ANÁLISE MULTILINEAR DISCRIMINANTE DE FORMAS FRONTAIS DE  
IMAGENS 2D DE FACE**

Dissertação de Mestrado apresentada ao Centro  
Universitário da FEI, como parte dos requisitos  
necessários para obtenção do título de Mestre  
em Engenharia Elétrica, orientada pelo professor  
Dr. Carlos Eduardo Thomaz.

São Bernardo do Campo

2012

Quiroga, Eliana Zacchi Tenório

Análise multilinear discriminante de formas frontais de imagens 2D de face / Eliana Zacchi Tenório Quiroga - São Bernardo do Campo, 2012.

82 f.:il.

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

Orientador Prof. Dr. Carlos Eduardo Thomaz.

1. Análise de componentes principais
2. Hiperplanos discriminantes
3. Formas 2D de face.

CDU: 681.32



Centro Universitário da **FEI**

# APRESENTAÇÃO DE DISSERTAÇÃO ATA DA BANCA JULGADORA

PGE- 10

## Programa de Mestrado de Engenharia Elétrica

Aluno: Eliana Zacchi Tenório Quiroga

Matrícula: 1081090

Título do Trabalho: **ANÁLISE MULTILINEAR DISCRIMINANTE DE FORMAS FRONTAIS DE IMAGENS 2D DE FACE.**

Área de Concentração: Inteligência Artificial Aplicada à Automação

Orientador: Prof. Dr. Carlos Eduardo Thomaz

Data da realização da defesa: 10 / fevereiro / 2012

**ORIGINAL ASSINADA**

A Banca Julgadora abaixo-assinada atribuiu ao candidato o seguinte:

APROVADO

REPROVADO

São Bernardo do Campo, 10 / fevereiro / 2012.

### MEMBROS DA BANCA JULGADORA

Prof. Dr. Carlos Eduardo Thomaz

Ass.: \_\_\_\_\_

Prof. Dr. José Carlos de Souza Júnior

Ass.: \_\_\_\_\_

Prof. Dr. Yossi Zana

Ass.: \_\_\_\_\_

### **VERSÃO FINAL DA DISSERTAÇÃO**

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

\_\_\_\_\_

Aprovação do Coordenador do Programa de Pós-graduação

\_\_\_\_\_  
Prof. Dr. Carlos Eduardo Thomaz

Aos meus pais e ao meu marido.

## **AGRADECIMENTOS**

Eu gostaria de agradecer, primeiramente a Deus, por me dar condições de concluir mais esta etapa em minha vida acadêmica.

Gostaria de agradecer também aos meus pais que me deram todo o apoio e incentivo necessário para não desistir nunca. Aos meus parentes e amigos por entender o porquê não estive presente em tantas reuniões, aniversários, passeios e por continuarem ao meu lado mesmo após tantas ausências de minha parte. Ao meu marido por ter me apoiado, ter estado ao meu lado mesmo quando eu estava estudando e não estava ao lado dele e por ter “automatizado” muitas das rotinas utilizadas durante os experimentos.

Ao meu orientador, Dr. Carlos Eduardo Thomaz, agradeço os conselhos, aprendizado, paciência e principalmente os incentivos. A todos os professores do curso de mestrado do Centro Universitário da FEI, agradeço pelo aprendizado e dedicação.

*“Knowledge will forever govern ignorance;  
and people who mean to be their own governors,  
must arm themselves with the power which knowledge gives.”*

James Madison

## RESUMO

Muitas atividades humanas e interações com o meio ambiente realizadas por máquinas (computadores ou robôs) são guiadas por análises da informação visual disponível. Na maioria das situações, formas 2D em imagens têm as mesmas características dos objetos originais. Esta dissertação descreve e implementa uma análise multilinear de formas 2D em imagens frontais de face com o objetivo de entender como esses dados se alteram no espaço original dos dados. São investigados não somente as componentes principais que representam maior variância, mas também as direções que melhor representam as diferenças discriminantes de acordo com a característica de interesse como, por exemplo, alterações relativas a gênero ou expressão facial. Os experimentos, utilizando bases de imagens de faces disponíveis publicamente para pesquisa, mostram que a análise multilinear discriminante baseada em hiperplanos paramétricos e não-paramétricos e componentes principais reordenadas tende a priorizar as direções que melhor explicam as diferenças de interesse. Portanto, acredita-se que pode-se estender todo e qualquer trabalho de análise e interpretação visual baseado em forma por essas novas transformações lineares de dados.

Palavras chave: Análise de componentes principais, Hiperplanos discriminantes, Formas 2D de face.



## **ABSTRACT**

Several human activities and interactions with the environment carried out by machines (computers or robots) are guided by analysis of the visual information available. In most situations, 2D shapes of images have the same characteristics of the original objects. This thesis describes and implements a multilinear shape analysis of 2D frontal face images. Our aim is not only to visualize and understand the data that represents the highest sample variance, but also that extract the most discriminant differences related to a specific characteristic of interest, such as changes owing to gender or facial expression. Our experiments, using face images databases public available for reserch, show that the multilinear discriminant analysis proposed based on parametric and nonparametric hyperplanes and reordered principal components tend to prioritize main directions that best explain the diferences of interest. Therefore, we believe that any shape analysis and visual interpretation can be extended using these new linear data transformation.

**Keywords:** Principal component analysis, Discriminant hyperplanes, 2D face shapes.

## LISTA DE FIGURAS

Figura 2-1 – Algumas formas 2D de fácil reconhecimento. Adaptado de [1].....	17
Figura 2-2 – Modos de variação da posição da mão. Adaptado de [4]. .....	19
Figura 3-1 - Pontos de controle segundo Maja (adaptado de [22]). .....	21
Figura 3-2 – Pontos de controle segundo Tim Cootes (adaptado de [20])......	22
Figura 3-3 - Gráfico que representa as 2 primeiras componentes principais e o autovalor correspondente a cada uma delas. Esta figura foi adaptada de [17]......	25
Figura 3-4 – Gráfico que representa o vetor $y$ que é a projeção mais discriminante do LDA e perpendicular ao vetor de separação dos dois grupos maximizando a distância entre classes e minimizando a distância intraclasses. Esta figura é uma adaptação de [29]. .....	28
Figura 3-5 – Hiperplano de separação com margem pequena em (a) e com margem máxima em (b) para um SVM treinado com duas classes. É esperada uma melhor capacidade de generalização de (b). Adaptado de [34].....	31
Figura 4-1 – Base AR. Exemplo das fotos tiradas, mostrando as características de cada uma das 13 posições tiradas em 2 sessões diferentes para cada pessoa (adaptado de [39]).....	37
Figura 4-2 - Localização dos 22 pontos de controle previamente marcados na AR_Face Database.....	38
Figura 4-3 - Base FEI. Exemplo das fotos tiradas, mostrando as características de cada uma das 14 posições. ....	40
Figura 4-4 – Imagem com os 46 pontos definidos no software am_tools [19]. .....	41
Figura 4-5 – Detalhamento das bases AR e FEI.....	43
Figura 4-6 - Reconstruções visuais variando-se as três primeiras componentes principais para a base AR.....	45
Figura 4-7 - Reconstruções visuais variando-se as três primeiras componentes principais para a base FEI. ....	46
Figura 4-8 – Taxa de acerto do PCA para a base AR.....	47
Figura 4-9 – Taxa de acerto do PCA para a base FEI. ....	48
Figura 4-10 – Taxa de classificação (%) e desvio padrão entre parênteses (%) para características gênero (homem versus mulher) e expressão facial (não sorrindo versus sorrindo) para a base AR. ....	48
Figura 4-11 – Taxas de classificação (%) e desvio padrão entre parênteses (%) para características gênero (homem versus mulher) e expressão facial (não sorrindo versus sorrindo) para a base FEI.....	49

Figura 4-12 - Reconstruções visuais navegando-se no hiperplano discriminante do LDA, MLDA e SVM (extrapolando os vetores suportes) e com relação ao gênero (de homens para mulheres) para a base AR. ....	49
Figura 4-13 - Reconstruções visuais navegando-se no hiperplano discriminante do LDA, MLDA e SVM (extrapolando os vetores suportes) e com relação à expressão facial (de sorrindo para não sorrindo) para a base AR. ....	50
Figura 4-14 - Reconstruções visuais navegando-se no hiperplano discriminante do LDA, MLDA e SVM (extrapolando os vetores suportes) com relação ao gênero (de homens para mulheres) para a base FEI. ....	50
Figura 4-15 - Reconstruções visuais navegando-se no hiperplano discriminante do LDA, MLDA e SVM (extrapolando os vetores suportes) e com relação à expressão facial (de sorrindo para não sorrindo) para a base FEI. ....	51
Figura 4-16 – As 20 primeiras componentes principais ordenadas pelos hiperplanos do LDA, MLDA e SVM usando a base AR. ....	52
Figura 4-17 – As 20 primeiras componentes principais ordenadas pelos hiperplanos do LDA, MLDA e SVM usando a base FEI. ....	53
Figura 4-18 - Variância total explicada pela componente principal do PCA padrão e pelas componentes principais mais discriminantes ordenadas pelo LDA, MLDA e SVM. Experimentos de gênero (superior) e expressão facial (inferior) usando a base AR. ....	54
Figura 4-19 – Variância total explicada pela componente principal do PCA padrão e pelas componentes principais mais discriminantes ordenadas pelo LDA, MLDA e SVM. Experimentos de gênero (superior) e expressão facial (inferior) usando a base FEI. ....	55
Figura 4-20 - Taxa de acerto do DPCA variando o número de componentes principais para gênero e expressão facial da base AR. ....	56
Figura 4-21 - Taxa de acerto do DPCA variando o número de componentes principais para gênero e expressão facial da base FEI. ....	57
Figura 4-22 - Visualização das mudanças descritas pelas direções principais usando a base AR. De cima para baixo: as 2 primeiras componentes principais do PCA padrão; 2 primeiras componentes principais mais discriminantes selecionadas pelo LDA sendo homens à esquerda e mulheres à direita; e as 2 primeiras componentes principais mais discriminantes pelo LDA sendo sorrindo à esquerda e não sorrindo à direita. ....	58
Figura 4-23 - Visualização das mudanças descritas pelas direções principais usando a base AR. De cima para baixo: as 2 primeiras componentes principais do PCA padrão; 2 primeiras componentes principais mais discriminantes selecionadas pelo MLDA sendo homens à	

esquerda e mulheres à direita; e as 2 primeiras componentes principais mais discriminantes pelo MLDA sendo sorrindo à esquerda e não sorrindo à direita. ....	59
Figura 4-24 - Visualização das mudanças descritas pelas direções principais usando a base AR. De cima para baixo: as 2 primeiras componentes principais do PCA padrão; 2 primeiras componentes principais mais discriminantes selecionadas pelo SVM sendo homens à esquerda e mulheres à direita; e as 2 primeiras componentes principais mais discriminantes pelo SVM sendo sorrindo à esquerda e não sorrindo à direita. ....	60
Figura 4-25 – Visualização das mudanças descritas pelas direções principais usando a base FEI. De cima para baixo: as 2 primeiras componentes principais do PCA padrão; 2 primeiras componentes principais mais discriminantes selecionadas pelo LDA sendo homens à esquerda e mulheres à direita; e as 2 primeiras componentes principais mais discriminantes pelo LDA sendo sorrindo à esquerda e não sorrindo à direita. ....	61
Figura 4-26 - Visualização das mudanças descritas pelas direções principais usando a base FEI. De cima para baixo: as 2 primeiras componentes principais do PCA padrão; 2 primeiras componentes principais mais discriminantes selecionadas pelo MLDA sendo homens à esquerda e mulheres à direita; e as 2 primeiras componentes principais mais discriminantes pelo MLDA sendo sorrindo à esquerda e não sorrindo à direita. ....	62
Figura 4-27 - Visualização das mudanças descritas pelas direções principais usando a base FEI. De cima para baixo: as 2 primeiras componentes principais do PCA padrão; 2 primeiras componentes principais mais discriminantes selecionadas pelo SVM sendo homens à esquerda e mulheres à direita; e as 2 primeiras componentes principais mais discriminantes pelo SVM sendo sorrindo à esquerda e não sorrindo à direita. ....	63
Figura 6-1 – Exemplo de rotação de face (adaptado de [38]). ....	65
Figura 6-2 - Reconstruções visuais variando-se as nove primeiras componentes principais para as imagens de perfil da base FEI. ....	66
Figura 6-3 – Reconstruções visuais variando-se as nove primeiras componentes principais discriminantes conforme MLDA para as imagens de perfil da base FEI. ....	67

## SUMÁRIO

1	INTRODUÇÃO .....	14
1.1	Organização .....	15
2	REVISÃO BIBLIOGRÁFICA.....	16
3	MÉTODOS.....	21
3.1	Modelo de Distribuição de Pontos (PDM) .....	21
3.2	Análise de Componentes Principais .....	23
3.3	Hiperplanos Discriminantes.....	26
3.3.1	Análise de Discriminantes Lineares (LDA / MLDA) .....	26
3.3.2	Máquina de Vetores de Suporte.....	29
3.4	Componentes Principais Discriminantes .....	31
3.5	Considerações complementares .....	34
4	EXPERIMENTOS E RESULTADOS .....	35
4.1	Bases de faces e representação das formas .....	35
4.1.1	Base de Faces AR .....	35
4.1.2	Banco de Faces da FEI .....	39
4.2	Experimentos .....	43
4.3	Resultados: Componentes Principais.....	45
4.4	Resultados: Discriminantes Lineares .....	48
4.5	Resultados: Componentes Discriminantes .....	52
5	CONCLUSÕES.....	64
6	TRABALHOS FUTUROS.....	65
	REFERÊNCIAS .....	68
	APÊNDICE A – Código do DPCA .....	72
	APÊNDICE B - Artigo do X SBAI.....	76

## 1 INTRODUÇÃO

O estudo sobre representação e interpretação de informações visuais em um sistema computacional é formalmente conhecido como Visão Computacional, tendo motivado aplicações em diversos campos do conhecimento como, robótica, medicina, segurança, entre outros.

Muitas atividades humanas e interações com o meio ambiente realizadas por máquinas (computadores ou robôs) são guiadas por análises da informação visual disponível [1]. Humanos detectam e interpretam faces e expressões faciais em uma cena com pouco ou nenhum esforço, desenvolver um sistema automatizado que realize esta tarefa é um tanto complicado. Existem vários problemas relacionados, pois além dos recursos computacionais necessários para interpretar automaticamente as imagens obtidas, há dificuldades adicionais na análise da informação visual, como a existência de ruído, oclusão, distorção e perda de informação em mapeamentos de objetos do espaço tridimensional (3D) para espaços bidimensionais (2D).

Na maioria das situações, formas 2D em imagens têm as mesmas características dos objetos originais com exceção fundamentalmente da profundidade e do movimento. Objetos representados por formas 2D podem ser prontamente reconhecidos na maioria dos casos. Por exemplo, a forma de um coração, de uma estrela, de uma nota musical, de uma ferramenta e até mesmo o formato da face de uma pessoa, assunto desta dissertação, são facilmente reconhecidos por nós.

Essencialmente, a análise computacional de forma envolve uma série de tarefas importantes desde a aquisição da imagem até a classificação da mesma. Estas tarefas podem ser divididas em três grandes etapas: pré-processamento, extração de características e classificação [1]. Esta dissertação descreve uma análise multilinear de formas 2D em imagens frontais de face com o objetivo de entender como esses dados estão se alterando no espaço original. Especificamente, investiga-se não somente as componentes principais com maior variância [2], [3], [4], [5], mas também as direções que melhor representam as diferenças discriminantes entre os dados dependendo do contexto da análise multivariada em questão [6], [7]. Pretende-se, portanto, extrair as direções mais discriminantes de acordo com as características de interesse, por exemplo, diferenças de gênero e expressão facial, e não as

direções mais expressivas [8] que não necessariamente representam exclusivamente as diferenças discriminantes [9].

Em outras palavras, o objetivo deste trabalho é estender o trabalho de Thomaz e Giraldi [7] para forma de faces, distinguir as diferenças entre formas de face de homens e mulheres (gênero) e formas de face entre pessoas sorrindo e não sorrindo (expressão facial).

## **1.1 Organização**

Esta dissertação está organizada da seguinte forma. No próximo capítulo, apresenta-se uma breve revisão bibliográfica dos principais trabalhos já publicados relacionados com a área de análise de imagens a partir de formas e que utilizaram abordagens lineares. No capítulo 3, descrevem-se as técnicas de estatística multivariada que serão utilizadas ao longo do trabalho. Depois, no capítulo 4, são discutidos os experimentos e resultados obtidos. Finalmente, no capítulo 5, são apresentadas as conclusões e no capítulo 6, possibilidades de trabalhos futuros.

## 2 REVISÃO BIBLIOGRÁFICA

A análise de imagens de faces é uma linha de pesquisa em Visão Computacional que tem motivado muitos pesquisadores dada a sua enorme abrangência e complexidade. Abrangência pela sua vasta aplicabilidade, principalmente em tarefas de reconhecimento 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 [10].

Visão requer um processamento real de muitas informações como forma, orientação espacial, cor, textura, movimento, entre outras. Existem dificuldades adicionais para o reconhecimento visual como, por exemplo, iluminação, oclusão e distorção [1]. Para evitar estas dificuldades, muitos pesquisadores optam por trabalhar com informação sobre as formas dos objetos ao invés das imagens das mesmas.

A forma de um objeto é uma descrição geométrica do espaço ocupado pelo objeto, conforme determinado pelo seu contorno externo. Formas simples podem ser descritas pela geometria dos objetos como um conjunto de dois ou mais pontos, uma linha, uma curva, uma figura plana ou uma figura sólida. Em outras palavras, a forma de um conjunto de pontos é toda a informação geométrica que é invariante a translações, rotações e mudanças de tamanho. A forma não depende do tamanho do objeto nem das mudanças na orientação ou direção. Mas as formas podem mudar se o objeto for dimensionado não uniformemente. Por exemplo, uma esfera pode se tornar uma elipse quando escalado diferentemente nas direções vertical e horizontal. Portanto, preservando-se os eixos de simetria, preserva-se também a forma.

Os principais modelos de Visão Computacional baseados em formas de imagens 2D, conforme ilustrados na Figura 2-1, foram propostos a partir do final da década de 1980. Em 1987, Kass, Witkin e Terzopoulos [11] descreveram modelos ativos de contornos, ou “snakes” flexíveis, que podem se esticar e deformar para se adequar às características das imagens. Logo depois, em 1989, Yuille [12] descreveu um modelo de contorno das mãos a partir de várias figuras geométricas que representam as características da imagem. Ele descreveu também métodos que ajustam seus modelos às imagens. No entanto, segundo Tim



Cootes [2], um dos pesquisadores pioneiros em trabalhar com formas como padrões multidimensionais, esta técnica tem que ser individualmente adaptada a cada aplicação.

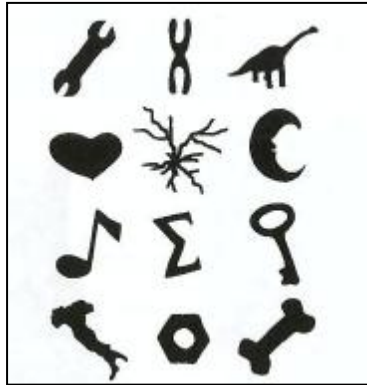


Figura 2-1 – Algumas formas 2D de fácil reconhecimento. Adaptado de [1].

Ainda em 1989, Staib e Duncan [13] desenvolveram um modelo de formas baseado em Fourier, representando uma fronteira fechada como uma soma de funções trigonométricas de várias frequências. Eles também desenvolveram uma técnica de minimização de energia iterativa para adequar o modelo à imagem. No entanto, este modelo é limitado a fronteiras fechadas e nem sempre funciona corretamente para formas que variam.

Após um ano, em 1990, David Cooper e C. Taylor [14] descreveram um método de representação de formas baseado num modelo de distribuição estatístico do comprimento entre dois pontos de controle, chamado *Chord Length Distribution* (CLD), o qual possui muitas das propriedades frequentemente citadas para a representação da forma. Também, no início da década de 1990, Lowe [15] descreveu uma técnica para adequar projeções de modelos tridimensionais parametrizados para imagens bidimensionais por iteratividade minimizando a distância entre linhas no modelo projetado e a imagem.

Em 1991, Turk e Pentland [16] apresentaram um trabalho pioneiro onde a idéia principal foi executar o reconhecimento de imagens de faces em um espaço de menor dimensionalidade, porém, para isso, o reconhecimento foi baseado no princípio da menor distância Euclidiana entre uma face de teste, projetada no espaço de faces, e cada face do conjunto de treinamento. Eles apresentaram, também, resultados do uso da técnica de Análise de Componentes Principais (PCA) [17], investigada neste trabalho, para localização de faces

em imagens de seqüência de vídeos e mostraram como resultado deste trabalho que o PCA necessita que o fundo da imagem seja estático, podendo ser afetado por variações de iluminação, pose e principalmente escala.

Em 1991, T. Cootes e colegas ([2], [3], [4], [5]), interessados em problemas onde os formatos dos objetos variam, por exemplo, interpretações médicas onde existem variações biológicas, desenvolveram um método novo de representação de formas por meio do qual pode-se, automaticamente, extrair as propriedades das formas dos dados e descrever sucintamente os parâmetros que representam maior variância desses dados [3]. Os objetos são representados por um conjunto de pontos de controle e examinando-se a variação estatística das coordenadas desses pontos para um conjunto de treinamento estima-se a posição inicial, o tamanho e o formato de um objeto em uma imagem e, conseqüentemente, os ajustes necessários para melhor adequar este modelo de forma à imagem correspondente [3].

Dando continuidade aos resultados impressionantes obtidos nos primeiros trabalhos, em 1992, Cootes e Taylor [3] desenvolveram um método de construção de modelos flexíveis, representando os objetos como um conjunto de pontos e examinando a estatística das coordenadas após vários conjuntos de treinamento. Esses modelos são conhecidos hoje como *Point Distribution Models* (PDM). Na verdade, os autores descreveram em [3] um esquema de otimização iterativo para PDMs, permitindo uma estimativa inicial de posição, tamanho e formato de um objeto em uma imagem a ser analisada e, com isso, parâmetros de ajustes podem ser calculados para melhor adequar o modelo à imagem. Os modelos sugeridos são calculados para cada ponto do modelo. Diferentemente do modelo de Yuille [12], proposto em 1989, um PDM pode ser utilizado em diferentes aplicações, pois o mesmo representa uma grande variedade de objetos.

Para avaliar modelos de reconhecimento de faces baseados em medidas geométricas e intensidade dos pixels das imagens, em 1993, Brunelli e Poggio [18] desenvolveram e implementaram dois algoritmos que comparam duas estratégias de análise simples para uma base de dados comum. Em ambos os algoritmos a base de dados utilizada continha imagens frontais de face de 47 pessoas sendo 26 homens e 21 mulheres e 4 imagens por pessoa. O primeiro algoritmo é baseado no cálculo de um conjunto de características geométricas, como largura e comprimento do nariz, boca, posição e forma do queixo. E o segundo é baseado no modelo de nível de cinza correspondente.

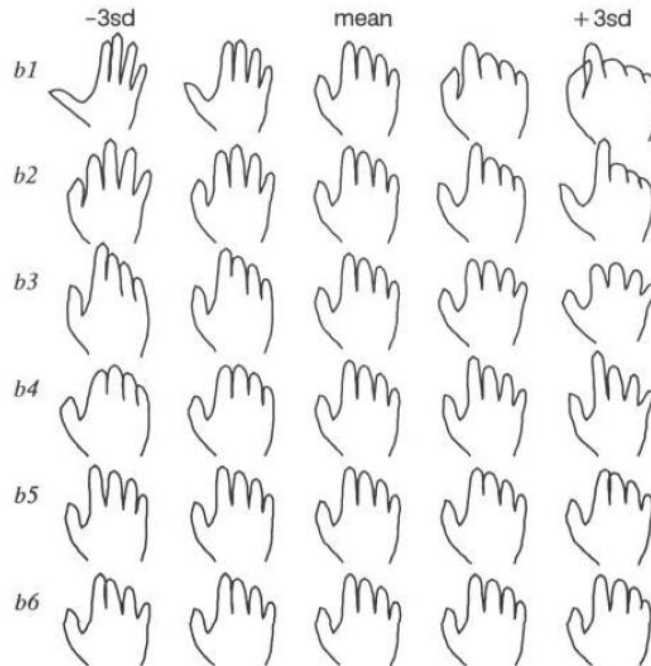


Figura 2-2 – Modos de variação da posição da mão. Adaptado de [4].

Em 1995, Tim Cootes e Taylor descrevem, formalmente pela primeira vez, no artigo “*Active Shape Models*” [3], a idéia inicial do software conhecido atualmente como *am\_Tools* [19] utilizando como exemplo imagem de um resistor e de uma mão. Em [4], os mesmos autores descreveram um procedimento para reconhecer gestos e monitorar em uma seqüência de imagens o movimento de uma mão humana quando opera-se um *mouse*. O *Point Distribution Models* (PDM) foi utilizado para prover uma descrição parametrizada compacta do formato da mão para qualquer gesto ou transição entre os mesmos. Os valores dos parâmetros dos formatos foram usados num classificador estatístico para identificar gestos. Para isto, o PDM foi gerado através da realização de alinhamentos quadrados dos membros do conjunto de treinamento e, a partir disto, foram definidos 89 pontos de controle para representar a mão. Após a aplicação do método PCA [17], o resultado foi um vetor que representava a forma média de uma mão e uma série de direções no espaço original que descrevia as variações das mãos em torno desta forma média, conforme apresentado na Figura 2-2. Esta técnica apresentou dificuldades para encontrar um modelo específico em seqüência de imagens, porém funcionou perfeitamente para modelos estatísticos [4]. Considerando que o presente trabalho é fundamentado em modelos estatísticos, optou-se por usar a mesma técnica

citada acima, que obteve um bom funcionamento para os modelos estatísticos que serão discutidos no próximo capítulo.

Depois, em 1998, Tim Cootes [5] colocou em prática a idéia dos pontos de controle utilizada na imagem de mão, porém, desta vez, ele aplicou em um modelo de face estatístico com 122 pontos de controle, aplicados manualmente em cada imagem de face. Para gerar um modelo estatístico com as variações da face, ele alinhou todas as imagens do conjunto de treinamento e aplicou o PCA para gerar uma imagem média. Mais recentemente, em 2005, Lanitis, Taylor e Cootes [20] descreveram o uso de modelos flexíveis para representar o formato e a aparência em nível de cinza do rosto humano. Esses modelos são controlados por um pequeno número de parâmetros que podem ser usados para codificar a aparência de um rosto para compreender as imagens e até mesmo classificá-las. Esse método utilizou apenas um modelo flexível para classificação de face e apresentou uma alta taxa de reconhecimento, o que permitiu considerá-lo como um sistema de identificação de faces automático o qual tolera mudanças de expressão, inclinação e iluminação.

Nos últimos anos, em 2009, Thomaz et. al. [6] desenvolveram um método linear discriminante que quantifica as mudanças em imagens 2D de face. Eles localizaram a direção mais discriminante de acordo com a característica de interesse, entre elas, gênero, expressão facial e idade [6]. Para isto, foi utilizado um método multivariado estatístico que é composto pelo PCA e hiperplanos separadores, reduzindo a dimensionalidade das imagens originais e extraindo as informações discriminantes das mesmas [6]. Devido à limitação supervisionada dos hiperplanos separadores, os autores propuseram uma análise multilinear de imagens limitada à direção mais discriminante da característica de interesse. Estendendo a idéia do trabalho descrito em [6], em 2010, Thomaz e Giraldi propuseram um novo método de análise multilinear de imagens que reordena as componentes principais de acordo com as características que mais discriminam. Sendo assim, pode-se agora então encontrar as direções discriminantes para cada característica de interesse sem a restrição da abordagem supervisionada.

Os trabalhos aqui revisados brevemente descrevem modelos flexíveis ou modelos baseados em forma e mostram que quase todos os estudos que utilizam modelos lineares, principalmente os que descrevem trabalhos com faces, utilizam métodos estatísticos multivariados de extração de características.

### 3 MÉTODOS

Este capítulo descreve os principais métodos de extração de características investigados neste trabalho. Esses métodos se restringem a análises multivariadas dos padrões de interesse que, neste estudo, se referem a formas frontais de imagens de faces.

#### 3.1 Modelo de Distribuição de Pontos (PDM)

O Modelo de Distribuição de Pontos (*Point Distribution Models*) foi desenvolvido por Tim Cootes [21] e representa a geometria média da forma de um determinado padrão de interesse e a variação geométrica do mesmo inferida a partir de um conjunto de treinamento. O PDM trabalha com pontos de referência. Por exemplo, um determinado ponto, como o canto direito da boca no rosto de uma pessoa, é referenciado em todas as imagens do conjunto de treinamento no mesmo lugar.

Segundo Maja [22], os pontos mais relevantes de características faciais são os cantos e pontos médios dos olhos, cantos das sobrancelhas, cantos e pontos médios dos lábios, ponta e cantos externos do nariz e ponto inferior central do queixo. Estes 20 pontos estão representados na Figura 3-1.

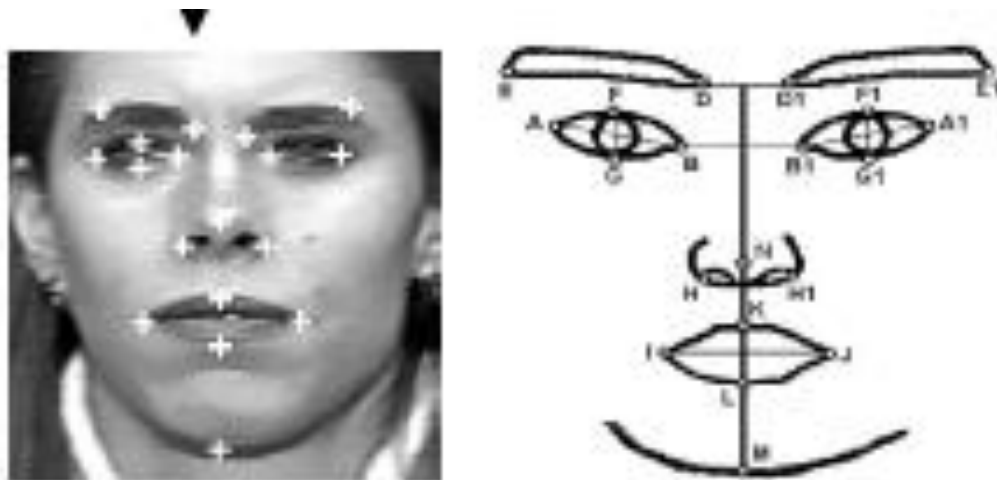


Figura 3-1 - Pontos de controle segundo Maja (adaptado de [22]).

Tim Cootes [20] sugere a utilização de 152 pontos, conforme Figura 3-2, para representar o formato da face.

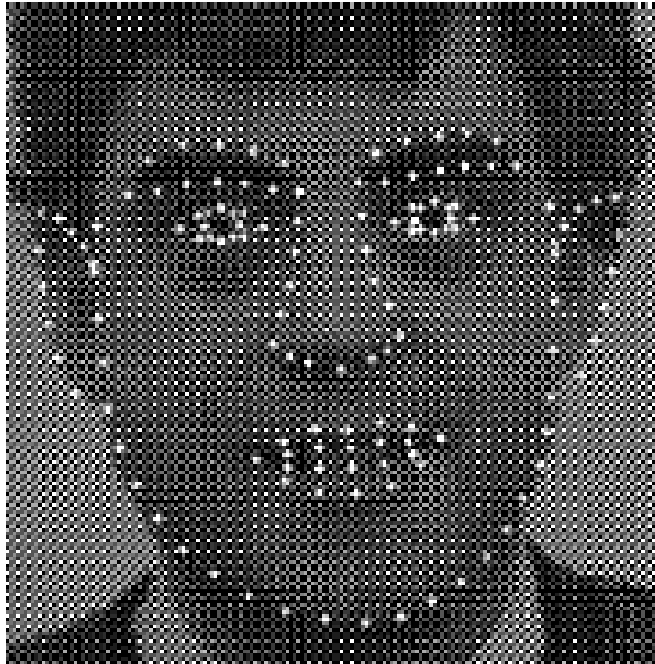


Figura 3-2 – Pontos de controle segundo Tim Cootes (adaptado de [20]).

No PDM, são definidos pontos de referência para cada imagem, como os exemplos ilustrados na Figura 3-1 e na Figura 3-2 e esses pontos de referência são compostos por coordenadas  $x$  e  $y$ . Ou seja, ao invés de trabalhar com a intensidade dos pixels este modelo permite analisar a forma descrita em pontos referenciados como coordenadas  $x$  e  $y$  da imagem original 2D.

Admitindo que cada coordenada de cada ponto de controle seja uma variável na imagem, cada forma pode ser representada no espaço  $n$ -dimensional a partir da matriz de dados, a qual é denominada  $X_{N \times n}$ , onde  $N$  é o número total de imagens disponíveis e  $n$  a dimensionalidade de cada forma descrita pelas coordenadas  $(x,y)$  de cada um dos  $n/2$  pontos de controle. Essa matriz  $X$  de representação de dados 2D pode ser, então, descrita por:

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,n/2} & Y_{1,1} & Y_{1,2} & \dots & Y_{1,n/2} \\ X_{2,1} & X_{2,2} & \dots & X_{2,n/2} & Y_{2,1} & Y_{2,2} & \dots & Y_{2,n/2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{N-1,1} & X_{N-1,2} & \dots & X_{N-1,n/2} & Y_{N-1,1} & Y_{N-1,2} & \dots & Y_{N-1,n/2} \\ X_{N,1} & X_{N,2} & \dots & X_{N,n/2} & Y_{N,1} & Y_{N,2} & \dots & Y_{N,n/2} \end{bmatrix},$$

onde  $x_{ij}$  e  $y_{ij}$  representam, respectivamente, as coordenadas  $x$  e  $y$  do ponto de controle  $j$  para a forma  $i$ .

Sendo assim, a dimensionalidade da matriz de formas 2D varia de acordo com o número de pontos de referência de cada forma.

### 3.2 Análise de Componentes Principais

A Análise de Componentes Principais, ou simplesmente *PCA* (*Principal Component Analysis*) [17], é um método estatístico multivariado de extração de informação, proposto por Karl Pearson [23] em 1901, que transforma um conjunto  $n$  de variáveis originais em um número  $m$  de combinações lineares dessas variáveis, onde  $m \leq n$ , por meio da seleção das características que representam maior variância [24] e conseqüente redução de dimensionalidade dos dados.

Seja uma amostra de treinamento contendo  $N$  formas de face representadas por vetores  $n$ -dimensionais que descrevem as coordenadas 2D dos pontos de controle dessas formas. A matriz de covariância desses dados pode ser escrita como:

$$S = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T, \quad (3.1)$$

onde:  $x_i$  é a forma onde todas as coordenadas de cada ponto de controle foram concatenadas e formaram um vetor  $n$ -dimensional e  $\bar{x}$  é a forma média global da amostra de treinamento. Pode-se perceber que a matriz de covariância  $S$  é uma função de  $(N-1)$  ou menos vetores linearmente independentes e tem dimensionalidade  $n \times n$ . Os autovalores de  $S$  devem satisfazer a equação (3.2):

$$|S - \lambda I| = 0, \quad (3.2)$$

onde  $I$  é a matriz identidade de dimensionalidade  $n \times n$ . Esta equação é conhecida como Equação Característica e qualquer valor de  $\lambda$  que satisfaça esta equação é chamado de autovalor e o vetor correspondente a este  $\lambda$  é chamado de autovetor [17].

Como  $S$  é uma matriz  $n \times n$  simétrica, tem-se os  $n$  autovalores reais  $\lambda_1, \dots, \lambda_n$  e os  $n$  autovetores reais  $\phi_1, \dots, \phi_n$ . Os autovetores correspondentes a 2 autovalores distintos são ortogonais entre si. A prova deste teorema pode ser encontrada em [17].

As componentes principais são os autovetores e autovalores da matriz de covariância  $S$  reordenados de forma decrescente, ou seja, onde o primeiro autovetor corresponde ao autovalor de maior valor.

É um resultado provado na literatura que o conjunto  $m$  de autovetores de  $S$ , que corresponde aos maiores autovalores, minimiza o erro quadrático médio de reconstrução desses dados para todas as possíveis escolhas de bases  $m$  vetoriais ortonormais [17]. Esse conjunto de autovetores que define um novo sistema de coordenadas para a amostra de treinamento é conhecido como componentes principais e pode ser descrito matematicamente como:

$$P^T S P = \Lambda. \quad (3.3)$$

onde  $P$  e  $\Lambda$  são as matrizes de autovetores e autovalores de  $S$ .

No contexto específico de reconhecimento de faces, a matriz  $P_{pca} = [p_1, p_2, \dots, p_m]$  de componentes principais ordenada decrescentemente ( $\lambda_1 \geq \lambda_2 \dots \geq \lambda_m$ ) é comumente denominada de autofaces [25], [16]. As  $m$  componentes principais podem então substituir as  $n$  variáveis e a amostra original de dados, que consiste em  $N$  exemplos de  $n$  variáveis, é reduzida para uma amostra de dados contendo  $N$  exemplos e  $m$  componentes principais.

De forma análoga aos trabalhos descritos em [4], [5] e [20], pode-se utilizar a forma média global  $\bar{x}$  representada pelo vetor  $n$ -dimensional para interpretar as alterações mais expressivas capturadas pelas componentes principais. Matematicamente, essa navegação nas componentes principais pode ser descrita por:

$$y = \bar{x} + j\sqrt{\lambda_i} p_i, \quad (3.4)$$

onde  $j \in \{-3, -2, -1, 0, 1, 2, 3\}$  e  $p_i$  são as  $m$  componentes principais ordenadas decrescentemente pelos seus autovalores correspondentes, isto é,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ .



A idéia geométrica desta navegação pode ser observada na Figura 3-3, onde a elipse representa o espalhamento dos dados descrito pela matriz de covariância e os eixos da elipse são as componentes principais.

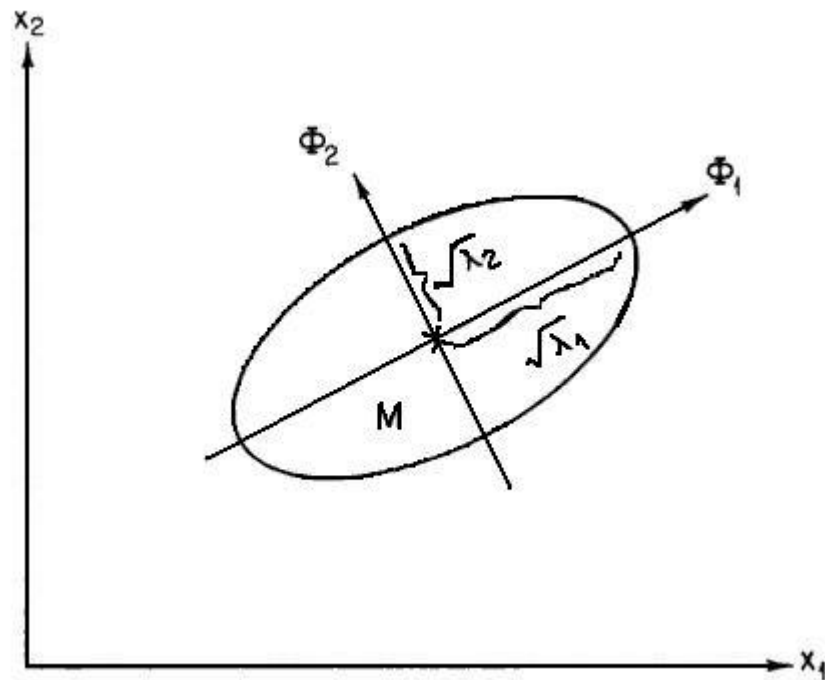


Figura 3-3 - Gráfico que representa as 2 primeiras componentes principais e o autovalor correspondente a cada uma delas. Esta figura foi adaptada de [17].

A proporção da variância total explicada pela  $k$  componente principal é :

$$(\text{proporção da variância total}) = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad k = 1, 2, \dots, p. \quad (3.5)$$

Devido à característica não-supervisionada de generalização linear do PCA, não é possível saber qual componente principal irá distinguir um padrão específico ou de interesse. Na próxima seção, são descritas outras transformações lineares que permitem levar em consideração essa informação.

### 3.3 Hiperplanos Discriminantes

Os classificadores discriminantes utilizam técnicas de classificação e discriminação com o objetivo de separar grupos de objetos (ou padrões) distintos ou alocar novos padrões em grupos previamente definidos.

#### 3.3.1 Análise de Discriminantes Lineares (LDA / MLDA)

A Análise de Discriminantes Lineares, ou simplesmente, LDA (*Linear Discriminant Analysis*), é um método estatístico multivariado de extração de informação, proposto por Ronald A. Fisher [26] em 1936, que separa, através de combinação linear, em duas ou mais classes de interesse, os dados originais. Essa separação em classes ou grupos distintos é feita por meio de uma transformação linear que maximiza a distância entre as classes e minimiza o espalhamento dentro de cada classe.

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

$$S_b = \sum_{j=1}^g N_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})^T \quad (3.6)$$

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

$$S_w = \sum_{i=1}^g \sum_{j=1}^{N_i} (x_{i,j} - \bar{x}_i)(x_{i,j} - \bar{x}_i)^T \quad (3.7)$$

onde:  $x_{i,j}$  é a forma n-dimensional j da classe  $\pi_i$ ,  $N_i$  é o número de exemplos de treinamento da classe  $\pi_i$ , e  $g$  representa o número total de classes ou grupos. O vetor  $\bar{x}_i$  é a média amostral da classe  $\pi_i$  [17]. O vetor da média global  $\bar{x}$  pode ser então reescrito como

$$\bar{x} = \frac{1}{N} \sum_{i=1}^g N_i \bar{x}_i = \frac{1}{N} \sum_{i=1}^g \sum_{j=1}^{N_i} x_{i,j} \quad (3.8)$$

onde  $N$  é, como descrito anteriormente, o número total de exemplos de treinamento, ou seja,  $N = N_1 + N_2 + \dots + N_g$ .

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

$$W_{lda} = \arg_w \max \frac{|W^T S_b W|}{|W^T S_w W|}. \quad (3.9)$$

O critério de Fisher descrito pela equação (3.9) é satisfeito quando a matriz de projeção  $W_{lda}$  é composta, no máximo, pelos  $(g-1)$  autovetores de  $S_w^{-1} S_b$ , cujos autovalores correspondentes são não-nulos [17] [27]. Por exemplo, em problemas envolvendo duas classes somente, a matriz de projeção do LDA contém, na verdade, um único autovetor  $w_{lda}$ , admitindo que  $S_w$  seja não-singular.

A idéia geométrica desta navegação pode ser observada na Figura 3-4, onde as elipses representam os espalhamentos dos dados de cada grupo, que são considerados iguais e calculados por  $S_w$ , e o vetor  $w$  é a projeção mais discriminante considerando a separação dos dois grupos.

A equação (3.10) expressa como se navega ao longo da projeção do LDA,

$$y = \bar{x} + w_{LDA} b_i. \quad (3.10)$$

onde  $\bar{x}$  é a forma média global,  $w_{LDA}$  é o hiperplano discriminante e  $b_i$  é o parâmetro que será variado para que se percorra o eixo da respectiva projeção do LDA, de maneira análoga à navegação nas componentes principais.

A projeção  $y$  do LDA na Figura 3-4 demonstra que existe uma determinada distribuição com média e variância. Com o intuito de percorrer o eixo de projeção do LDA, o parâmetro  $b_i$  pode, iniciando pela média, variar 3 desvios padrões para cada lado, considerando espalhamentos com distribuição normal [7].

O LDA tem sido utilizado como uma técnica de extração de características em diversos problemas de classificação. Porém, ao aplicar o LDA em casos com poucas amostras, tem-se dois possíveis problemas: singularidade e instabilidade da matriz intra-classes

[28]. Na prática, isto pode ser observado comumente em problemas de reconhecimento de padrões em imagens onde existe um grande número de pixels por imagem, mas o número total de imagens (conjunto de treinamento) é limitado.

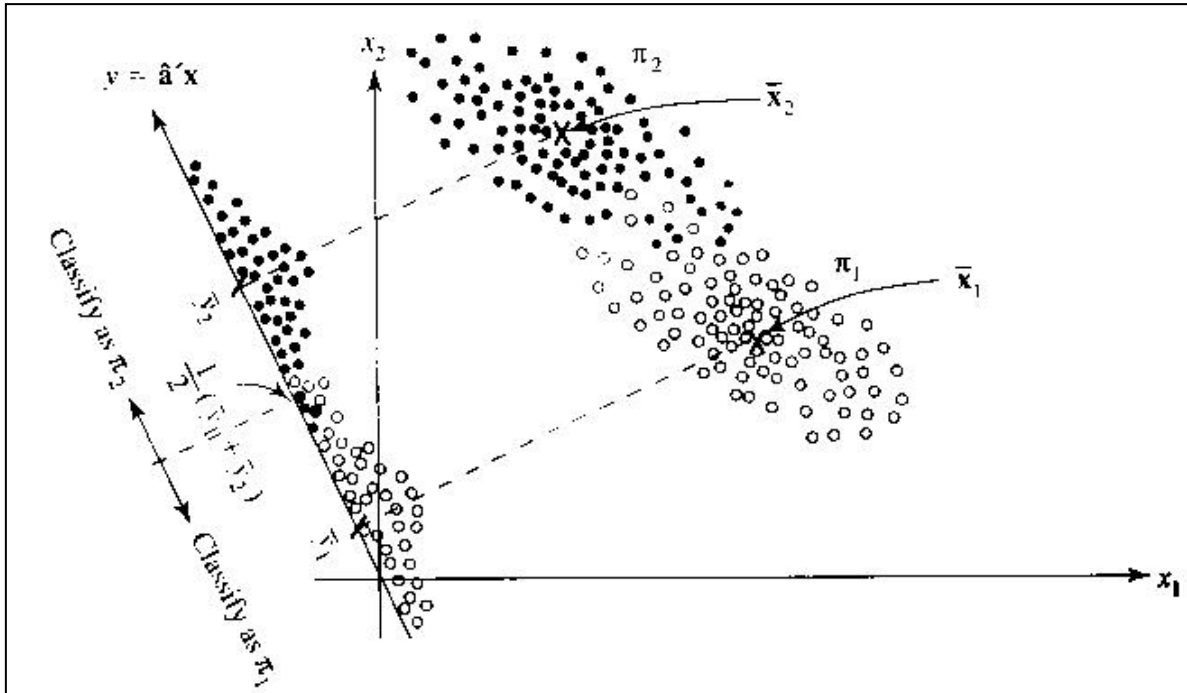


Figura 3-4 – Gráfico que representa o vetor  $y$  que é a projeção mais discriminante do LDA e perpendicular ao vetor de separação dos dois grupos maximizando a distância entre classes e minimizando a distância intraclasses. Esta figura é uma adaptação de [29].

Portanto, quando o número total de exemplos  $N$  não é consideravelmente maior que o número de variáveis  $n$ , ocorre uma instabilidade no cálculo da matriz inversa de  $S_w$  [17]. A quantidade de exemplos de treinamento para evitar essa instabilidade no cálculo da matriz inversa de  $S_w$  deve ser igual ou superior a 5 vezes a quantidade de variáveis que definem essas amostras [30]. Uma alternativa para o tratamento do problema de instabilidade no cálculo da inversa da matriz  $S_w$  é usar o método denominado MLDA (*Maximum uncertainty Linear Discriminant Analysis*) [28]. Esta técnica consiste em substituir a matriz de espalhamento  $S_w$  por outra matriz regularizada  $S_w^*$ . A nova matriz regularizada  $S_w^*$  pode ser calculada por meio dos seguintes passos:

- a) Seleciona-se os autovetores  $\phi$  e autovalores  $\lambda$  de  $S_p$ , onde:

$$S_p = \frac{S_w}{N-g}; \quad (3.11)$$

b) Calcula-se a média dos autovalores  $\bar{\lambda}$ , ou seja,

$$\bar{\lambda} = \frac{1}{n} \sum_{j=1}^n \lambda_j = \frac{\text{Tr}(S_p)}{n}; \quad (3.12)$$

c) Constrói-se uma nova matriz de autovalores baseada na dispersão dos maiores autovalores, isto é,

$$\Lambda^* = \text{diag}[\max(\lambda_1, \bar{\lambda}), \dots, \max(\lambda_n, \bar{\lambda})]; \quad (3.13)$$

d) Calcula-se a matriz de espalhamento intraclasses regularizada

$$S_w^* = S_p^* (N - g) = (\Phi \Lambda^* \Phi^T)(N - g). \quad (3.14)$$

Com a matriz  $S_w^*$  calculada, substitui-se  $S_w$  da equação (3.9) por  $S_w^*$  e regulariza-se o critério de Fisher para problemas onde o número total de exemplos de treinamento é limitado e equivalente à dimensionalidade dos dados.

### 3.3.2 Máquina de Vetores de Suporte

A Máquina de Vetores de Suporte ou *Support Vector Machines* (SVM) é uma técnica de aprendizagem desenvolvida por V. Vapnik e sua equipe na AT&T [31], [32], [33]. Esta técnica foi originalmente criada para classificação binária, mas pode ser estendida para problemas de múltiplas classes.

O SVM é um caso particular de métodos baseado em kernel, que tem como objetivo mapear vetores de características em um espaço de alta dimensão usando uma função kernel e, a partir daí, criar um novo espaço linear onde se alinha um hiperplano de separação de padrões. Os vetores de características que estão localizados mais próximos do hiperplano são chamados de vetores de suporte.

O classificador SVM também é chamado de classificador não-paramétrico, pois ao contrário da inferência estatística clássica, os parâmetros de distribuição de probabilidade não

são pré-definidos e seu número de vetores de suporte depende dos dados de treinamento utilizados.

A idéia do método SVM pode ser resumida da seguinte forma. Seja um conjunto de dados composto por dados de duas classes. O SVM determina o hiperplano que separa os pontos de forma a colocar o maior número de pontos da mesma classe do mesmo lado, maximizando a distância entre as classes e a imagem de separação.

Para uma classificação linear binária, seja o seguinte conjunto de treinamento formado por  $N$  pontos, onde cada entrada  $x_i$  tem  $n$  atributos e está em uma das duas classes, ou seja,  $y_i = -1$  ou  $y_i = +1$ . O hiperplano de separação pode ser descrito pela equação (3.15):

$$y_i = x_i * w + b = 0 \quad (3.15)$$

onde  $w$  é o vetor normal ao hiperplano e  $\frac{b}{w}$  é a distância perpendicular do hiperplano até a origem.

Os dados de treinamento podem ser discriminados em:

$$y_i = +1 : x_i * w + b \geq +1 , \quad (3.16)$$

$$y_i = -1 : x_i * w + b \leq -1 . \quad (3.17)$$

Considerando os pontos mais próximos da separação do hiperplano, os vetores de suporte, pode-se descrever os planos como:

$$\text{para } H_1: x_i * w + b = +1, \quad (3.18)$$

$$\text{para } H_2: x_i * w + b = -1. \quad (3.19)$$

Para obter a separação ótima de hiperplano é necessário maximizar a margem entre  $H_1$  e  $H_2$ , ou seja, manter os vetores de suporte mais distantes possíveis do hiperplano discriminante.

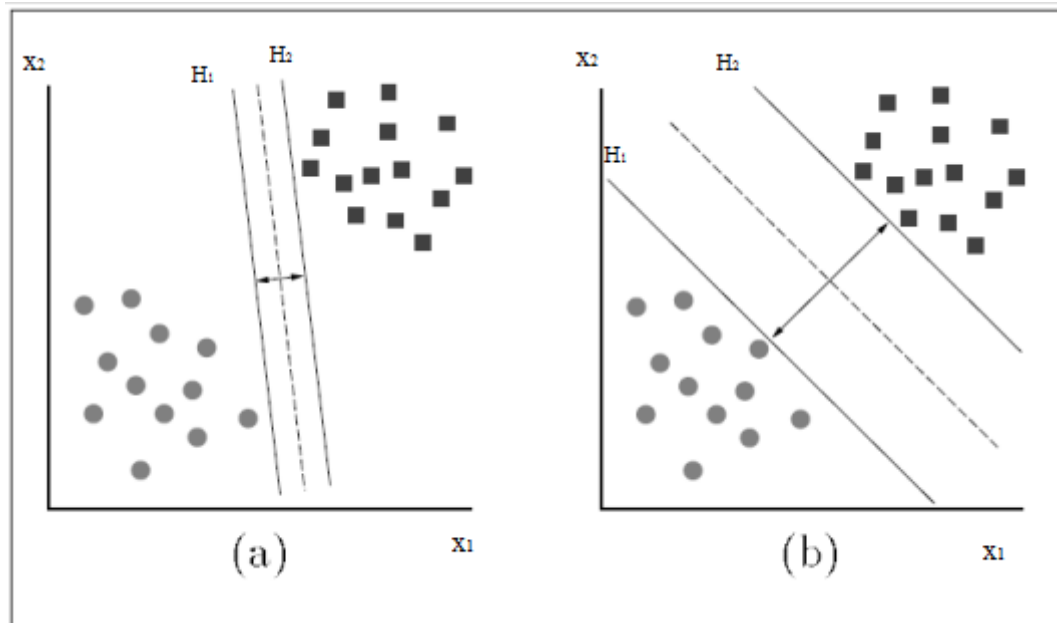


Figura 3-5 – Hiperplano de separação com margem pequena em (a) e com margem máxima em (b) para um SVM treinado com duas classes. É esperada uma melhor capacidade de generalização de (b). Adaptado de [34].

O vetor  $w_{svm}$  é uma combinação dos vetores de suporte, ou seja, aqueles pontos de treinamento cujo  $\alpha \neq 0$ :

$$w_{svm} = \sum_{i=1}^N \alpha_i y_i x_i. \quad (3.20)$$

Os métodos LDA / MLDA e o SVM estão limitados ao número de grupos. Sendo assim, tendo dois grupos para classificar, embora utilize-se informação supervisionada, tem-se apenas uma direção para interpretar as alterações mais discriminantes capturadas [35].

### 3.4 Componentes Principais Discriminantes

Thomaz e Giraldo [7] propuseram a idéia de usar os pesos discriminantes obtidos por meio de hiperplanos de separação para selecionar as componentes principais mais discriminantes, ou seja, a seleção será feita pela contribuição discriminante de cada uma delas. Este método é conhecido como *Discriminant Principal Components Analysis* (DPCA) e pode ser aplicado tanto em hiperplanos de separação paramétricos como não paramétricos.

O número de direções discriminantes não é limitado ao número de grupos, isso se deve ao fato de trabalhar no espaço do PCA, porém reorientado com relação a outros hiperplanos, como o LDA, MLDA e SVM. Conforme mencionado anteriormente, tanto o LDA, MLDA como o SVM são capazes de identificar apenas uma direção discriminante para separação de dois grupos somente e, conseqüentemente, outras informações importantes para caracterizar diferenças entre os grupos podem ser perdidas [36].

Sejam  $N$  exemplos de treinamento linearmente independentes que descrevem duas classes de padrões. Os vetores de dados com média zero são projetados nas componentes principais e reduzidos a vetores  $m$  dimensionais, onde  $m=N-1$ , que representam as características mais expressivas de cada um dos vetores de dados  $n$ -dimensional. Após isso, a matriz de dados  $N \times m$  e seus correspondentes são usados como dados de entrada para calcular os hiperplanos de separação.

A característica mais discriminante de cada um dos vetores  $m$ -dimensionais é obtida através da multiplicação dos  $N \times m$  características mais expressivas da matriz por  $m \times 1$  vetores discriminantes. Assim, o conjunto de treinamento inicial composto por  $N$  medições em  $n$  atributos é reduzido ao conjunto de dados composto por  $N$  medições com apenas 1 característica mais discriminante dado por:

$$\begin{aligned}\tilde{y}_1 &= x_{11}w_1 + x_{12}w_2 + \dots + x_{1m}w_m, \\ \tilde{y}_2 &= x_{21}w_1 + x_{22}w_2 + \dots + x_{2m}w_m, \\ &\dots \\ \tilde{y}_N &= x_{N1}w_1 + x_{N2}w_2 + \dots + x_{Nm}w_m,\end{aligned}\tag{3.21}$$

onde  $[w_1, w_2, \dots, w_m]$  são os pesos correspondentes às características das componentes principais calculadas pelos hiperplanos de separação e  $[x_{i1}, x_{i2}, \dots, x_{im}]$  são atributos de cada vetor de dados  $i$  projetado no espaço do PCA, onde  $i=1, \dots, N$ .

Os pesos  $w_T = [w_1, w_2, \dots, w_m]$  determinam a contribuição discriminante de cada característica. Esses pesos descrevem o hiperplano de separação no espaço original de características porque:

$$\begin{aligned}w_1 &= (P_{pca}w)^T * p_1, \\ w_2 &= (P_{pca}w)^T * p_2, \\ &\dots \\ w_m &= (P_{pca}w)^T * p_m,\end{aligned}\tag{3.22}$$



onde  $p_1, p_2, \dots, p_m$  são as componentes principais. Os pesos que são zero ou próximo de zero têm contribuição mínima no valor de  $\tilde{y}_i$  descrito na equação (3.21) e indicam que a característica a ele correspondente é insignificante para a separação do grupo de amostras. Por outro lado, os pesos com alto valor (em módulo) indicam que a característica a ele correspondente é muito importante para caracterizar a diferença entre os grupos.

Sendo assim, ao invés de selecionar as componentes principais em ordem decrescente de autovalores, selecionam-se as primeiras componentes principais com os maiores pesos discriminantes, ou seja:

$$P_{lda/svm} = [p_1, p_2, \dots, p_m] = \arg_p \max |P^T S P|, \quad (3.23)$$

onde  $p_i$  são os autovetores de  $S$  correspondentes aos pesos discriminantes em ordem decrescente, isto é,  $|w_1| > |w_2| > \dots > |w_m|$ . Esses pesos podem ser descritos pelo hiperplano de separação do LDA:

$$w_{lda} = \arg_p \max \frac{|w^T P_{pca}^T S_b P_{pca} w|}{|w^T P_{pca}^T S_w P_{pca} w|}, \quad (3.24)$$

ou pelo hiperplano de separação do MLDA:

$$w_{mlda} = \arg_p \max \frac{|w^T P_{pca}^T S_b P_{pca} w|}{|w^T P_{pca}^T S_w^* P_{pca} w|}, \quad (3.25)$$

ou pelo hiperplano de separação do SVM:

$$w_{svm} = \sum_{i=1}^N \alpha_i y_i (x_i P_{pca}). \quad (3.26)$$

Selecionam-se as componentes principais que são eficientes para discriminar os grupos ao invés de representar todos os padrões.

Esta técnica tem todas as características do PCA, como, por exemplo, base de transformação ortonormal, porém, orientada com alguma informação à priori extraída da base de treinamento classificada. Portanto, pode-se estender todo e qualquer trabalho de classificação baseado em forma, como os que foram descritos na revisão bibliográfica, para essa nova transformação linear de dados.

### 3.5 Considerações complementares

Neste capítulo foram descritos os métodos PCA, LDA, MLDA e SVM. Chamou-se a atenção das limitações de cada transformação linear tradicionais na literatura. Apresentou-se também um modelo proposto recentemente denominado DPCA que tenta combinar as vantagens de um método não supervisionado como o PCA com os métodos de hiperplano de separação paramétricos e não paramétricos como os que foram apresentados. Na próxima seção, descrevem-se os principais experimentos e resultados obtidos neste estudo utilizando bases públicas de imagens de faces a fim de fazer uma comparação dos métodos aqui descritos.

## 4 EXPERIMENTOS E RESULTADOS

Os experimentos foram desenvolvidos com o objetivo de distinguir faces femininas de masculinas, de pessoas sorrindo e não sorrindo, a partir de pontos que representam a forma das mesmas, sem levar em consideração cor ou textura. Para isto, foram utilizadas bases de imagens de faces disponíveis publicamente para pesquisa.

Todos os experimentos foram realizados utilizando programas desenvolvidos em Matlab. Para marcar os pontos e automatizar a representação de formas características foi utilizado o software *am\_tools* [19] desenvolvido e disponibilizado por Tim Cootes. Este software gera um arquivo de coordenada para cada imagem com as coordenadas x,y dos pontos que foram marcados nas mesmas.

### 4.1 Bases de faces e representação das formas

As bases de faces utilizadas neste trabalho são públicas e descrevem comportamentos distintos de imagens de face. São elas: AR [37] e FEI [38].

Dentro do escopo deste trabalho, estas bases foram escolhidas sabendo-se que a base FEI é uma base normalizada e a base AR possui algum ruído. Essa diferença será analisada neste trabalho.

#### 4.1.1 Base de Faces AR

A base de dados AR [37] foi criada por Aleix Martinez e Robert Benavente no Centro de Visão Computacional (CVC) da Universidade Autônoma de Barcelona (UAB). Há, nesta base, 4000 imagens coloridas que correspondem a 126 faces (sendo 70 de homens e 56 de mulheres).

As fotos foram tiradas com diferentes expressões faciais, condições de iluminação e oclusão (óculos escuros e lenço). As fotos foram realizadas no CVC sob condições controladas. Nenhuma restrição de roupa, óculos, maquiagem, penteado ou outra foi imposta aos participantes. Cada pessoa participou de 2 sessões de fotos, com intervalo de 14 dias entre elas.

As imagens foram armazenadas como arquivos RGB RAW de dimensão 768x576 pixels e 24 bits de profundidade. Os arquivos foram nomeados de forma a facilitar a identificação das imagens: A-xx-yy.raw, sendo A = W (mulher) ou M (homem), xx o número de identificação da pessoa (00 a 70 para homens e 00 a 56 para mulheres) e yy a característica de cada imagem conforme abaixo:

- 1 – Expressão natural;
  - 2- Sorrindo;
  - 3- Com raiva;
  - 4- Gritando;
  - 5- Com iluminação projetada à esquerda;
  - 6- Com iluminação projetada à direita;
  - 7- Com iluminação projetada em ambos os lados;
  - 8- Com óculos escuro;
  - 9- Com óculos escuro e iluminação projetada à esquerda;
  - 10- Com óculos escuro e iluminação projetada à direita;
  - 11- Com lenço;
  - 12- Com lenço e iluminação projetada à esquerda;
  - 13- Com lenço e iluminação projetada à direita;
  - 14 a 26 – Segunda sessão. Seguindo as mesmas condições das imagens 1 a 13.
- Esta seqüência de nomeação dos arquivos pode ser observada na Figura 4-1.

Neste trabalho, foram utilizadas apenas as fotos com características expressão natural e sorrindo.

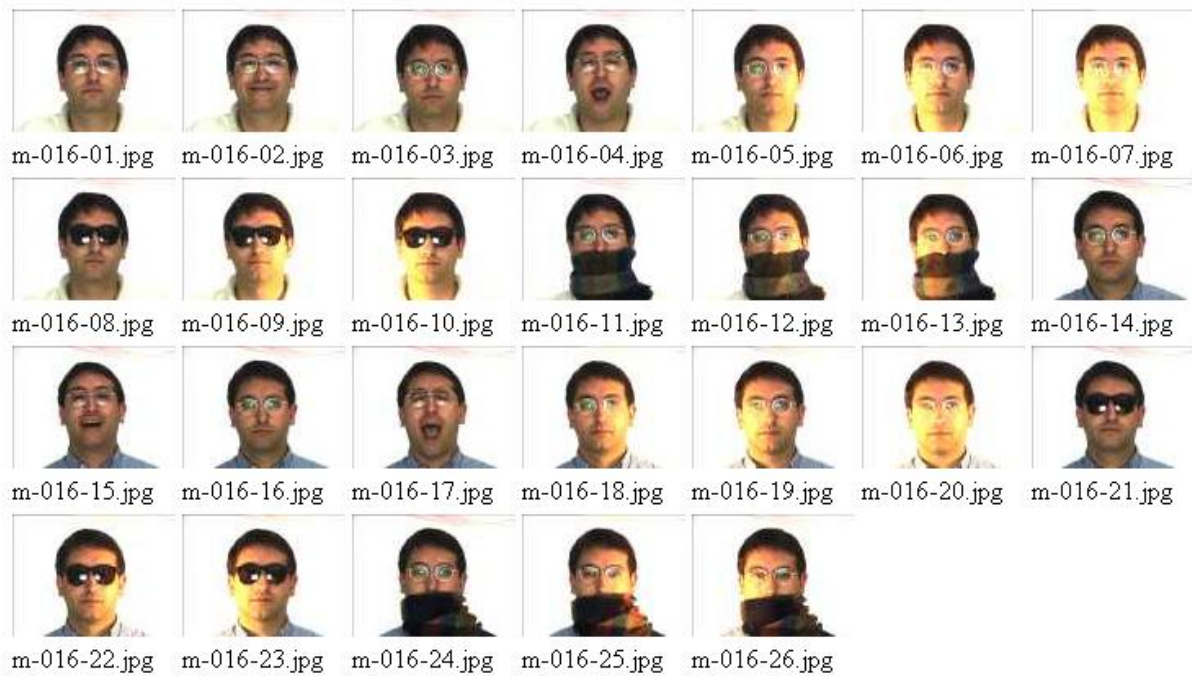


Figura 4-1 – Base AR. Exemplo das fotos tiradas, mostrando as características de cada uma das 13 posições tiradas em 2 sessões diferentes para cada pessoa (adaptado de [39]).

Todas essas imagens foram previamente marcadas com 22 pontos de controle, conforme ilustrado na Figura 4-2.

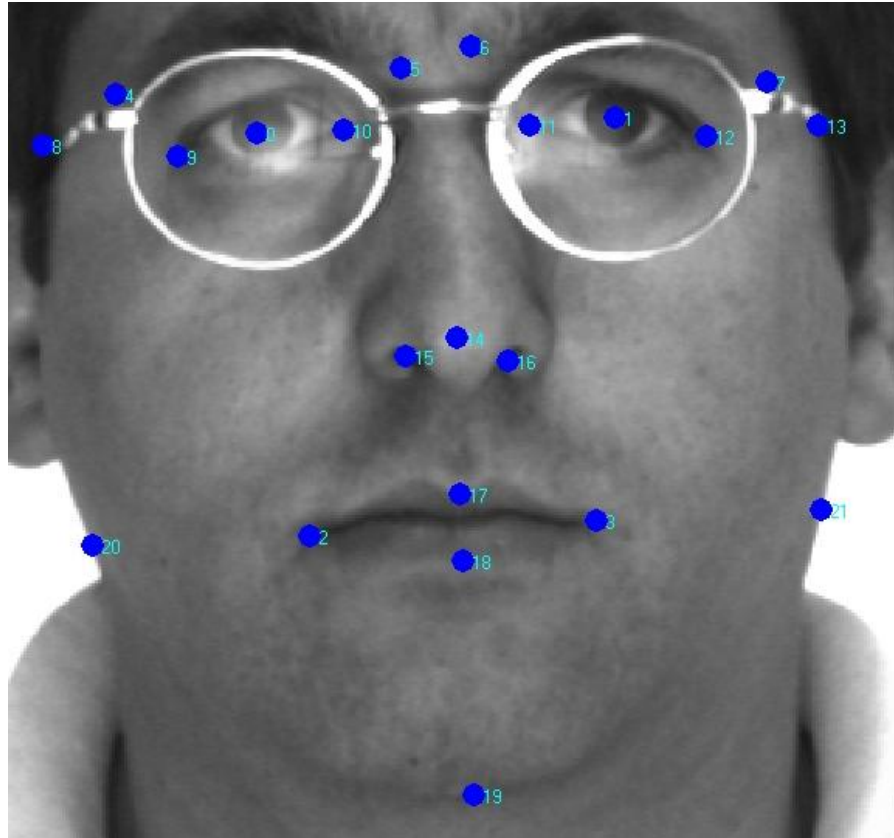


Figura 4-2 - Localização dos 22 pontos de controle previamente marcados na AR\_Face Database.

Conforme pode-se observar na Figura 4-2, os 22 pontos descrevem:

- 0- Centro do olho direito
- 1- Centro do olho esquerdo
- 2- Canto externo direito da boca
- 3- Canto externo esquerdo da boca
- 4- Canto externo inferior da sobrancelha direita
- 5- Canto interno inferior da sobrancelha direita
- 6- Canto interno inferior da sobrancelha esquerda
- 7- Canto externo inferior da sobrancelha esquerda
- 8- Contorno lateral direito do rosto

- 9- Canto externo do olho direito
- 10- Canto interno do olho direito
- 11- Canto interno do olho esquerdo
- 12- Canto interno do olho direito
- 13- Contorno lateral esquerdo do rosto
- 14- Ponta do nariz
- 15- Cavidade nasal direita
- 16- Cavidade nasal esquerda
- 17- Centro superior da boca
- 18- Centro inferior da boca
- 19- Queixo
- 20- Contorno lateral direito do rosto
- 21- Contorno lateral esquerdo do rosto

#### 4.1.2 Banco de Faces da FEI

Esta base de imagens [38] é um banco de faces que contém um conjunto de imagens de faces capturadas entre junho/2005 e março/2006 no Laboratório de Inteligência Artificial da FEI em São Bernardo do Campo, São Paulo, Brasil. Foram capturadas 14 imagens de cada uma das 200 pessoas, totalizando 2800 imagens. Todas as imagens são coloridas e tiradas em um fundo branco com uma rotação de perfil de 180° a partir da imagem frontal, conforme descrito em [40]. A escala varia aproximadamente em 10% e o tamanho original de cada imagem é de 640x480 pixels. Todas as faces são de estudantes e funcionários da FEI, entre 19 e 40 anos, com aparências distintas, tipo de cabelo e adornos. O número de homens e mulheres é exatamente o mesmo. A Figura 4-3 ilustra um exemplo de fotos tiradas de um indivíduo que pertence a base.

O banco de faces da FEI é composto, também, por um aditivo que contém apenas imagens frontais previamente alinhadas a partir de um modelo padrão, sendo assim, uma característica encontrada em um pixel de uma determinada imagem pode ser encontrada no mesmo pixel das demais imagens. Este alinhamento manual foi feito a partir da direção dos olhos e nariz como referência [40]. O tamanho de cada uma dessas imagens é 360x260 pixels.



Figura 4-3 - Base FEI. Exemplo das fotos tiradas, mostrando as características de cada uma das 14 posições.

Nos experimentos realizados com esta segunda base de dados, foram utilizados 46 pontos, conforme Figura 4-4, dentre eles, os 19 dos 20 pontos mais relevantes de características frontais conforme Maja [22] e mais 27 pontos dos 152 sugeridos por Tim Cootes [20]. Estes 26 pontos foram adicionados para permitir que o software *am\_tools* [19], desenvolvido por Tim Cootes, tivesse uma quantidade de pontos suficiente para gerar uma representação da face adequada para a visualização das formas.



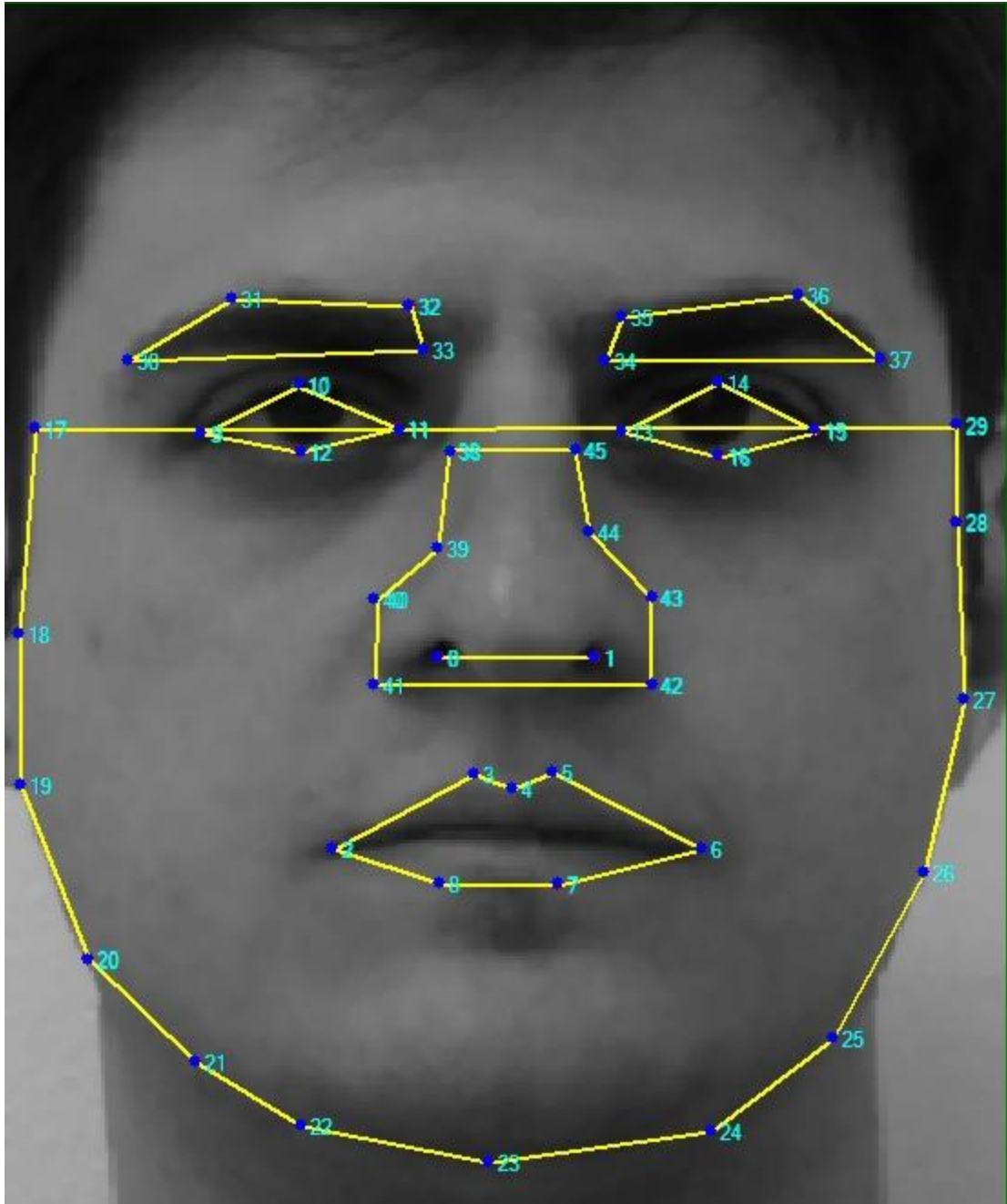


Figura 4-4 – Imagem com os 46 pontos definidos no software am\_tools [19].

Conforme pode-se observar na Figura 4-4, foram criados 46 pontos para cada imagem, sendo eles:

- 0- Cavidade nasal esquerda
- 1- Cavidade nasal direita
- 2- Canto externo esquerdo da boca

- 3 a 5- Contorno da parte central superior da boca
- 6- Canto externo direito da boca
- 7 e 8- Contorno da parte central inferior da boca
- 9- Canto externo do olho esquerdo
- 10- Centro superior do olho esquerdo
- 11- Canto interno do olho esquerdo
- 12- Centro inferior do olho esquerdo
- 13- Canto interno do olho direito
- 14- Centro superior do olho direito
- 15- Canto externo do olho direito
- 16- Centro inferior do olho direito
- 17 a 23– Contorno lateral esquerdo do rosto
- 23- Queixo
- 23 a 29 – Contorno lateral direito do rosto
- 30- Canto externo inferior da sobrancelha esquerda
- 31- Canto externo superior da sobrancelha esquerda
- 32- Canto interno superior da sobrancelha esquerda
- 33- Canto interno inferior da sobrancelha esquerda
- 34- Canto interno inferior da sobrancelha direita
- 35- Canto interno superior da sobrancelha direita
- 36- Canto externo superior da sobrancelha direita
- 37- Canto externo inferior da sobrancelha direita
- 38 a 41 – Contorno lateral esquerdo do nariz

42 a 45 – Contorno lateral direito do nariz.

## 4.2 Experimentos

Neste trabalho foram utilizadas 238 imagens frontais sendo 148 imagens masculinas e 90 imagens femininas existentes na base AR e 400 imagens frontais sendo 100 mulheres e 100 homens para a base FEI, sendo que cada individuo foi fotografado duas vezes – com expressão facial sorrindo e não sorrindo. Os experimentos visaram avaliar quais são as características que representam maior variância quando nos referimos ao formato do rosto.

Para os experimentos das bases AR e FEI, utilizou-se o número de amostras conforme Figura 4-5.

	Base AR	Base FEI
Número total de imagens	238	400
Gênero (homens / mulheres)	148 / 90	200 / 200
Expressão Facial (sorrindo / não sorrindo)	119 / 119	200 / 200
Total de variáveis	44	92
Matriz de dados	238x44	400x92

Figura 4-5 – Detalhamento das bases AR e FEI

Foram geradas então matrizes de dados de dimensionalidade 238x44 para a base AR, onde as linhas 1 a 148 representam as imagens do grupo 1 e as linhas 149 a 238 representam as imagens do grupo 2 para gênero, e as linhas 1 a 119 representam as imagens do grupo 1 e as linhas 120 a 238 representam as imagens do grupo 2 para expressão facial. Para a base FEI, foram geradas matrizes de dados de dimensionalidade 400x92, onde as linhas de 1 a 200 representam as imagens do grupo 1 e as linhas 201 a 400 representam as imagens do grupo 2 tanto para gênero como para expressão facial. As colunas destas matrizes representam as coordenadas x,y de cada ponto.

Embora esta ordenação preserve a questão de separar as formas pelo grupo ao qual pertence, esta ordem não é levada em consideração pelo PCA. Porém, as informações da

classe que cada amostra pertence será levada em consideração utilizando o LDA / MLDA e o SVM.

Adotou-se o método de *cross-validation* [41] que é geralmente utilizado para estimar a capacidade de generalização de um classificador estatístico. O *cross-validation* ocorre por meio da divisão dos dados em subconjuntos e a execução desses subconjuntos de teste-treino em diferentes combinações, fazendo um bom uso dos dados disponíveis em cada padrão mesmo quando os dados são utilizados tanto como treinamento quanto teste. Este método é especialmente útil quando não há dados suficientes para formar um conjunto exclusivo de treinamento, validação e teste, evitando sobreposição de dados.

Por meio do *k-fold* é possível estimar o desempenho estatístico do *cross-validation*, medindo a execução de cada um dos subconjuntos [41]. Para realizar o *k-fold*, a amostra original é dividida aleatoriamente em k subgrupos de tamanho igual. Dos k subgrupos, um é separado para ser o grupo de teste e os outros k-1 grupos são utilizados como grupos de treinamento. Este processo é repetido k vezes, para que todos os subgrupos sejam usados uma vez como grupo de teste. Após isso, calcula-se a média dos erros estimados a fim de produzir uma estimativa de erro geral.

A vantagem deste método é que todos os subgrupos são usados tanto para treinamento como para teste. De acordo com [42], é usual realizar o *cross-validation* com k=10, pois experimentos extensivos, realizados por outros autores, mostraram que, quando possível, esta é a melhor escolha a fim de se conseguir uma estimativa precisa.

Para estes experimentos, foi aplicado o método *k-fold cross validation* para k=10, ou seja, dividiu-se o conjunto de treinamento em 10 grupos. Para classificação foi utilizado o classificador de distância de Mahalanobis. Assim, a média de cada classe i foi calculada a partir das imagens de treinamento correspondentes e a distância de Mahalanobis de cada classe  $\bar{x}_i$  foi usada para atribuir um teste de observação  $x_t$  a cada grupo de cada experimento. Isto é, fez-se a atribuição de  $x_t$  para a classe i que minimiza:

$$d_i(x_t) = \sum_{j=1}^k \frac{1}{\lambda_j} (x_{tj} - \bar{x}_{ij})^2, \quad (4.1)$$

onde  $\lambda_j$  é o autovalor correspondente e k é o número de componentes principais retidas. Em todos os experimentos de reconhecimento, foi considerado um número diferente de

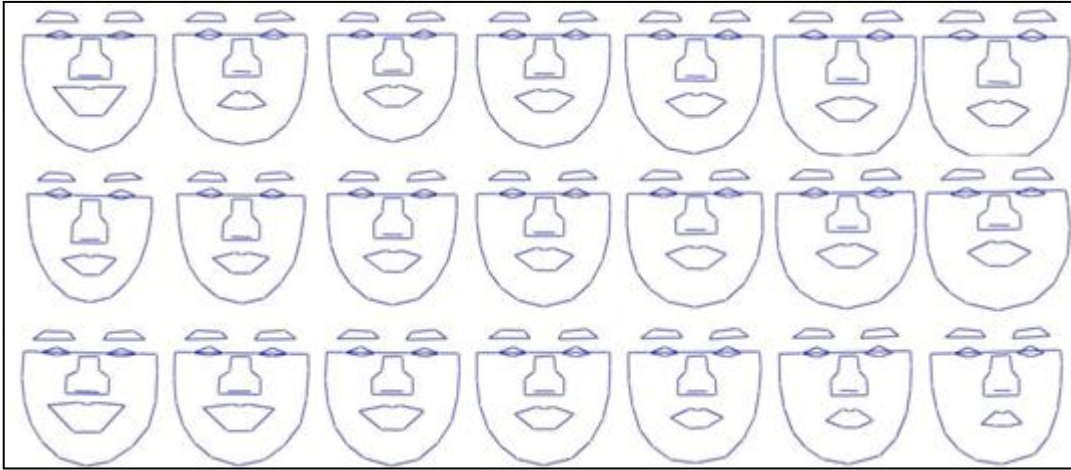
componentes principais para calcular as taxas de reconhecimento dos métodos correspondentes.

### 4.3 Resultados: Componentes Principais

Na Figura 4-6 e na Figura 4-7 pode-se observar as reconstruções visuais das três primeiras componentes principais para as bases AR e FEI, respectivamente. Na Figura 4-6, pode-se verificar, para as bases AR, que a primeira e terceira componentes principais (1ª e 3ª linhas) modelaram principalmente variações no tamanho da face. Não é possível dizer o que a segunda componente principal modelou exatamente. Na Figura 4-7, pode-se observar, para a base FEI, que a primeira componente principal (1ª linha) modelou principalmente variações no tamanho da face, a segunda componente principal modelou o formato da face e a terceira componente principal modelou expressão facial.



$-3\sqrt{\lambda}$      $-2\sqrt{\lambda}$      $\sqrt{\lambda}$     Face média     $+\sqrt{\lambda}$      $+2\sqrt{\lambda}$      $+3\sqrt{\lambda}$   
 Figura 4-6 - Reconstruções visuais variando-se as três primeiras componentes principais para a base AR.



$-3\sqrt{\lambda}$      $-2\sqrt{\lambda}$      $-\sqrt{\lambda}$     Face média     $+\sqrt{\lambda}$      $+2\sqrt{\lambda}$      $+3\sqrt{\lambda}$   
 Figura 4-7 - Reconstruções visuais variando-se as três primeiras componentes principais para a base FEI.

Foi calculada a proporção da variância total para ambas as bases. Para a base AR, as 3 primeiras componentes principais são responsáveis por 89% da variância total. Já para a base FEI, as 3 primeiras componentes principais são responsáveis por 62,7% da variância total. Sabendo que as primeiras componentes da base FEI representam uma quantidade menor de variância quando comparada com a base AR, pode-se concluir que existe uma redundância menor nos dados da FEI quando comparadas com a base AR.

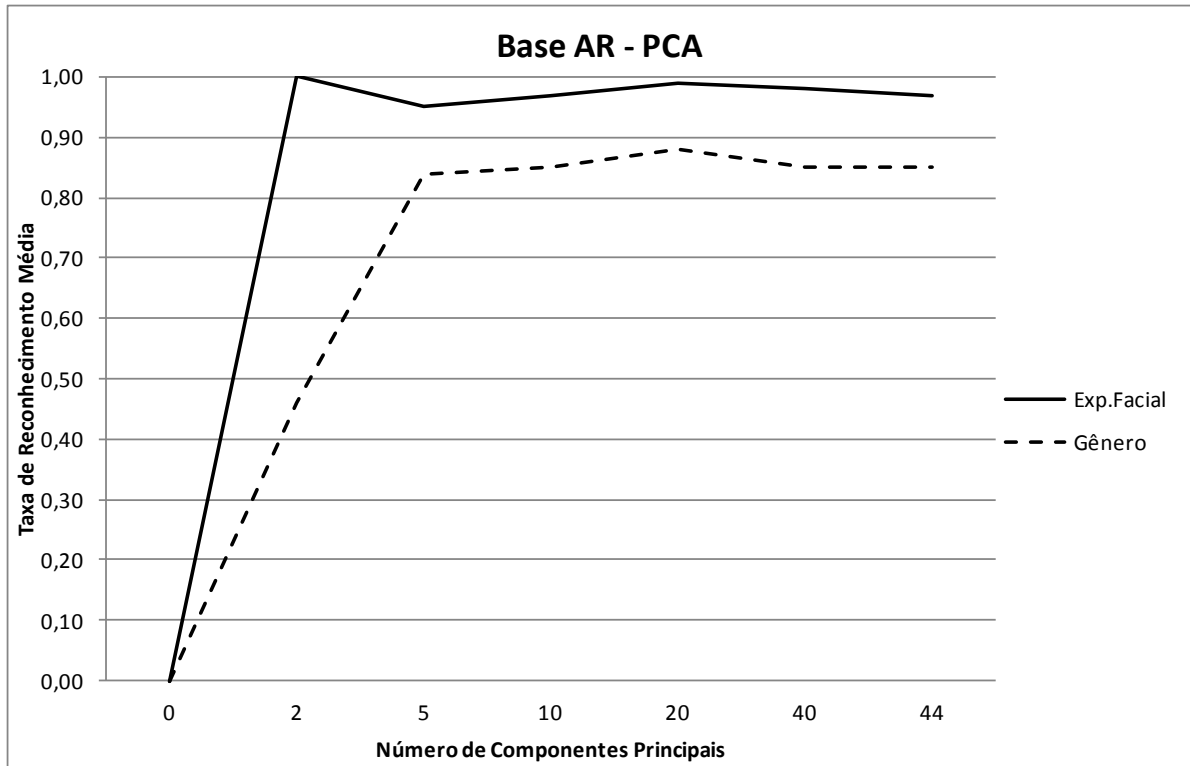


Figura 4-8 – Taxa de acerto do PCA para a base AR.

Em termos de taxa de reconhecimento média, pode-se concluir pela Figura 4-8 que na base AR é possível atingir uma taxa de reconhecimento de 100% com apenas 2 componentes principais para expressão facial. Para gênero, é possível atingir uma taxa de reconhecimento de no máximo 88%. Essa taxa é atingida com 20 componentes principais.

Observando a Figura 4-9, pode-se concluir que na base FEI é possível atingir uma taxa de reconhecimento de 99% com apenas 10 componentes principais para expressão facial. Para gênero, é possível atingir uma taxa de reconhecimento de 98% com 30 componentes principais.

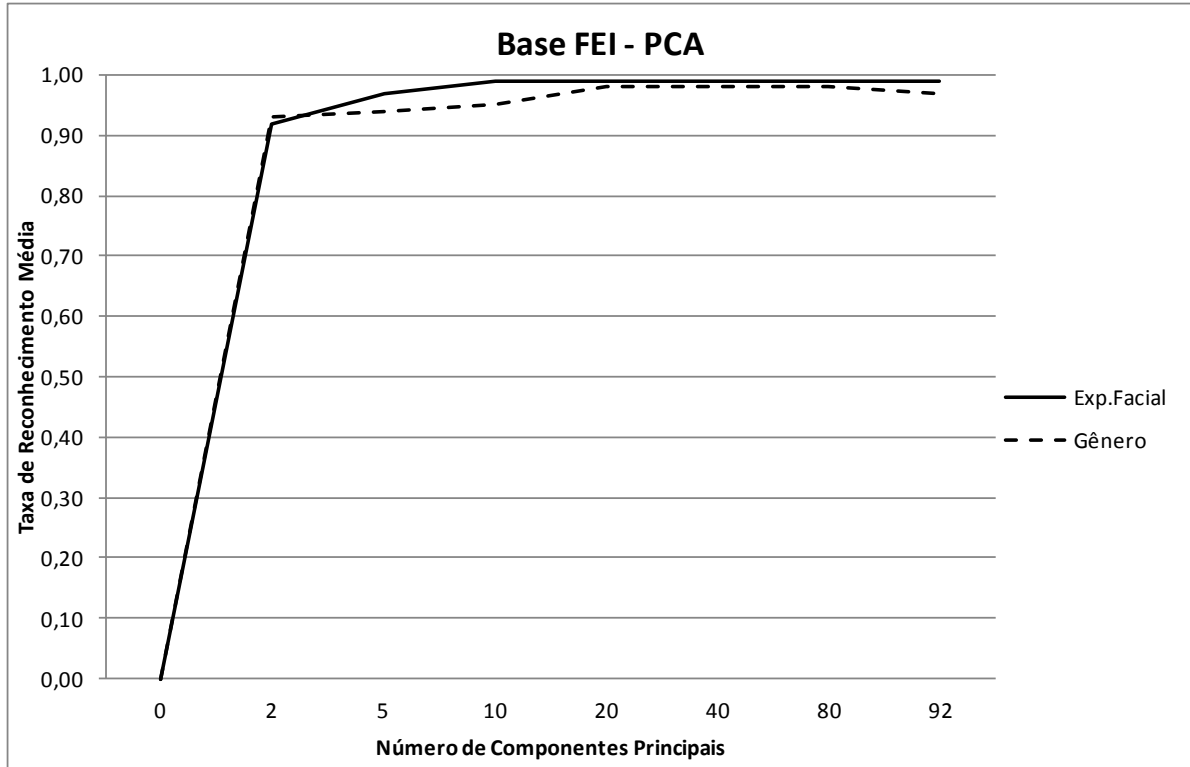


Figura 4-9 – Taxa de acerto do PCA para a base FEI.

#### 4.4 Resultados: Discriminantes Lineares

Antes da análise visual de como essas formas estão discriminadas no espaço  $n$ -dimensional e da interpretação dessa reconstrução pelos hiperplanos de separação, apresentase o desempenho de separação linear desses hiperplanos.

	Gênero		Expressão Facial	
	Homem	Mulher	Sorrindo	Não sorrindo
LDA	94,26 (1,50)	93,11 (0,87)	99,92 (0,30)	99,66 (0,66)
MLDA	86,62 (1,40)	87,22 (1,02)	97,48 (0,45)	96,39 (0,78)
SVM	90,74 (11,54)	98,67 (0,87)	99,75 (0,46)	99,83 (0,40)

Figura 4-10 – Taxa de classificação (%) e desvio padrão entre parênteses (%) para características gênero (homem versus mulher) e expressão facial (não sorrindo versus sorrindo) para a base AR.



	Gênero		Expressão Facial	
	Homem	Mulher	Sorrindo	Não sorrindo
LDA	96,10 (1,08)	96,45 (1,14)	95,90 (1,47)	98,75 (0,81)
MLDA	87,95 (1,87)	91,95 (2,34)	92,20 (0,66)	97,15 (1,35)
SVM	98,70 (0,80)	98,50 (2,05)	98,95 (1,14)	99,35 (0,90)

Figura 4-11 – Taxas de classificação (%) e desvio padrão entre parênteses (%) para características gênero (homem versus mulher) e expressão facial (não sorrindo versus sorrindo) para a base FEI.

O resultado deste experimento para a base AR pode ser observado na Figura 4-10 e, para a base FEI, na Figura 4-11. Pode-se verificar que, em linhas gerais, todos os métodos têm uma boa taxa de reconhecimento tanto para discriminação de expressão facial como de gênero para as duas bases utilizadas, ou seja, acima de 80%. Esses hiperplanos têm potencial para descrever características visuais discriminantes que apresentam uma boa separação dos dados no espaço original, mas estão limitados a uma única direção de reconstrução das formas [35].

Para visualizar as mudanças descritas pelos hiperplanos de separação do LDA, MLDA e SMV foram utilizados os desvios padrão de cada grupo a partir da média dos grupos considerando uma distribuição normal [35]. Essas mudanças podem ser vistas nas figuras a seguir.

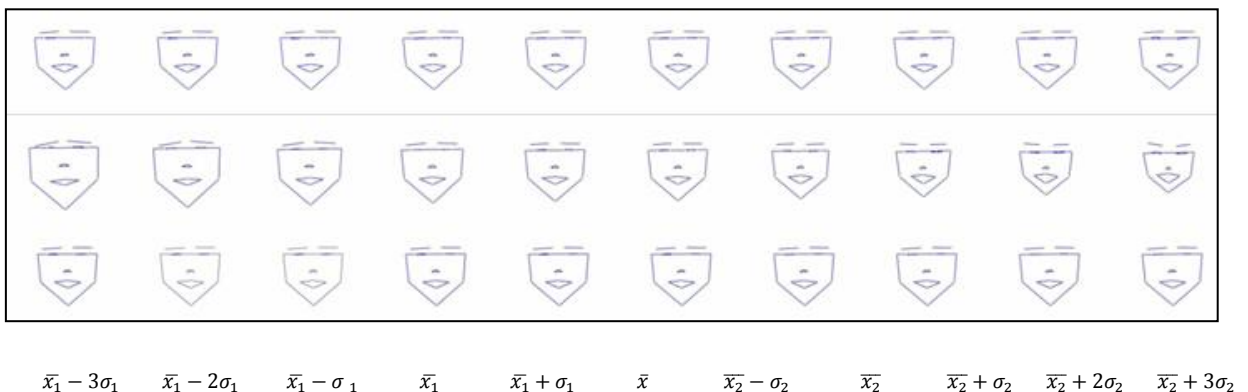


Figura 4-12 - Reconstruções visuais navegando-se no hiperplano discriminante do LDA, MLDA e SVM (extrapolando os vetores suportes) e com relação ao gênero (de homens para mulheres) para a base AR.

Observando a Figura 4-12, pode-se verificar a partir da reconstrução do MLDA que aparentemente os homens têm um rosto maior e mais largo que os rostos das mulheres. Pode-se observar, também, que as mulheres têm a sobrancelha mais fina e mais inclinada.

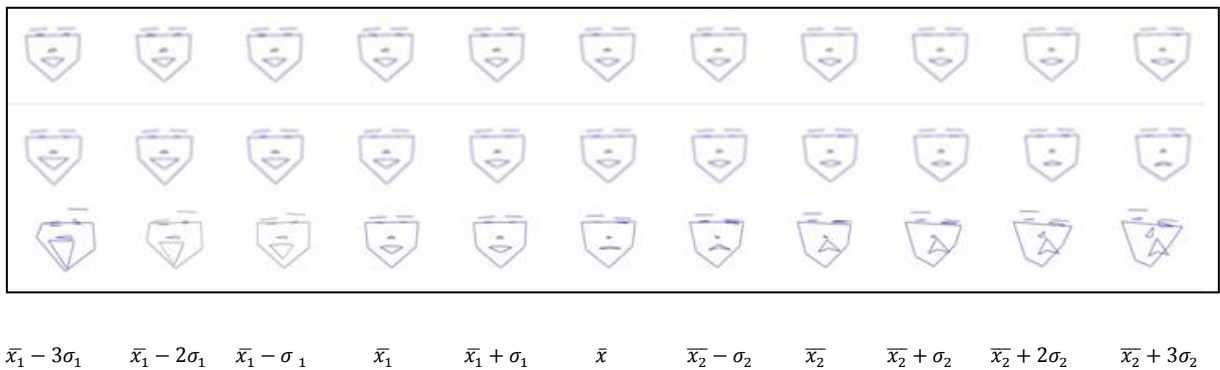


Figura 4-13 - Reconstruções visuais navegando-se no hiperplano discriminante do LDA, MLDA e SVM (extrapolando os vetores suportes) e com relação à expressão facial (de sorrindo para não sorrindo) para a base AR.

Observando a Figura 4-13, pode-se verificar que o que mais varia para expressão facial é o formato da boca. O formato do rosto e nariz variam, mas de forma bem mais sutil.

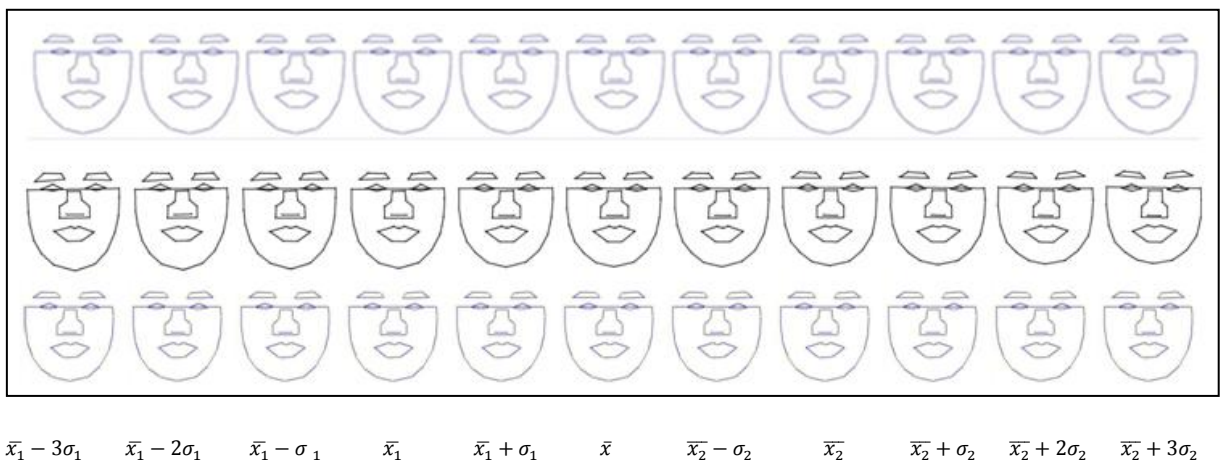
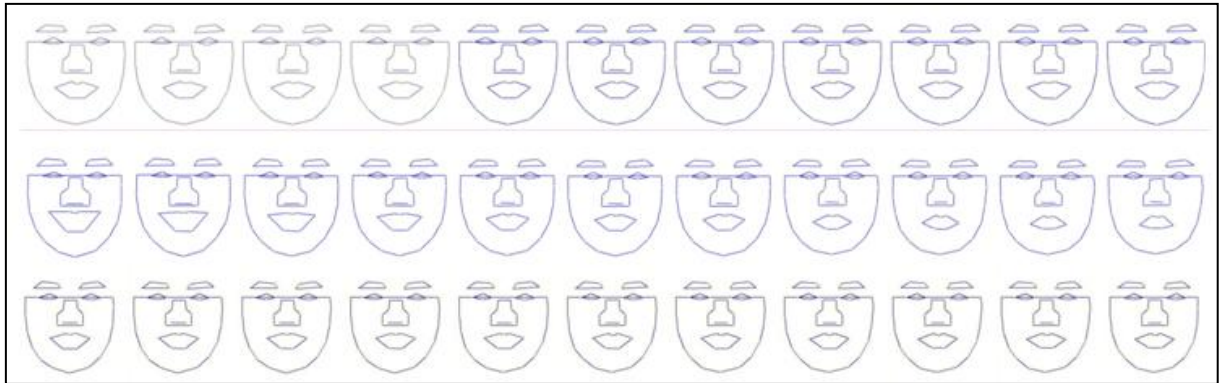


Figura 4-14 - Reconstruções visuais navegando-se no hiperplano discriminante do LDA, MLDA e SVM (extrapolando os vetores suportes) com relação ao gênero (de homens para mulheres) para a base FEI.

A partir da Figura 4-14, é possível visualizar diferenças muito sutis na largura e comprimento do rosto, formato da sobrancelha e largura do nariz para a base FEI. Essas diferenças podem ser melhor observadas na reconstrução da navegação discriminante do MLDA.



$\bar{x}_1 - 3\sigma_1$     $\bar{x}_1 - 2\sigma_1$     $\bar{x}_1 - \sigma_1$     $\bar{x}_1$     $\bar{x}_1 + \sigma_1$     $\bar{x}$     $\bar{x}_2 - \sigma_2$     $\bar{x}_2$     $\bar{x}_2 + \sigma_2$     $\bar{x}_2 + 2\sigma_2$     $\bar{x}_2 + 3\sigma_2$

Figura 4-15 - Reconstruções visuais navegando-se no hiperplano discriminante do LDA, MLDA e SVM (extrapolando os vetores suportes) e com relação à expressão facial (de sorrindo para não sorrindo) para a base FEI.

De forma análoga, a partir da Figura 4-15, é possível visualizar diferenças principalmente no formato da boca para a base FEI também. Essa diferença pode ser facilmente visualizada na reconstrução da navegação discriminante do MLDA.

Observando a Figura 4-12, Figura 4-13, Figura 4-14 e Figura 4-15 pode-se observar pouca variação entre as reconstruções visuais do LDA. Isso ocorre, pois neste trabalho temos limitação do número de amostras, o que acarreta em degradação do desempenho do LDA devido à instabilidade da matriz de espalhamento intraclasse.

Para o caso de gênero do SVM, observamos que reconstruindo as imagens apenas entre os vetores suportes não foi possível observar diferenças entre as imagens obtidas, pois os vetores suportes estão muito próximos um do outro. Sendo assim, decidimos por extrapolar as margens (vetores suportes) e reconstruir outras imagens fora deste limite. No entanto, em ambas as bases, mesmo após a extrapolação, não é possível visualizar mudanças significativas. Já para o caso de expressão facial do SVM, pode-se observar que existem distorções quando extrapolamos os hiperplanos (vetores suportes) para a base AR. Isso ocorre, pois o método SVM é um método baseado exclusivamente nos vetores de suporte e não é necessariamente possível descrever as possíveis distorções após esses limites. Já para a base FEI, mesmo extrapolando, não é possível observar diferenças significativas entre as imagens.

Conclui-se a partir da Figura 4-12, Figura 4-13, Figura 4-14 e Figura 4-15 que o método MLDA representa visualmente a melhor discriminação das formas, evidenciando a viabilidade de conduzir experimentos regularizados em situações como essa de número de formas limitada.

#### 4.5 Resultados: Componentes Discriminantes

Foram listadas as 20 primeiras componentes principais com o maior peso em valor absoluto de discriminancia, tanto para gênero como para expressão facial, conforme apresentado na Figura 4-16 para a base AR e na Figura 4-17 para a base FEI.

Componentes Principais ordenadas por discriminancia - Base AR					
Genero			Exp Facial		
LDA	MLDA	SVM	LDA	MLDA	SVM
44	3	29	35	5	15
39	6	31	5	6	5
31	10	41	42	8	16
27	9	44	39	7	12
17	4	28	28	1	27
42	8	38	16	2	8
33	17	32	25	16	28
10	15	17	15	9	6
28	19	37	30	15	23
19	5	10	20	12	14
24	14	21	44	10	36
26	1	24	8	14	21
25	13	22	14	13	17
22	11	19	12	4	37
38	27	33	6	3	31
15	22	3	36	20	35
41	24	15	29	25	10
9	25	9	13	28	25
32	26	11	33	19	22
37	31	43	34	35	32

Figura 4-16 – As 20 primeiras componentes principais ordenadas pelos hiperplanos do LDA, MLDA e SVM usando a base AR.

Componentes Principais ordenadas por discriminancia - Base FEI					
Genero			Exp Facial		
LDA	MLDA	SVM	LDA	MLDA	SVM
89	13	42	89	3	37
84	5	62	77	6	45
85	7	13	91	4	26
74	15	59	81	9	32
90	16	53	61	12	66
92	11	74	50	10	70
75	6	35	86	13	3
79	14	57	88	11	49
60	2	25	66	2	35
40	3	80	65	7	67
80	12	63	80	26	60
13	20	89	76	16	6
78	17	40	70	25	51
77	4	39	92	17	61
76	18	64	29	29	13
65	25	49	59	5	4
66	1	71	72	8	41
81	26	16	26	27	77
53	23	87	83	19	81
62	22	85	84	15	53

Figura 4-17 – As 20 primeiras componentes principais ordenadas pelos hiperplanos do LDA, MLDA e SVM usando a base FEI.

Pode-se verificar que o LDA e o SVM selecionaram algumas das últimas componentes principais para compor o conjunto das 20 componentes principais mais discriminantes em ambas as bases. Sabe-se que as últimas componentes principais descrevem dados com menor variância e conseqüentemente relacionadas a poucas amostras. Já os hiperplanos de separação do MLDA selecionaram as componentes principais discriminantes baseados nos dados que aparecem em mais exemplos e conseqüentemente são as que melhor caracterizam as diferenças entre grupos.

A variância total explicada pelo PCA padrão e pelas componentes principais mais discriminantes para gênero e expressão facial pode ser observada na Figura 4-18 para a base AR e na Figura 4-19 para a base FEI. Constata-se por essas figuras que a variância das componentes principais do PCA padrão decresce exponencialmente conforme o número de componentes principais aumenta. Esta é uma característica do PCA. No entanto, as

características explicadas pelas componentes principais mais discriminantes não seguem o mesmo comportamento. Observa-se na Figura 4-18 que a ordenação para o LDA priorizou componentes com menor variância, pois as mesmas não representaram diferenças associadas a gênero de forma significativa que pudesse constar no início da ordenação padrão. Como são diferenças sutis, a ordenação tende a deixar componentes com menor variância no final. Isso acontece em todos métodos de separação em ambas as bases, porém é mais evidente no caso de gênero para a base AR.

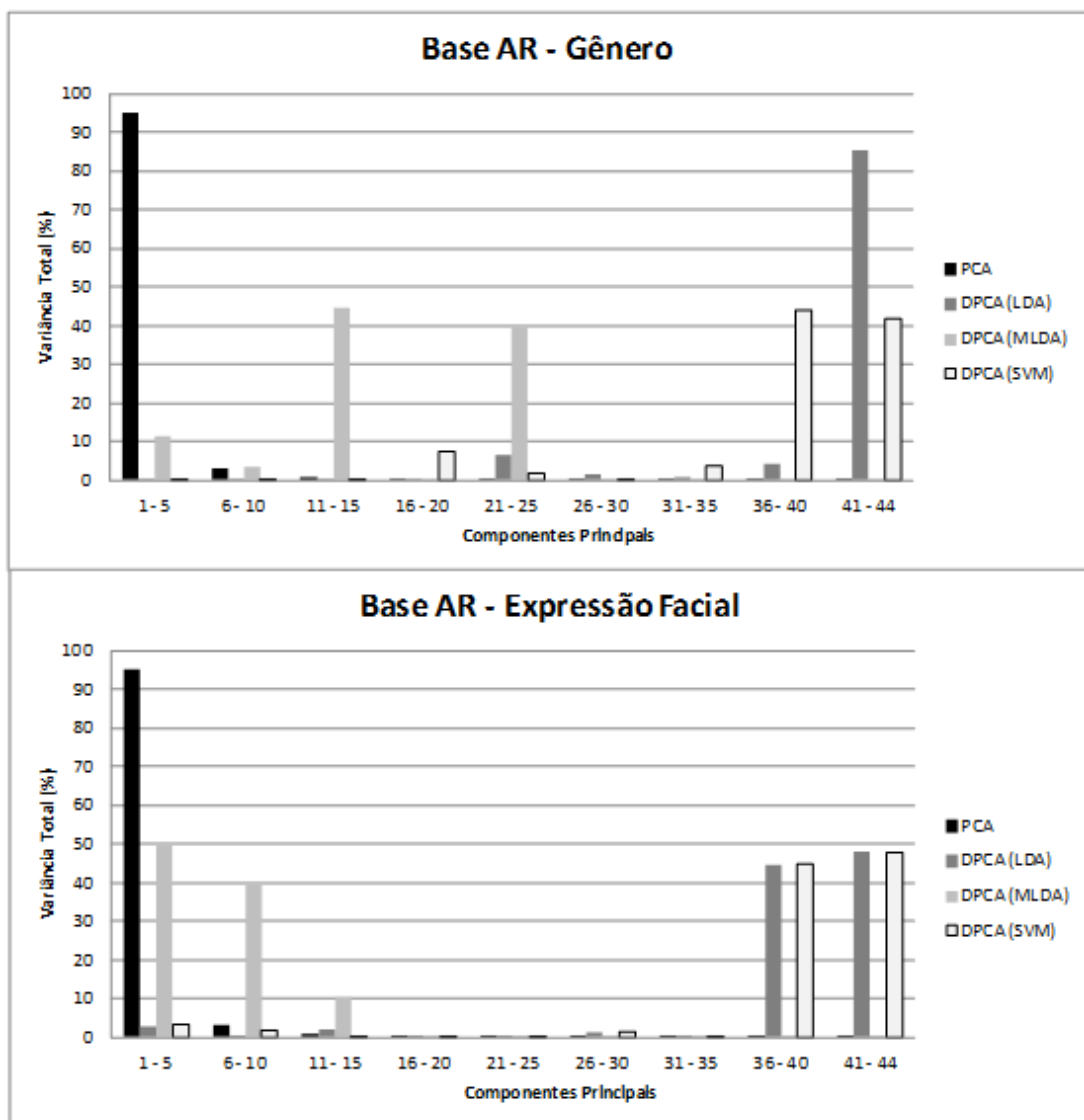


Figura 4-18 - Variância total explicada pela componente principal do PCA padrão e pelas componentes principais mais discriminantes ordenadas pelo LDA, MLDA e SVM. Experimentos de gênero (superior) e expressão facial (inferior) usando a base AR.

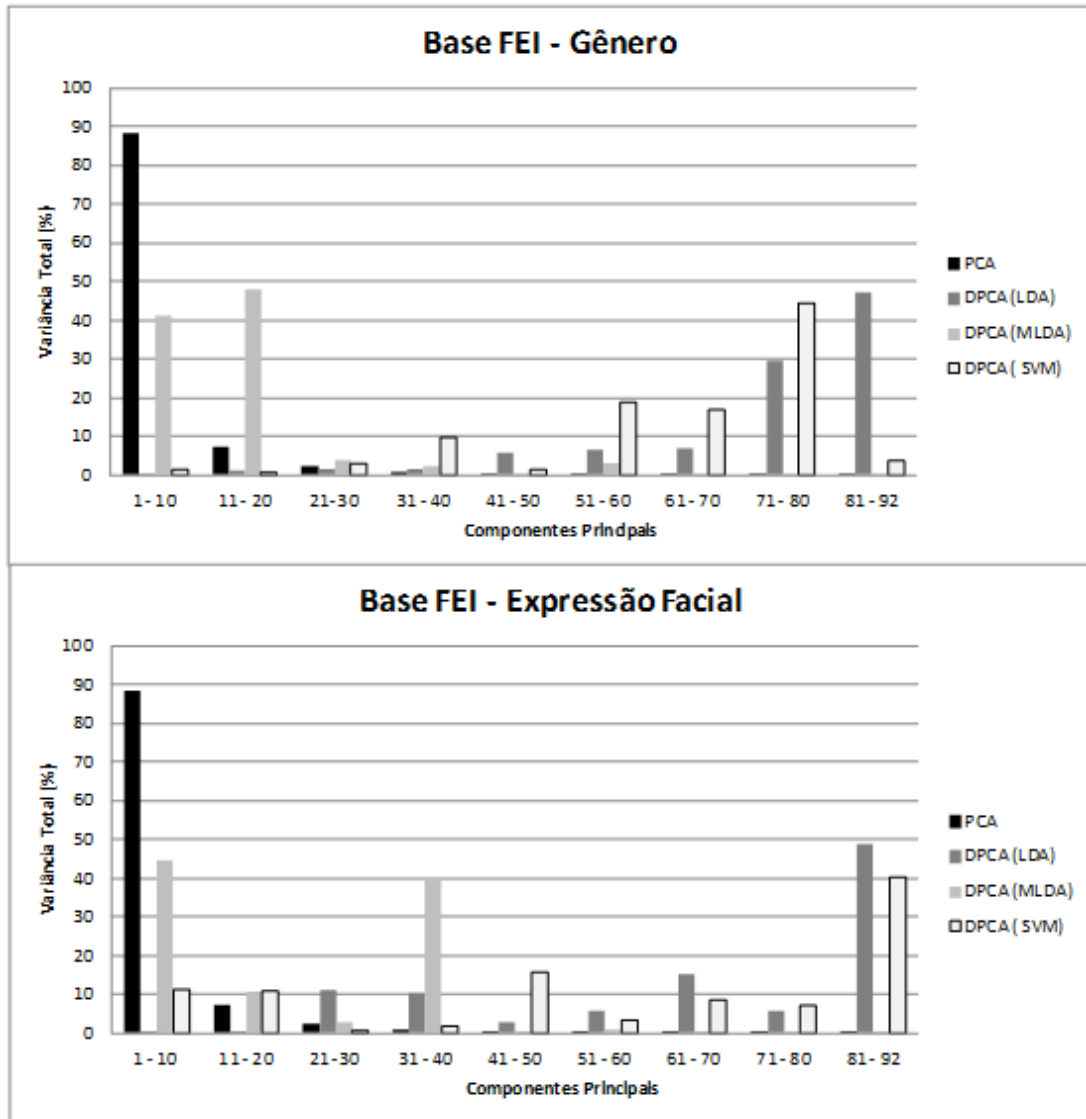


Figura 4-19 – Variância total explicada pela componente principal do PCA padrão e pelas componentes principais mais discriminantes ordenadas pelo LDA, MLDA e SVM. Experimentos de gênero (superior) e expressão facial (inferior) usando a base FEI.

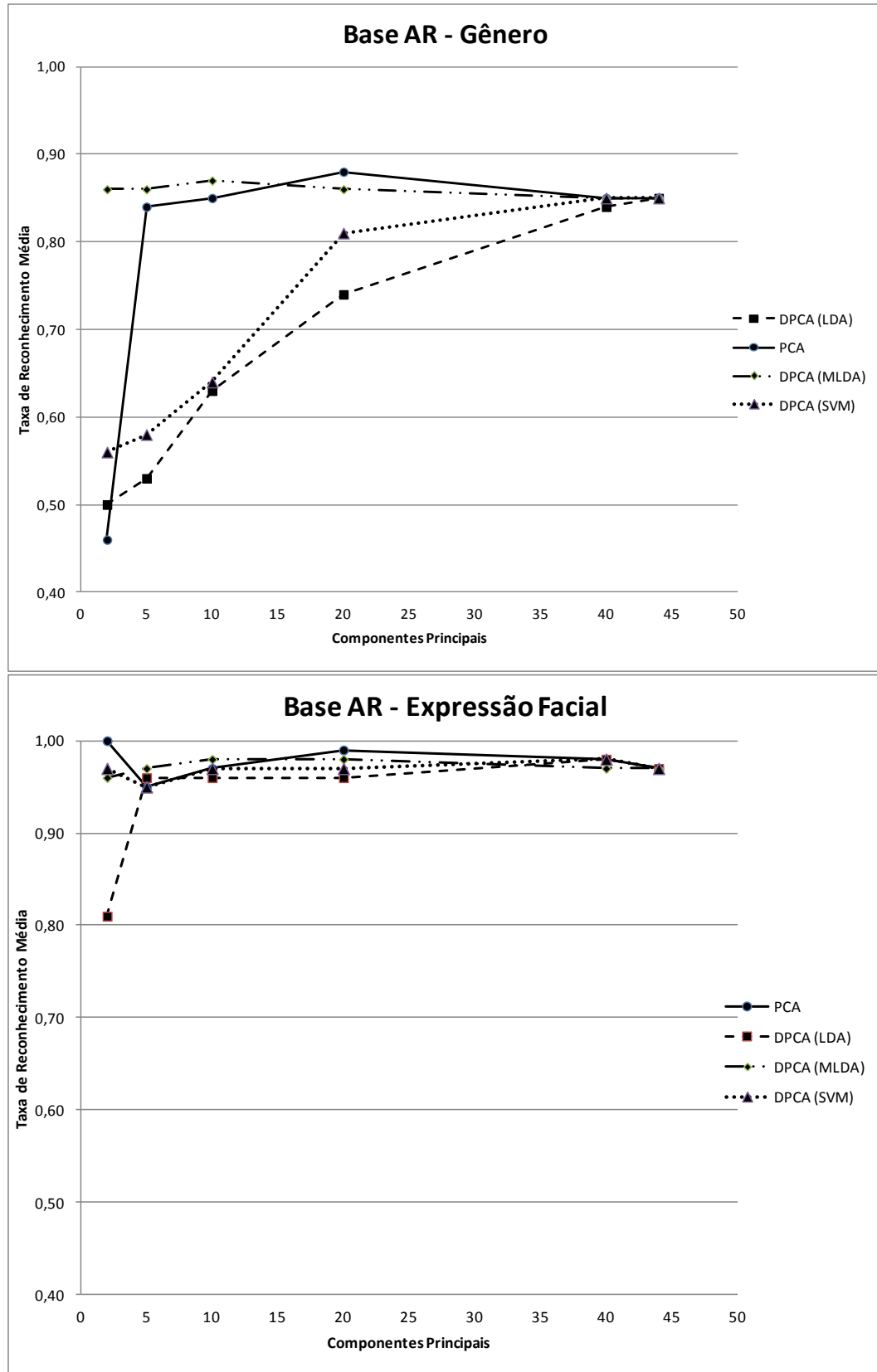


Figura 4-20 - Taxa de acerto do DPCA variando o número de componentes principais para gênero e expressão facial da base AR.



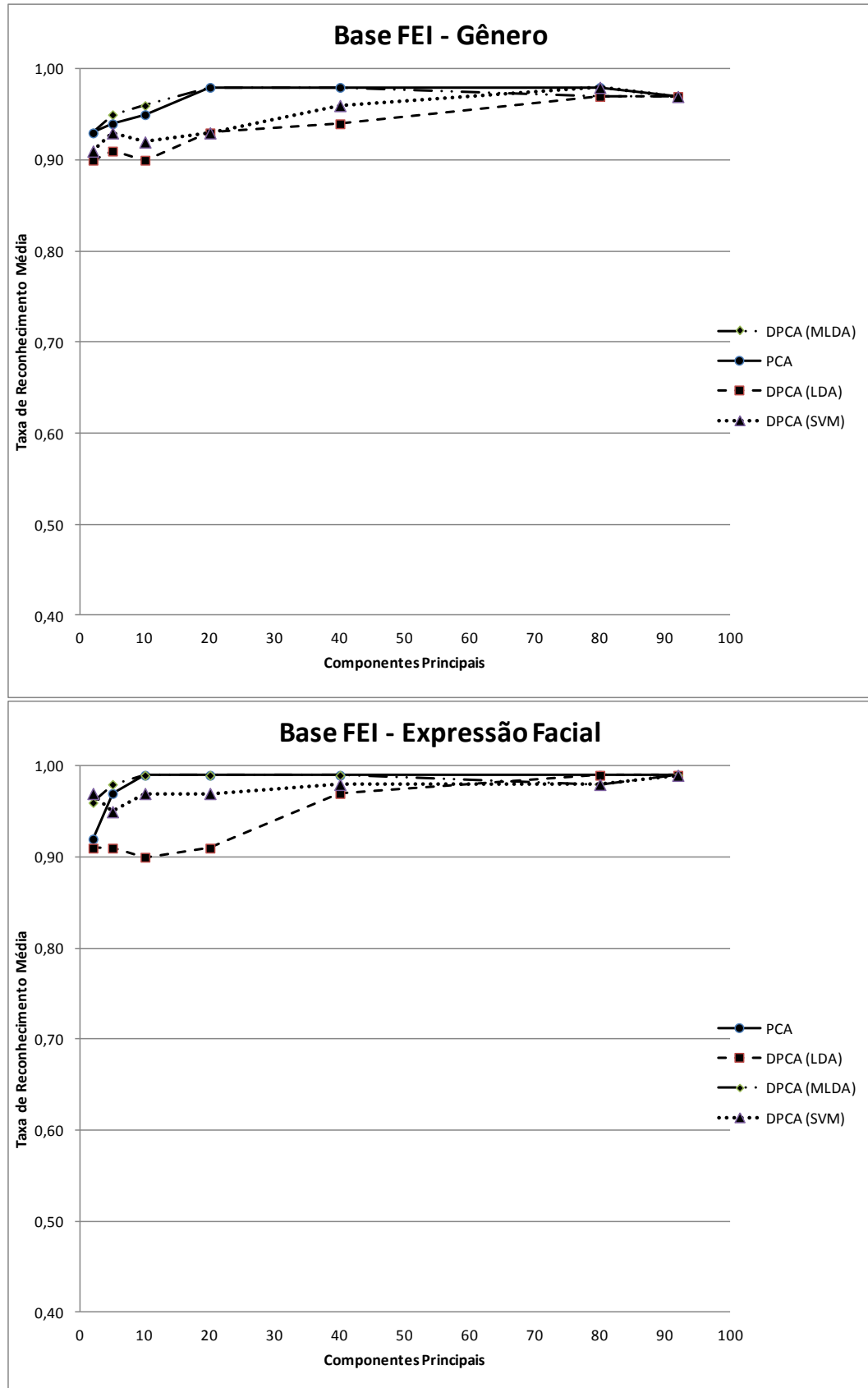


Figura 4-21 - Taxa de acerto do DPCA variando o número de componentes principais para gênero e expressão facial da base FEI.

Observando a Figura 4-20 e a Figura 4-21, pode-se observar que essa reordenação pelas componentes principais aumenta também o poder de discriminação dessas componentes principais. Pode-se observar também que a taxa de acerto para o PCA reorientado a partir do MLDA é superior as outras com exceção dos experimentos de expressão facial na base AR. Isso se deve ao fato de estarmos trabalhando com poucas amostras nesses experimentos.

Para visualizar as mudanças nas formas descritas por essas direções foram utilizados os limites de  $\pm 3\sqrt{\max(\lambda_i, \bar{\lambda})}$  para os hiperplanos de separação do LDA, MLDA e SVM, onde  $\bar{\lambda}$  é a média dos autovalores da matriz de covariância total S, pois alguns  $\lambda_i$  podem ser muito pequenos, indicando nenhuma mudança entre as reconstruções quando move-se ao longo da componente principal correspondente [35]. Essas mudanças podem ser analisadas na Figura 4-22, Figura 4-23 e Figura 4-24 para a base AR e na Figura 4-25, Figura 4-26 e Figura 4-27 para a base FEI.

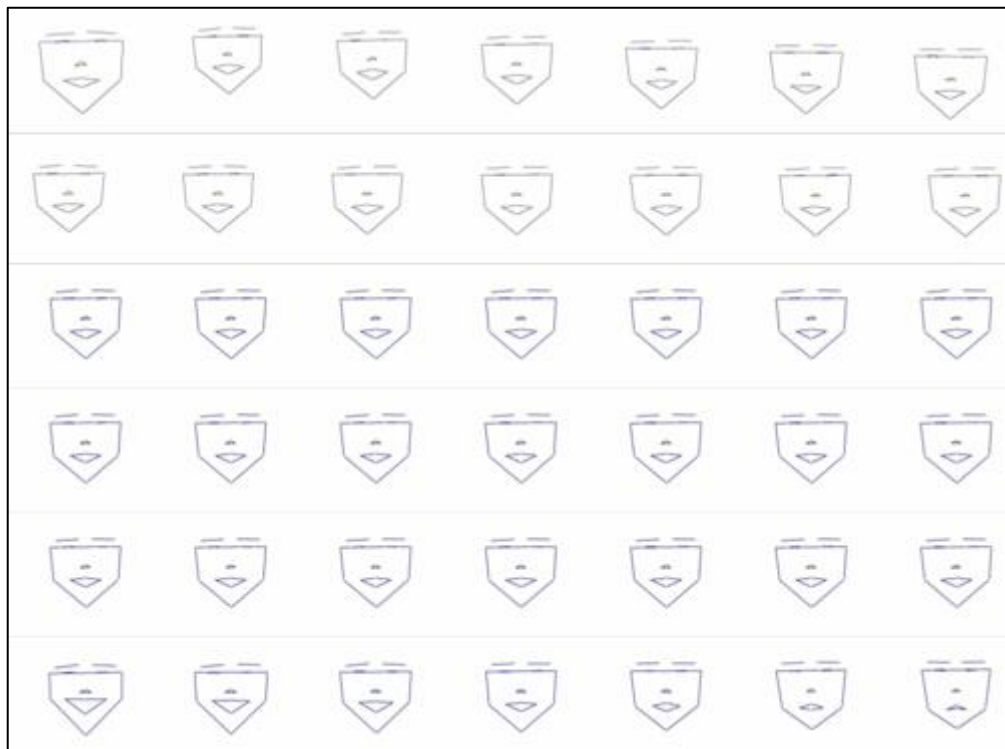


Figura 4-22 - Visualização das mudanças descritas pelas direções principais usando a base AR. De cima para baixo: as 2 primeiras componentes principais do PCA padrão; 2 primeiras componentes principais mais discriminantes selecionadas pelo LDA sendo homens à esquerda e mulheres à direita; e as 2 primeiras componentes principais mais discriminantes pelo LDA sendo sorrindo à esquerda e não sorrindo à direita.

Observando a Figura 4-22, não é possível observar alterações significativas nas duas primeiras componentes principais mais discriminantes selecionadas pelo LDA para gênero

para a base AR. Já para expressão facial, pode-se observar variações no formato da boca, formato e altura do rosto e formato do nariz.

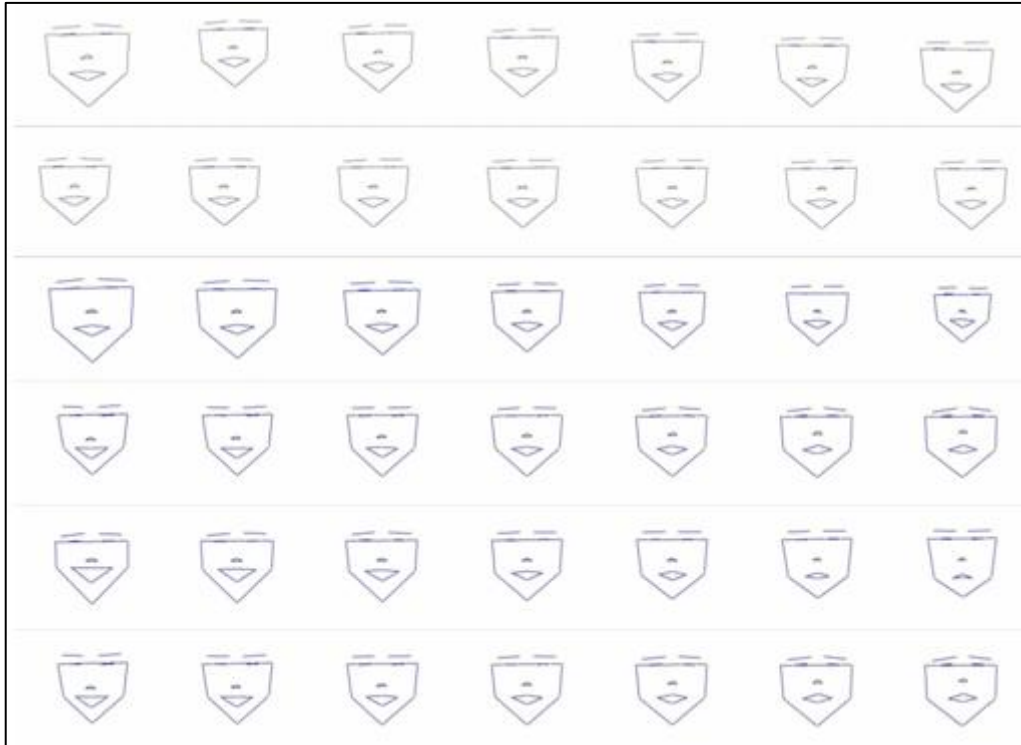


Figura 4-23 - Visualização das mudanças descritas pelas direções principais usando a base AR. De cima para baixo: as 2 primeiras componentes principais do PCA padrão; 2 primeiras componentes principais mais discriminantes selecionadas pelo MLDA sendo homens à esquerda e mulheres à direita; e as 2 primeiras componentes principais mais discriminantes pelo MLDA sendo sorrindo à esquerda e não sorrindo à direita.

Podem-se observar alterações significativas nas reconstruções das duas primeiras componentes principais mais discriminantes selecionadas pelo MLDA. Para gênero da base AR é possível observar alteração no tamanho, largura e altura do rosto, formato da boca e sobrancelha, já para expressão facial (base AR), é possível observar alteração na largura e altura do rosto, formato da boca, do nariz e posição da sobrancelha, conforme Figura 4-23 e nas duas componentes.

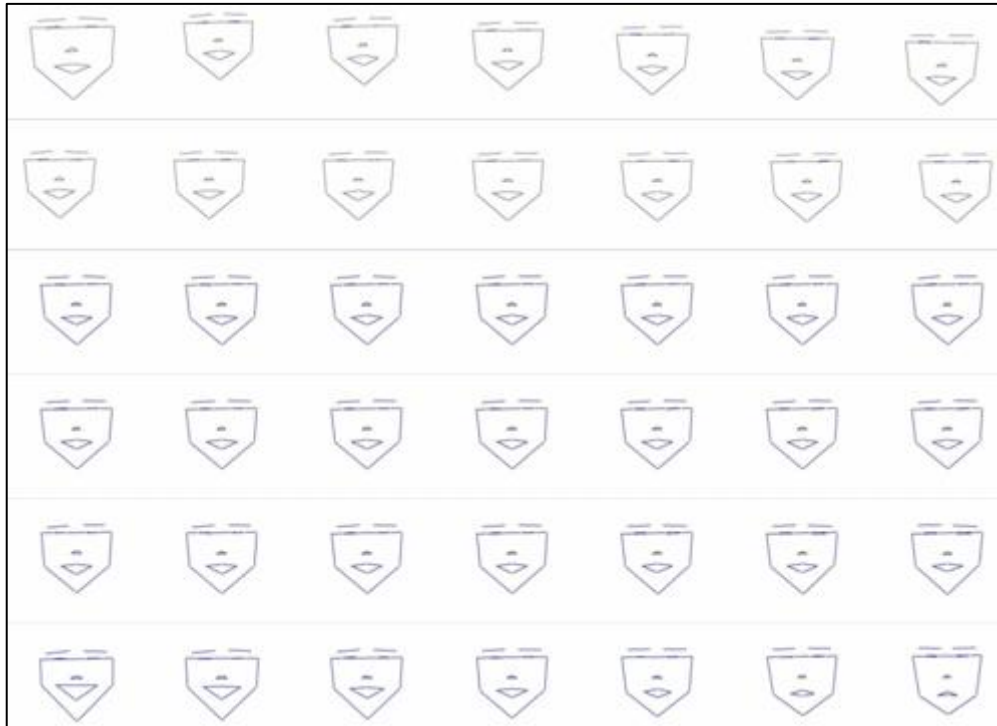


Figura 4-24 - Visualização das mudanças descritas pelas direções principais usando a base AR. De cima para baixo: as 2 primeiras componentes principais do PCA padrão; 2 primeiras componentes principais mais discriminantes seleccionadas pelo SVM sendo homens à esquerda e mulheres à direita; e as 2 primeiras componentes principais mais discriminantes pelo SVM sendo sorrindo à esquerda e não sorrindo à direita.

A Figura 4-24 ilustra alterações nariz, boca e formato da face para expressão facial. Para gênero, não é possível observar alterações significativas.

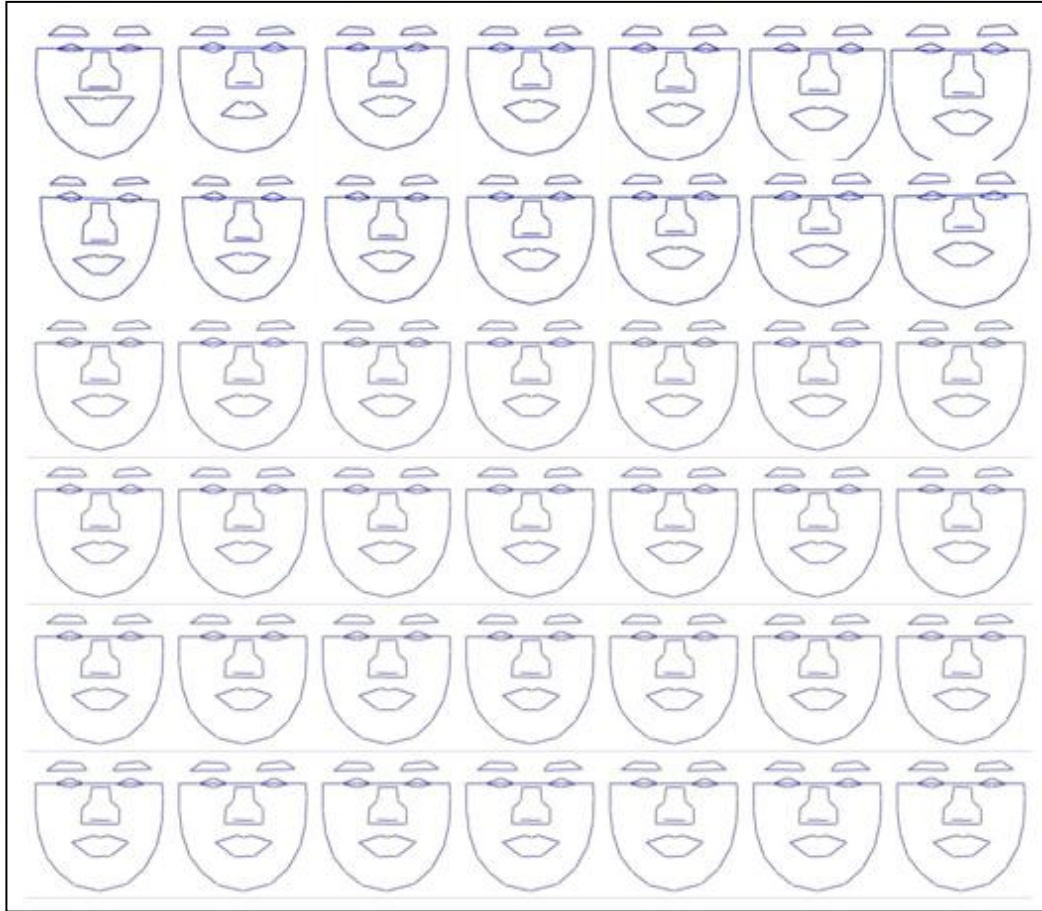


Figura 4-25 – Visualização das mudanças descritas pelas direções principais usando a base FEI. De cima para baixo: as 2 primeiras componentes principais do PCA padrão; 2 primeiras componentes principais mais discriminantes selecionadas pelo LDA sendo homens à esquerda e mulheres à direita; e as 2 primeiras componentes principais mais discriminantes pelo LDA sendo sorrindo à esquerda e não sorrindo à direita.

Observando a Figura 4-25, não é possível observar alterações significativas nas duas primeiras componentes principais mais discriminantes selecionadas pelo LDA tanto para gênero como para expressão facial para a base FEI. Já para expressão facial da base AR, foi possível observar variações no formato da boca, formato e altura do rosto e formato do nariz.

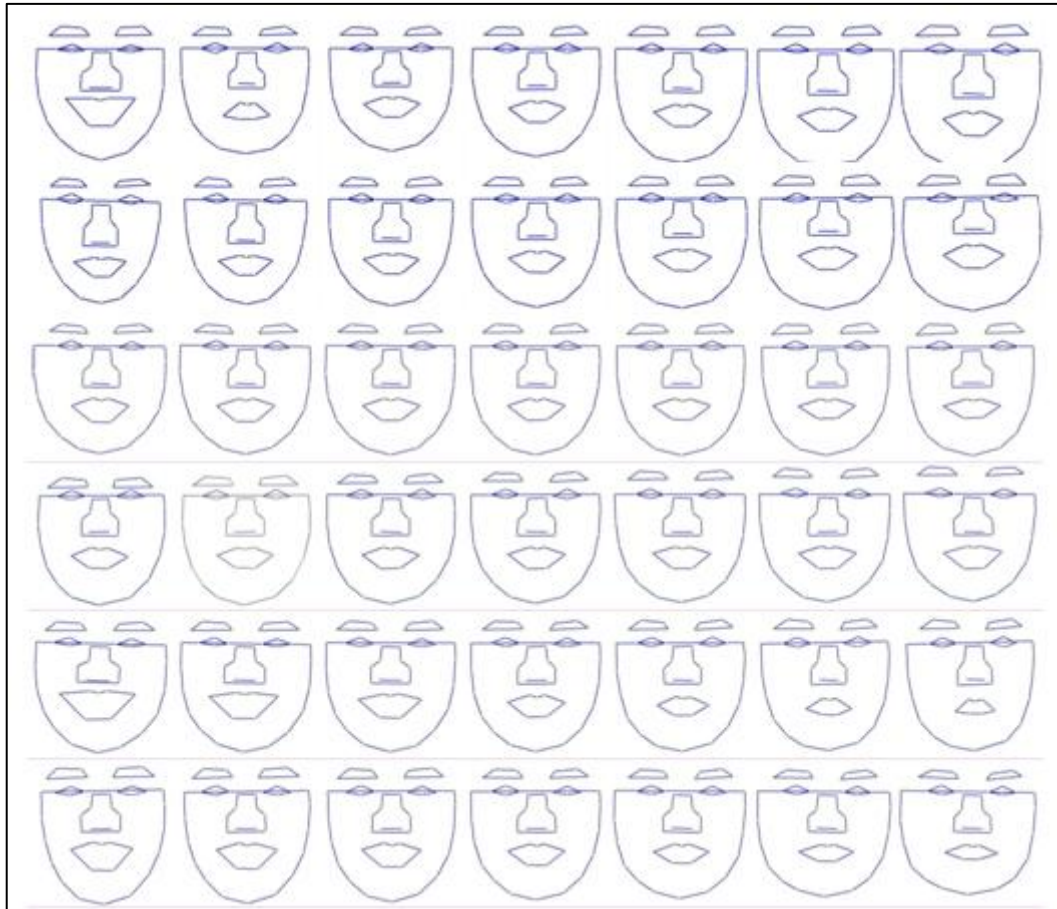


Figura 4-26 - Visualização das mudanças descritas pelas direções principais usando a base FEI. De cima para baixo: as 2 primeiras componentes principais do PCA padrão; 2 primeiras componentes principais mais discriminantes selecionadas pelo MLDA sendo homens à esquerda e mulheres à direita; e as 2 primeiras componentes principais mais discriminantes pelo MLDA sendo sorrindo à esquerda e não sorrindo à direita.

Para a base FEI e seleção das componentes discriminantes pelo MLDA, é possível observar alteração na largura da face e formato da boca para ambos os experimentos, porém só é possível observar variações na altura da face e formato do nariz para expressão facial e só é possível observar alteração na largura e altura da sobrancelha para gênero, conforme mostrado na Figura 4-26.



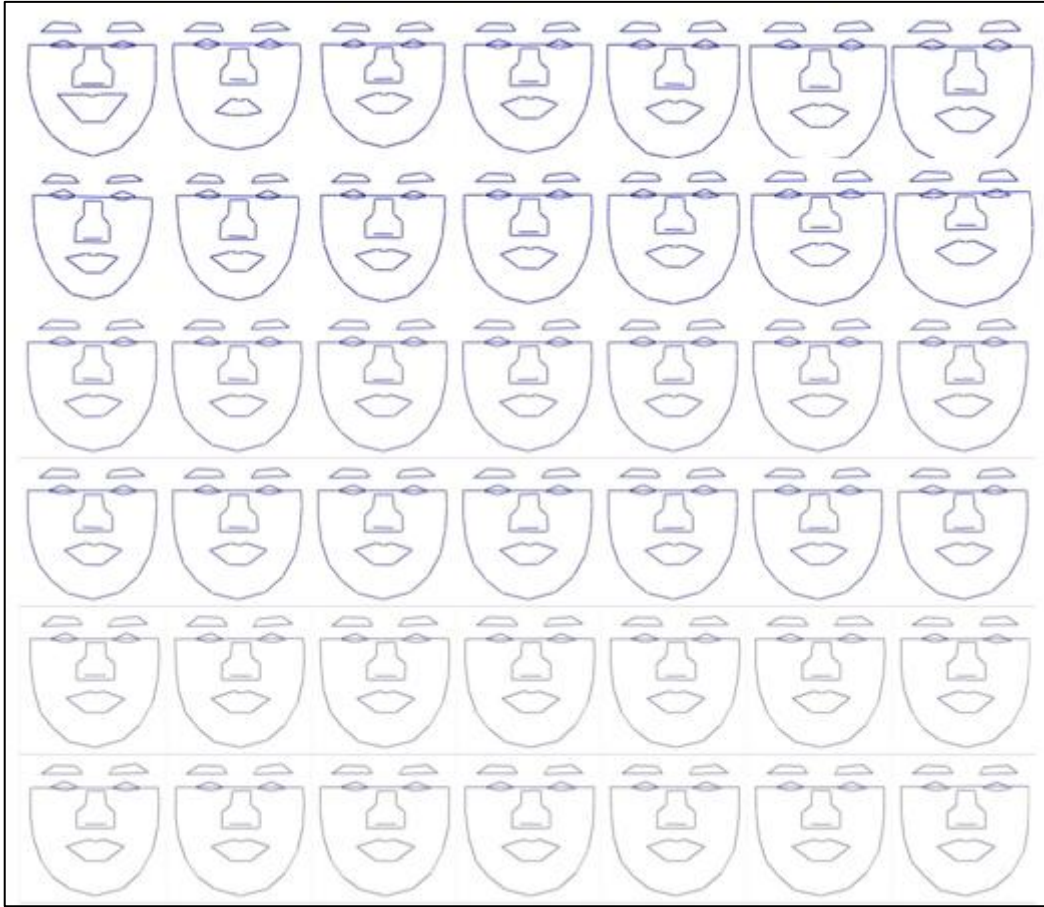


Figura 4-27 - Visualização das mudanças descritas pelas direções principais usando a base FEI. De cima para baixo: as 2 primeiras componentes principais do PCA padrão; 2 primeiras componentes principais mais discriminantes selecionadas pelo SVM sendo homens à esquerda e mulheres à direita; e as 2 primeiras componentes principais mais discriminantes pelo SVM sendo sorrindo à esquerda e não sorrindo à direita.

A dificuldade de visualizar alterações significativas nas reconstruções das duas primeiras componentes principais mais discriminantes pelo SVM também ocorre para a base FEI. Não é possível visualizar alterações significativas tanto para gênero como para expressão facial, conforme Figura 4-27.

## 5 CONCLUSÕES

Todos os modelos aqui descritos são baseados em modelos lineares e propuseram descrever as direções no espaço original nas formas que representam maior variância de acordo com o padrão de interesse. Pudemos observar que as componentes principais podem distinguir tanto expressão facial quanto gênero, que são os objetivos deste trabalho, como também tamanho e largura de face. Salienta-se que foi utilizada uma base que já havia sido pré-processada e/ou normalizada. A necessidade desta normalização deve-se ao fato de ter-se trabalhado com o PCA, que é um método estatístico multivariado não supervisionado. A utilização de uma base não normalizada com certeza traria variações, por exemplo, na altura dos olhos, deslocamento lateral, entre outras diferenças que não são o objeto de estudo deste trabalho.

Variando-se a amostra poderá haver variação na ordem das componentes principais, sendo impossível prever qual componente principal irá descrever as características de interesse, pois essas diferenças podem ser descritas tanto na primeira como na última componente principal. Isso se deve à característica não supervisionada do método.

Foi mostrado que, com estes grupos de amostras, é possível navegar nas componentes principais e que tanto o LDA, como o MLDA e o SVM, localizam a direção que mais discrimina, porém, como este problema está limitado a 2 grupos, eles encontram uma única direção, direção esta que separa os dois grupos. No entanto, os resultados obtidos com o DPCA mostram que existem potencialmente outras direções a serem analisadas para melhor interpretação e compreensão dos dados.

O método DPCA, proposto recentemente, tenta combinar as vantagens de um método não supervisionado como o PCA com os métodos de hiperplano de separação paramétricos e não paramétricos como os que foram apresentados. Esta técnica também está baseada em componentes principais e tem todas as características do PCA, como por exemplo, base de transformação ortonormal, porém, orientada com alguma informação à priori extraída da base de treinamento classificada. Acredita-se que pode-se estender todo e qualquer trabalho de análise e interpretação visual baseado em forma para essa nova transformação linear de dados.



## 6 TRABALHOS FUTUROS

Com base nos resultados obtidos através dos experimentos deste estudo, uma idéia de trabalho futuro seria estender os estudos com relação a outros tipos de imagens, não restritos a faces frontais. Seria interessante investigar, por exemplo, se é possível reconstruir um lado do rosto a partir apenas de imagens do outro lado. Portanto, nesta seção são descritos experimentos preliminares com o objetivo de verificar como os métodos aplicados neste trabalho se comportam para imagens de perfil.



Figura 6-1 – Exemplo de rotação de face (adaptado de [38]).

Devido aos resultados obtidos nos experimentos desta dissertação e a disponibilidade de imagens de perfil da base FEI, os experimentos (PCA, MLDA e PCA reorientado com relação aos hiperplanos do MLDA) foram repetidos para imagens de perfil conforme ilustrado na Figura 6-1. Esta reconstrução para as imagens de perfil da base FEI pode ser visualizada na Figura 6-2. Nesta mesma figura, pode-se observar que a primeira e quarta componente principal (1ª e 4ª linhas) modelaram principalmente a rotação da face. A segunda componente principal modelou a altura e comprimento da face, além da posição da face na imagem. A terceira componente principal modelou o tamanho da face. A quinta, oitava e nona componentes principais modelaram a inclinação da face, sendo que a oitava e a nona modelaram uma inclinação mais sutil. A sexta componente principal modelou a largura da face.

Para as componentes principais discriminantes apresentadas na Figura 6-3, pode-se verificar que houve uma reordenação das primeiras componentes, priorizando as direções que explicam fundamentalmente os movimentos de rotação das formas de face. Portanto, parece também ser possível descrever linearmente o movimento de rotação da face a partir apenas do formato das mesmas com resultados promissores.

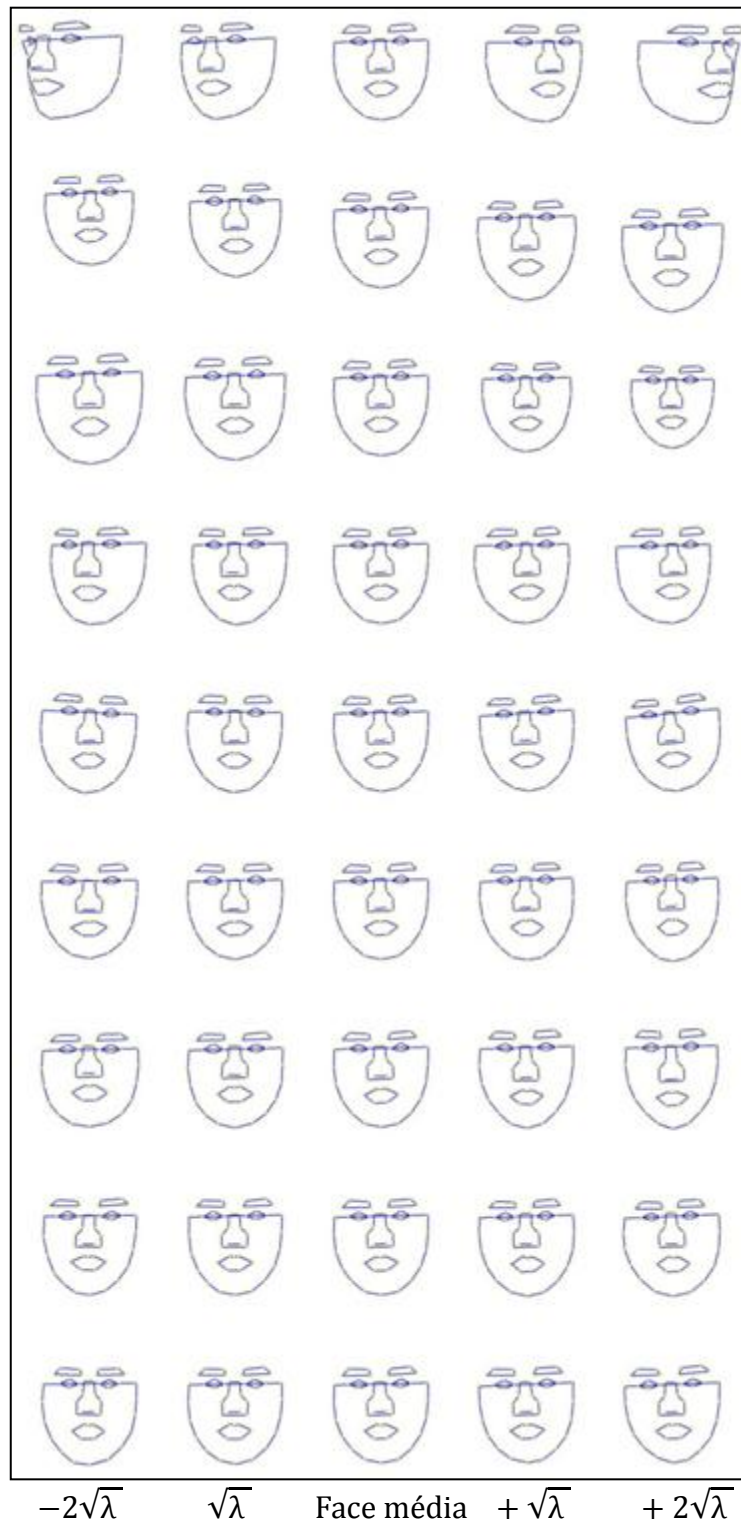


Figura 6-2 - Reconstruções visuais variando-se as nove primeiras componentes principais para as imagens de perfil da base FEI.

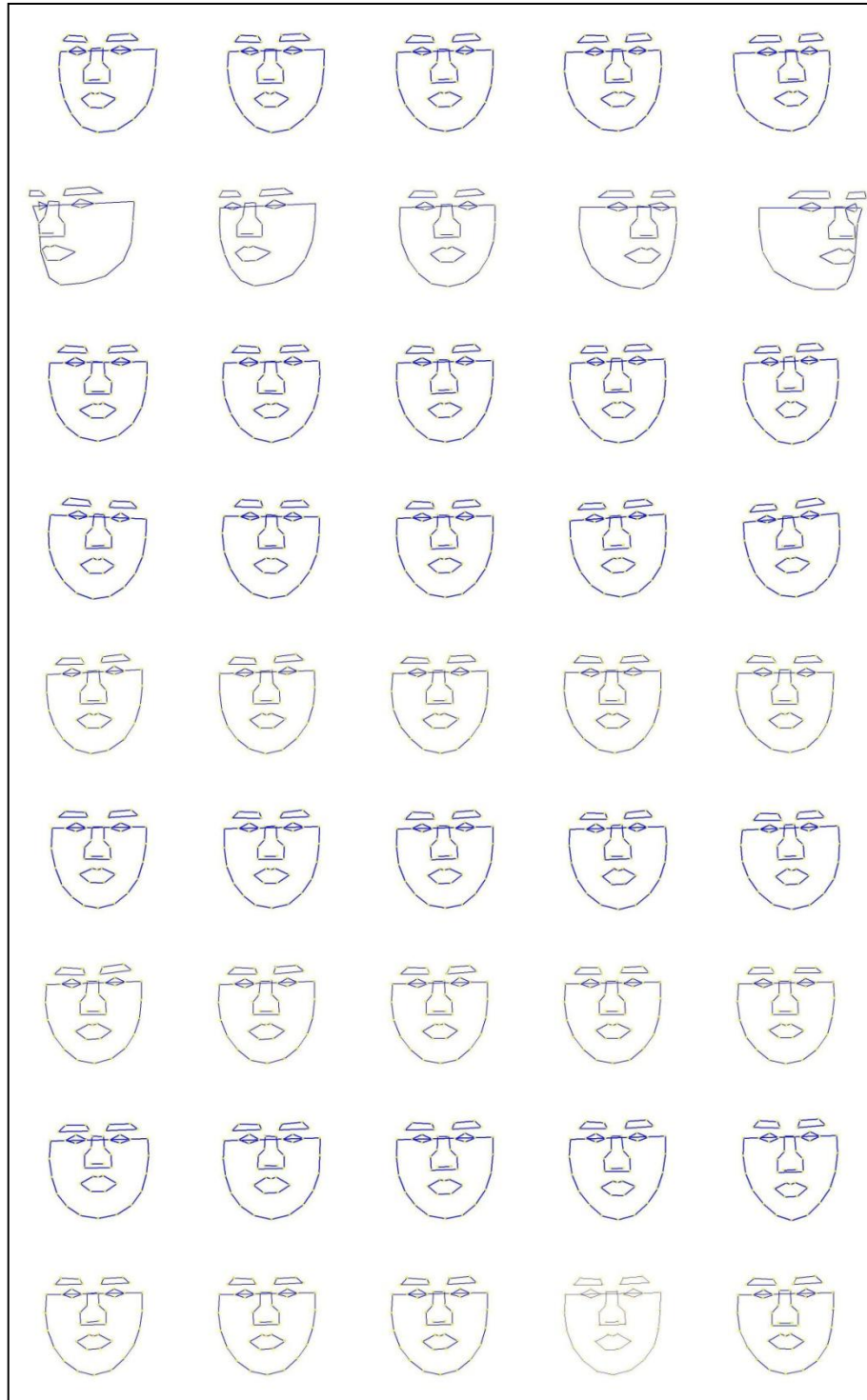


Figura 6-3 – Reconstruções visuais variando-se as nove primeiras componentes principais discriminantes conforme MLDA para as imagens de perfil da base FEI.

Nesta seção foram apresentados alguns resultados preliminares com a intenção de mostrar que é possível estender o trabalho aqui desenvolvido para imagens de perfil.

## REFERÊNCIAS

1. JUNIOR, R. M. C.; COSTA, L. D. F. **Shape Analysis and Classification Theory and Practice**. [S.l.]: CRC Press, 2001.
2. COOTES, T. F. et al. A Trainable Method of Parametric Shape Description. **2nd British Machine Vision Conference, P. Mowforth editor**, p. 54-61, 1991.
3. COOTES, T. F.; TAYLOR, C. J. Active Shape Models. **3rd British Machine Vision Conference, D. Hogg and R. Boyle editors**, p. 266-275, 1992.
4. AHMAD, T.; TAYLOR, C. J.; COOTES, T. F. Tracking and Recognising Hand Gestures using Statistical Shape Models. **Proceedings of the 6th British Machine Vision Conference, BMVA Press**, v. 2, p. 403-412, 1995.
5. EDWARDS, G. J.; COOTES, T. F.; TAYLOR, C. J. Face recognition using active appearance models. **Proceedings of the 5th European Conference Computer Vision**, v. 2, p. 581-695, 1998.
6. THOMAZ, C. E. et al. A multilinear discriminant analysis of 2D frontal face images. **In proceedings of SIBGRAP'09, IEEE CS Press**, Rio de Janeiro, Rio de Janeiro, Brazil, 11th-14th October 2009.
7. THOMAZ, C. E.; GIRALDI, G. A. A new ranking method for principal components analysis and its application to face image analysis. **Image and Vision Computing**, n. 28, p. 902-913, 2010.
8. SWETS, D.; WENG, J. Using Discriminants eigenfeatures for image retrieval. **IEEE Trans. Patterns Anal. Mach Intell.**, p. 831-836, 1996.
9. KITANI, E.; GILLIES, D.; THOMAZ, C. A statistical discriminant model for face interpretation and reconstruction. **In Proceedings of SIBGRAP'06. IEEE CS Press**, p. 247-254, October 2006.
10. THOMAZ, C. E. Estudo de classificadores para o reconhecimento automático de faces. **Dissertação Mestrado em Engenharia Elétrica - PUC**, Rio de Janeiro, 1999.
11. KASS, M.; WITKIN, A.; TERZOPOULOS, D. Snakes: Active Contour Models. **Fisrt International Conference on Computer Vision, IEEE Computer Society Press**, p. 259-268, 1987.
12. YUILLE, A. L.; COHEN, D. S.; HALLINAM, P. Feature Extraction from Faces using Deformable Templates. **Proceedings of Computer Vision**, p. 104-109, 1989.

13. STAIB, L. H.; DUNCAN, J. S. Parametrically Deformable Contour Models. **IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, San Diego, p. 98-103, jun. 1989.
14. TAYLOR, C. J.; COOPER, D. H. Shape Verification Using Belief Updating Proceedings. **British Machine Vision Conference BMVC90 (Oxford)**, p. 61-66, 1990.
15. LOWE, D. W. Fitting Parameterized Three Dimensional Models to Images. **IEEE PAMI**, p. 441-450, 1991.
16. TURK, M.; PENTLAND, A. Eingefaces for Recognition. **Journal of Cognitive Neuroscience, MIT**, v. 3, n. 1, p. 71-86, 1991.
17. FUKUNAGA, K. **Introduction to statistical pattern recognition**. 2. ed. Boston: Academic Press, 1990.
18. BRUNELLI, R.; POGGIO, T. Face Recognition: Features versus Templates. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 15, n. 10, p. 1042-1052, out. 1993.
19. Disponível em:  
<[http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/software/am\\_tools\\_doc/download\\_win.html](http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/software/am_tools_doc/download_win.html)>. Acesso em: 05 out. 2009.
20. LANITIS, A.; TAYLOR, C. J.; COOTES, T. F. An Automatic Identification System Using Flexible Appearance Models. **Image and Vision Computing**, v. 13, p. 393-401, 2005.
21. COOPER, D. H. et al. Active shape models—their training and application. **Computer Vision and Image Understanding**, v. 61, p. 38-59, 1995.
22. PANTIC, M.; VUKADINOVIC, D. Fully Automatic Facial Feature Point Detection Using Gabor Feature Based Boosted Classifiers. **2005 IEEE International Conference on Systems, Man and Cybernetics**, Waikoloa, Hawaii, p. 1692-1698, 10-12 October 2005.
23. PEARSON, K. On lines and planes of closest fit to system of point in space. **Philosophical Magazine** 2, p. 559-572, 1901. disponível em <http://stat.smmu.edu.cn/history/pearson1901.pdf>, acessado em 15/10/2009.
24. SWETS, D.; WENG, J. Using discriminants eigenfeatures for image retrieval. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 18, n. 8, p. 831 - 836, 1996.

25. KIRBY, M.; SIROVICH, L. Low-dimensional procedure for the characterization of human faces. **Journal of Optical Society of America**, v. 4, p. 519-524, March 1987.
26. FISHER, R. A. The use of Multiple Measurements in Taxonomic Problems. **Annals of Eugenics**, n. 7, p. 179-188, 1936.
27. KITTLER, J.; DEVIJVER, P. **Pattern Classification: A Statistical Approach**. [S.l.]: Prentice-Hall.
28. 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. **extended version of the SIBGRAPI'05 paper, Journal of the Brazilian Computer Society**, v. 12, n. 2, p. 7-18, September 2006.
29. JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 5. ed. [S.l.]: Prentice Hall.
30. JAIN, A. K.; CHANDRASEKARAN, B. Dimensionality and sample size considerations in pattern recognition practice. **Handbook of Statistics**, n. 2, p. 835-855, 1982.
31. VAPNIK, V. N.; GUYON, I. M.; BOSER, B. E. A training algorithm for optimal margin classifier. **In Proc. 5th ACM Workshop on Computational Learning Theory**, Pittsburg, PA, p. 144-152, Julho 1992.
32. CORTES, C.; VAPNIK, V. Support vector networks. **Machine Learning**, v. 20, p. 1-25, 1995.
33. VAPNIK, V. **The Nature of Statistical Learning Theory**, Springer, New York, 1995.
34. OSUNA, E.; FREUD, R.; GIROSI, F. Training Support Vector Machines: an Application to Face Detection. **Proceedings of CVPR'97**, Junho 1997.
35. THOMAZ, C. E.; TENORIO, E. Z. Análise multilinear discriminante de formas frontais de imagens 2D de face. **X SBAI - Simpósio Brasileiro de Automação Inteligente**, São João Del Rei - MG, p. 266 - 271, set. 2011.
36. COOK, R. D.; YIN, X. Dimension reduction and visualization in discriminant analysis (with discussion). **Australian and New Zealand Journal of Statistics**, n. 43, p. 147-199, 2001.
37. [HTTP://PERSONALPAGES.MANCHESTER.AC.UK/STAFF/TIMOTHY.F.COOTES/DATA/TARFD\\_MARKUP/TARFD\\_MARKUP.HTML](http://PERSONALPAGES.MANCHESTER.AC.UK/STAFF/TIMOTHY.F.COOTES/DATA/TARFD_MARKUP/TARFD_MARKUP.HTML).. Acesso em: 21 abr. 2010.
38. Disponível em: <<http://www.fei.edu.br/~cet/facedatabase.html>>. Acesso em: 05 set.

- 2009.
39. [HTTP://PERSONALPAGES.MANCHESTER.AC.UK/STAFF/TIMOTHY.F.COOTES/DATA/TARFD\\_MARKUP/TARFD\\_MARKUP.HTML](http://PERSONALPAGES.MANCHESTER.AC.UK/STAFF/TIMOTHY.F.COOTES/DATA/TARFD_MARKUP/TARFD_MARKUP.HTML). Acesso em: 21 abr. 2010.
  40. AMARAL, V.; THOMAZ, C. E. Normalização Espacial de Imagens Frontais de Face. **Relatório Técnico: 2008/01**, Departamento de Engenharia Elétrica, FEI, São Bernardo do Campo, São Paulo, Brasil, 2008.
  41. CAWLEY, G. C.; TALBOT, N. L. C. Efficient Leave-One-Out Cross-Validation of Kernel Fisher Discriminant Classifiers. **Pattern Recognition**, n. 36 (11), p. 2585-2592, 2003.
  42. RAFAELZADEH, P.; TANG, L.; LIU, H. **Cross Validation**. Arizona State University. [S.l.].

**APÊNDICE A – Código do DPCA**



```

function rr = KFOLDPCA(npca,Classif)

% npca - Number of principal components
%classif = 1 (homem_mulher); classif = 2 (sorrindo_ao_sorrindo)

%-----

ng = 2;      % Number of groups
rpath = 'C:\Arquivos\Mestrado\Pontinhos';
ns_g1 = 'group1.smp';
ns_g2 = 'group2.smp';
trein=1;

for trein=1:10
    grupo1_trein=sprintf('%s%d%s','group1_treinamento',trein, '.smp');
    grupo2_trein=sprintf('%s%d%s','group2_treinamento',trein, '.smp');
    grupo1_valid=sprintf('%s%d%s','group1_validacao',trein, '.smp');
    grupo2_valid=sprintf('%s%d%s','group2_validacao',trein, '.smp');

    lpath = sprintf('%s/pontos',rpath);
    if (Classif==1)
        spath = sprintf('%s/base_comparacao/homem_mulher',rpath);
    elseif (Classif==2)
        spath = sprintf('%s/base_comparacao/sorrindo_ao_sorrindo',rpath);
    end;

    s_g1_trein = geramatriz(sprintf('%s/%s',spath,grupo1_trein));
    s_g2_trein = geramatriz(sprintf('%s/%s',spath,grupo2_trein));
    s_g1_valid = geramatriz(sprintf('%s/%s',spath,grupo1_valid));
    s_g2_valid = geramatriz(sprintf('%s/%s',spath,grupo2_valid));

    matriz(1:180,1:92) = s_g1_trein;
    matriz(181:360,1:92) = s_g2_trein;
    matriz_valid(1:20,1:92) = s_g1_valid;
    matriz_valid(21:40,1:92) = s_g2_valid;

    fclose('all');

%-----

nX1=180;
nY1=20;
nX2=180;
nY2=20;
Ny=200;

rr = zeros(1,4);
Rid = zeros((nY1+nY2),4);
Rid(:,1) = [ones(nY1,1);2.*ones(nY2,1)];
Ridsvm = [ones(nX1,1);-ones(nX2,1)];

```

```

media = mean(matriz);
for i = 1:size(matriz,1), matriz(i,:) = matriz(i,:) - media; end

[P,Kpca,Vpca] = pca(matriz,92);
Xpca = matriz * P;

Kstd = Kpca(1:npca);      % most expressive eigenvalues
Pstd = P(:,1:npca);      % most expressive components

Xpca = matriz * Pstd;
Ypca = (matriz_valid - ones(size(matriz_valid,1),1)*media) * Pstd;

[RidX,RidY] = classmah(Kstd,Xpca,Ypca,ng,[nX1 nX2]);
Rid(:,2) = RidY;

%-----

disp('Calculating LDA hyper-plane...');
Xpca = matriz * P;
wlda = lda(Xpca,[nX1 nX2]);

[v,ilda] = sort(abs(wlda));
ilda = flipud(ilda)

Klda = Kpca(ilda(1:npca)); % most discriminant eigenvalues (MLDA)
Plda = P(:,ilda(1:npca)); % most discriminant components (MLDA)

Xpca = matriz * Plda;
Ypca = (matriz_valid - ones(size(matriz_valid,1),1)*media) * Plda;
[RidX,RidY] = classmah(Klda,Xpca,Ypca,ng,[nX1 nX2]);
Rid(:,3) = RidY;

%-----

disp('Calculating MLDA hyper-plane...');
Xpca = matriz * P;
wmlda = mllda(Xpca,ng,[nX1 nX2],ng-1);

[v,imlda] = sort(abs(wmlda));
imlda = flipud(imlda)

Kmllda = Kpca(imlda(1:npca)); % most discriminant eigenvalues (MLDA)
Pmllda = P(:,imlda(1:npca)); % most discriminant components (MLDA)

Xpca = matriz * Pmllda;
Ypca = (matriz_valid - ones(size(matriz_valid,1),1)*media) * Pmllda;
[RidX,RidY] = classmah(Kmllda,Xpca,Ypca,ng,[nX1 nX2]);
Rid(:,4) = RidY;

```

```

%-----
disp('Calculating SVM hyper-plane...');
Xpca = matriz * P;

[nsv,alpha,b0] = svc(Xpca,Ridsvm);
[wc,wr] = svcw(Xpca,Ridsvm,alpha,b0);
wsvm = wr';

[v,isvm] = sort(abs(wsvm));
isvm = flipud(isvm)
Ksvm = Kpca(isvm(1:npc)); % most discriminant eigenvalues (SVM)
Psvm = P(:,isvm(1:npc)); % most discriminant components (SVM)

Xpca = matriz * Psvm;
Ypca = (matriz_valid - ones(size(matriz_valid,1),1)*media) * Psvm;
[RidX,RidY] = classmah(Ksvm,Xpca,Ypca,ng,[nX1 nX2]);
Rid(:,4) = RidY;

rr(1,1) = 1 - sum(Rid(:,1)~=Rid(:,2))/Ny;
rr(1,2) = 1 - sum(Rid(:,1)~=Rid(:,3))/Ny;
rr(1,3) = 1 - sum(Rid(:,1)~=Rid(:,4))/Ny;

end

```

**APÊNDICE B - Artigo do X SBAI**

## ANÁLISE MULTILINEAR DISCRIMINANTE DE FORMAS FRONTAIS DE IMAGENS 2D DE FACE

ELIANA ZACCHI TENORIO\*, CARLOS EDUARDO THOMAZ\*

\* *Centro Universitário da FEI*

*Av. Humberto de Alencar Castelo Branco, 3972 - Assunção  
São Bernardo do Campo, SP, Brasil*

Emails: [eliana@ultrabit.com.br](mailto:eliana@ultrabit.com.br), [cet@fei.edu.br](mailto:cet@fei.edu.br)

**Abstract**— This paper describes and implements a multilinear shape analysis of 2D frontal face images. Our aim is not only to visualize and understand the information that most varies given the sample data, but also to extract the most discriminant differences related to a specific characteristic of interest, such as changes owing to gender or facial expression. The experiments carried out have showed that the multilinear discriminant analysis, although limited by the number of existing supervised classes, is a promising approach to extend well-known methods of extracting shape properties of 2D objects in general.

**Keywords**— Multilinear discriminant analysis, 2D face shapes.

**Resumo**— Este artigo descreve e implementa uma análise multilinear de formas 2D em imagens frontais de face com o objetivo de entender como essas informações se alteram no espaço original dos dados. Investiga-se não somente as componentes principais que mais variam mas também a direção que melhor representa as diferenças discriminantes de acordo com a característica de interesse como, por exemplo, alterações relativas a gênero ou expressão facial. Os experimentos mostram que a análise multilinear discriminante, embora limitada ao número de classes existentes, é promissora e permite a extensão dos métodos existentes na literatura para extração de propriedades das formas de objetos em geral.

**Keywords**— Análise multilinear discriminante, formas 2D de face.

### 1 Introdução

Muitas atividades humanas e interações com o meio ambiente realizadas por máquinas (computadores ou robôs) são guiadas por análises da informação visual disponível (Costa and Júnior, 2001). Entretanto, além dos recursos computacionais necessários para interpretar automaticamente as imagens obtidas, há dificuldades adicionais na análise da informação visual, como a existência de ruído, oclusão, distorção e perda de informação em mapeamentos de objetos do espaço tridimensional (3D) para espaços bidimensionais (2D).

Na maioria das situações, formas 2D em imagens têm as mesmas características dos objetos originais com exceção fundamentalmente da profundidade e do movimento. Objetos representados por formas 2D podem ser prontamente reconhecidos na maioria dos casos. Por exemplo, a forma de um coração, de uma estrela, de uma nota musical, de uma ferramenta e até mesmo o formato da face de uma pessoa, assunto deste trabalho, são facilmente reconhecidos por nós.

Essencialmente, a análise computacional de forma envolve uma série de tarefas importantes desde a aquisição da imagem até a classificação da mesma. Estas tarefas podem ser divididas em três grandes etapas: pré-processamento, extração de características e classificação (Costa and Júnior, 2001). T. Cootes e colegas (Cootes et al., 1991; Cootes and Taylor, 1992; Ahmad et al., 1995; Edwards et al., 1998), interessados em problemas onde os formatos dos objetos po-

dem variar, desenvolveram um método pioneiro de representação de formas por meio do qual pode-se, automaticamente, extrair as propriedades das formas dos dados e descrever sucintamente os parâmetros que mais variam desses dados (Cootes et al., 1991). Os objetos são representados por um conjunto de pontos de controle e examinando-se a variação estatística das coordenadas desses pontos para um conjunto de treinamento estima-se a posição inicial, o tamanho e o formato de um objeto em uma imagem e, conseqüentemente, os ajustes necessários para melhor adequar esse modelo de forma à imagem correspondente (Cootes and Taylor, 1992). Por exemplo, em (Ahmad et al., 1995) os autores descreveram um procedimento para reconhecer gestos e monitorar em uma seqüência de imagens o movimento de uma mão humana quando opera-se um mouse. Para esta aplicação foram definidos 89 pontos de controle para representar a mão e, após a aplicação do método de Análise de Componentes Principais (PCA) (Fukunaga, 1990), o resultado foi um vetor que representava a forma média de uma mão e uma série de direções no espaço original que descrevia as variações das mãos em torno desta forma média. Resultados semelhantes, utilizando-se um número maior de pontos de controle, também foram implementados com sucesso em um modelo de face estatístico baseado em imagens frontais (Edwards et al., 1998; Lanitis et al., 2005).

Este artigo investiga os trabalhos recentes de (Kitani et al., 2006; Kitani and Thomaz, 2007; Thomaz et al., 2009) e descreve uma análise

multilinear de formas 2D em imagens frontais de face com o objetivo de entender como essas informações estão se alterando no espaço original. Especificamente, investiga-se não somente as componentes principais que mais variam (Cootes et al., 1991; Cootes and Taylor, 1992; Ahmad et al., 1995; Edwards et al., 1998), mas também a direção que melhor representa as diferenças discriminantes entre os dados dependendo do contexto da análise multivariada em questão. Pretende-se, portanto, extrair a direção mais discriminante de acordo com a característica de interesse, por exemplo, diferenças de gênero e expressão facial, e não as direções mais expressivas (Swets and Weng, 1996) que não necessariamente representam exclusivamente as diferenças discriminantes (Kitani et al., 2006).

## 2 Análise de Componentes Principais

A Análise de Componentes Principais, ou simplesmente PCA (Principal Component Analysis) (Fukunaga, 1990) como é amplamente conhecida, é um método estatístico multivariado de extração de informação, proposto originalmente por Karl Pearson (Pearson, 1901), que transforma um conjunto  $n$  de variáveis originais em um número  $m$  de combinações lineares destas variáveis, onde  $m \leq n$ , por meio da seleção das características que mais variam (ou mais expressivas (Swets and Weng, 1996)) e conseqüente redução de dimensionalidade dos dados.

Seja uma amostra de treinamento contendo  $N$  formas de face representadas por vetores  $n$ -dimensionais que descrevem as coordenadas 2D dos pontos de controle dessas formas. A matriz de covariância desses dados pode ser escrita como:

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

onde  $\mathbf{x}_i$  e  $\bar{\mathbf{x}}$  são, respectivamente, a forma  $i$  e média global dessa amostra de treinamento. É um resultado provado na literatura que o conjunto  $m$  de autovetores de  $S$ , que corresponde aos maiores autovalores, minimiza o erro quadrático médio de reconstrução desses dados para todas as possíveis escolhas de bases  $m$  vetoriais ortonormais (Fukunaga, 1990). Esse conjunto de autovetores que define um novo sistema de coordenadas para a amostra de treinamento é conhecido como componentes principais e pode ser descrito matematicamente como:

$$P^T S P = \Lambda, \quad (2)$$

onde  $P$  e  $\Lambda$  são as matrizes de autovetores e autovalores de  $S$ . No contexto específico de reconhecimento de faces, a matriz  $P_{pca} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m]$

de componentes principais ordenada decrescentemente ( $\lambda_1 \geq \lambda_2 \dots \geq \lambda_m$ ) é comumente denominada de autofaces (Sirovich and Kirby, 1987; Turk and Pentland, 1991). As  $m$  componentes principais podem então substituir as  $n$  variáveis e a amostra original de dados, que consiste de  $N$  exemplos e  $n$  variáveis, é reduzida para uma amostra de dados contendo  $N$  exemplos e  $m$  componentes principais.

## 3 Análise de Discriminantes Lineares

A Análise de Discriminante Lineares, ou simplesmente LDA (Linear Discriminant Analysis) (Devijver and Kittler, 1982) como é também amplamente conhecida, é um método estatístico multivariado de extração de informação, proposto originalmente por Ronald A. Fisher (Fisher, 1936), que separa em duas ou mais classes de interesse os dados originais. Essa separação em classes ou grupos distintos é feita através de uma transformação linear que maximiza a distância entre as classes e minimiza o espalhamento dentro de cada grupo.

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

$$S_b = \sum_{i=1}^g N_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \quad (3)$$

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

$$S_w = \sum_{i=1}^g \sum_{j=1}^{N_i} (\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i)(\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i)^T \quad (4)$$

onde  $\mathbf{x}_{i,j}$  é a forma  $n$ -dimensional  $j$  da classe  $\pi_i$ ,  $N_i$  é o número de exemplos de treinamento da classe  $\pi_i$ , e  $g$  representa o número total de classes ou grupos. O vetor  $\bar{\mathbf{x}}_i$  é a média amostral da classe  $\pi_i$  (Fukunaga, 1990). O vetor de média global  $\bar{\mathbf{x}}$  pode ser então re-escrito como

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^g N_i \bar{\mathbf{x}}_i = \frac{1}{N} \sum_{i=1}^g \sum_{j=1}^{N_i} \mathbf{x}_{i,j}, \quad (5)$$

onde  $N$  é, como descrito anteriormente, o número total de exemplos de treinamento, ou seja,  $N = N_1 + N_2 + \dots + N_g$ .

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

$$W_{lda} = \arg \max_W \frac{|W^T S_b W|}{|W^T S_w W|}, \quad (6)$$

O critério de Fisher descrito pela equação (6) é satisfeito quando a matriz de projeção  $W_{lda}$  é composta, no máximo, pelos  $(g - 1)$  autovetores de  $S_w^{-1}S_b$ , cujos autovalores correspondentes são não-nulos (Fukunaga, 1990; Devijver and Kittler, 1982). Por exemplo, em problemas envolvendo duas classes somente, a matriz de projeção do LDA contém, na verdade, um único autovetor  $w_{lda}$ , admitindo que  $S_w$  seja não-singular.

No entanto, quando o número total de exemplos  $N$  não é consideravelmente maior que o número de variáveis  $n$ , ocorre uma instabilidade no cálculo da matriz inversa de  $S_w$  (Fukunaga, 1990). A quantidade de exemplos de treinamento para evitar essa instabilidade no cálculo da matriz inversa de  $S_w$  deve ser igual ou superior a 5 vezes a quantidade de variáveis que definem essas amostras (Jain and Chandrasekaran, 1982). Portanto, para o tratamento do problema de instabilidade no cálculo da inversa da matriz  $S_w$ , pode-se utilizar o método denominado MLDA (*Maximum uncertainty Linear Discriminant Analysis*) (Thomaz et al., 2006). Essa técnica consiste em substituir a matriz de espalhamento  $S_w$  por outra matriz regularizada  $S_w^*$ . A nova matriz regularizada  $S_w^*$  pode ser calculada por meio dos seguintes passos:

1. Seleciona-se os autovetores  $\Phi$  e autovalores  $\Lambda$  de  $S_p$ , onde  $S_p = \frac{S_w}{N-g}$ ;
2. Calcula-se a média dos autovalores  $\bar{\lambda}$ , ou seja,

$$\bar{\lambda} = \frac{1}{n} \sum_{j=1}^n \lambda_j = \frac{Tr(S_p)}{n}; \quad (7)$$

3. Constrói-se uma nova matriz de autovalores baseada na dispersão dos maiores autovalores, isto é,

$$\Lambda^* = diag[max(\lambda_1, \bar{\lambda}), \dots, max(\lambda_n, \bar{\lambda})]; \quad (8)$$

4. Calcula-se a matriz de espalhamento intraclasses regularizada

$$S_w^* = S_p^*(N - g) = (\Phi \Lambda^* \Phi^T)(N - g). \quad (9)$$

Com a matriz  $S_w^*$  calculada, substitui-se  $S_w$  da equação (6) por  $S_w^*$  e regulariza-se o critério de Fisher para problemas onde o número total de exemplos de treinamento é limitado e equivalente a dimensionalidade dos dados.

## 4 Experimentos e Resultados

Os experimentos foram desenvolvidos com o objetivo de distinguir gênero e expressão facial em imagens de faces a partir apenas dos formatos das mesmas, sem levar em consideração cor ou textura. Para isto, utilizou-se duas bases de dados disponíveis publicamente para pesquisa: AR

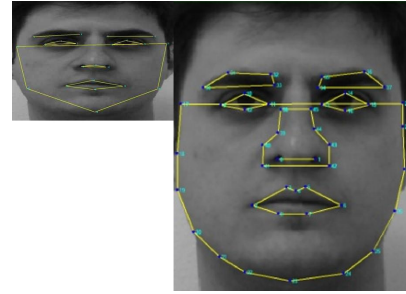


Figura 1: Ilustração da localização dos pontos de controle: 22 pontos da base AR (esquerda); 46 pontos da base FEI (direita).

(Martinez and Benavente, 1998) e FEI (Thomaz and Giraldo, 2010).

Da base AR, utilizou-se 238 imagens frontais, todas com expressão facial sorrindo e não-sorrindo, sendo 148 masculinas e 90 femininas. Os formatos dessas imagens foram descritos por 22 pontos de controle previamente marcados<sup>1</sup>, conforme o exemplo ilustrado na Figura 1.

Para a base de faces da FEI, utilizou-se 400 imagens frontais, todas com expressão facial sorrindo e não-sorrindo, sendo 200 masculinas e 200 femininas. Os formatos dessas imagens foram descritos por 46 pontos de controle, dentre eles 19 dos 20 pontos mais relevantes de características frontais de acordo com (Vukadinovic and Pantic, 2005) e mais 27 pontos dos 152 sugeridos por (Lanitis et al., 2005). Estes 27 pontos foram adicionados para permitir que a ferramenta de referencia na literatura (Lanitis et al., 2005) tivesse uma quantidade de pontos suficiente para gerar uma representação da face por meio de linhas, conforme o exemplo ilustrado na Figura 1.

### 4.1 Reconstrução das Componentes Principais

De forma análoga aos experimentos de (Ahmad et al., 1995; Edwards et al., 1998; Lanitis et al., 2005), pode-se utilizar a forma média global  $\bar{x}$  representada pelo vetor  $n$ -dimensional para interpretar as alterações mais expressivas capturadas pelas componentes principais. Matematicamente, essa navegação nas componentes principais pode ser descrita por

$$\mathbf{y} = \bar{\mathbf{x}} + j\sqrt{\lambda_i} \cdot \mathbf{p}_i, \quad (10)$$

onde  $j \in \{-3, -2, -1, 0, 1, 2, 3\}$  e  $\mathbf{p}_i$  são as  $m$  componentes principais ordenadas decrescentemente pelos seus autovalores correspondentes, isto é,  $\lambda_1 \geq \lambda_2 \dots \geq \lambda_m$ .

As Figuras 2 e 3 mostram essas transformações lineares para as 3 primeiras componentes principais das bases de faces AR e FEI, respectivamente. Devido à característica não-supervisionada da análise de componentes prin-

<sup>1</sup><http://personalpages.manchester.ac.uk/staff/timothy.f.cootes>

cipais, não é possível prever ou garantir qual componente principal irá distinguir as diferenças relativas a gênero ou expressão facial exclusivamente. Navegando nas primeiras componentes, percebe-se as alterações mais expressivas que descrevem, em ambas bases de faces, fundamentalmente diferenças associadas ao tamanho, formato e posição dos rostos. No entanto, informações mais sutis, como diferenças de expressão facial, podem ser capturadas nas componentes principais cujos autovalores não são máximos, conforme pode-se observar na reconstrução da 3a. componente principal da base de faces da FEI, ilustrada na Figura 3.

#### 4.2 Reconstrução dos Discriminantes Lineares

Antes da reconstrução dos discriminantes lineares, estimou-se a habilidade de generalização dos vetores  $\mathbf{w}_{mlda}$  para discriminar gênero e expressão facial dos exemplos de treinamento das bases de dados AR e FEI por meio do método cross-validation e do classificador de distancia Euclideana (Fukunaga, 1990). A Tabela 1 mostra as taxas de classificação obtidas pelos vetores discriminantes correspondentes. Pode-se verificar que em todos os experimentos realizados a taxa de acurácia foi alta, atingindo valores superiores a 85% para ambas características de interesse e ambas bases de dados.

Tabela 1: Taxas de classificação (%) para as características gênero (homem versus mulher) e expressão facial (sorrindo versus não-sorrindo).

Base de Faces	Gênero		Exp. F.	
	h	m	s	n
AR	86.8	87.2	97.5	96.4
FEI	87.9	91.9	92.2	97.1

Para a análise multilinear discriminante, calculou-se o vetor  $\mathbf{w}_{mlda}$  de projeção considerando todos os exemplos de treinamento e as separações dos formatos das faces com relação a gênero e expressão facial. Admitindo que o espalhamento das classes no vetor  $\mathbf{w}_{mlda}$  discriminante seja descrito por uma distribuição Gaussiana e impondo limites para a variação de cada classe, como por exemplo  $\pm 3\sigma_i$  (ou 99.73%), onde  $\sigma_i$  é o desvio padrão de cada classe  $i \in \{1, 2\}$ , a navegação nos discriminantes lineares pode ser calculada matematicamente por

$$\mathbf{y} = \bar{\mathbf{x}} + j\sigma_i \cdot \mathbf{w}_{mlda}, \quad (11)$$

onde  $j \in \{-3, -2, -1, 0, 1, 2, 3\}$ .

As Figuras 4 e 5 apresentam, de cima para baixo, as navegações nos vetores discriminantes de gênero e expressão facial para as bases de dados AR e FEI, respectivamente. Pode-se observar, claramente, que os rostos das imagens femininas

da base AR são menores e mais sorridentes do que os rostos masculinos, e as sobrancelhas possuem simetrias distintas dependendo do gênero. Além disso, a Figura 4 mostra que independentemente do gênero, há diferenças relevantes associadas exclusivamente a expressão facial. Para a base de faces da FEI, os rostos masculinos parecem ser mais finos ou alongados e conter sobrancelhas mais grossas que os rostos femininos, mas tão sorridentes quanto, independentemente do gênero. A reconstrução das alterações mais discriminantes relativas a expressão facial da base de faces da FEI ilustra diferenças somente no formato da boca e na abertura dos olhos, representando exclusivamente as mudanças fundamentais dessa característica de interesse.

## 5 Conclusão

Os modelos descritos e implementados neste trabalho estão baseados em transformações lineares e descrevem no espaço original dos dados as diferenças nas formas frontais de imagens 2D de face. Constatou-se, no entanto, que não é possível prever qual componente principal irá descrever as características discriminantes de interesse, pois essas características podem ser representadas tanto pela primeira como pela última componente principal e de maneira não necessariamente exclusiva. Uma alternativa para essa limitação inerente a métodos não-supervisionados é realizar uma análise discriminante multilinear dos dados, utilizando exemplos já classificados e disponíveis. Os experimentos desse trabalho mostram que tal análise, embora limitada ao número de classes existentes, é promissora e permite a extensão dos métodos existentes na literatura para extração de propriedades das formas de objetos em geral.

## Referências

- Ahmad, T., Taylor, C. J. and Cootes, T. F. (1995). Tracking and recognising hand gestures using statistical shape models, *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 403–412.
- Cootes, T. F., Cooper, D., Taylor, C. J. and Graham, J. (1991). A trainable method of parametric shape description, *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 54–61.
- Cootes, T. F. and Taylor, C. J. (1992). Active shape models, *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 266–275.
- Costa, L. D. and Júnior, R. M. (2001). *Shape Analysis and Classification: Theory and Practice*, CRC Press.





Figura 2: Reconstrução das alterações mais expressivas da base de faces AR. Da esquerda para direita:  $[-3\sqrt{\lambda_i}, -2\sqrt{\lambda_i}, -1\sqrt{\lambda_i}, \bar{x}, +1\sqrt{\lambda_i}, +2\sqrt{\lambda_i}, +3\sqrt{\lambda_i}]$ , onde  $i = \{1, 2, 3\}$  de cima para baixo.

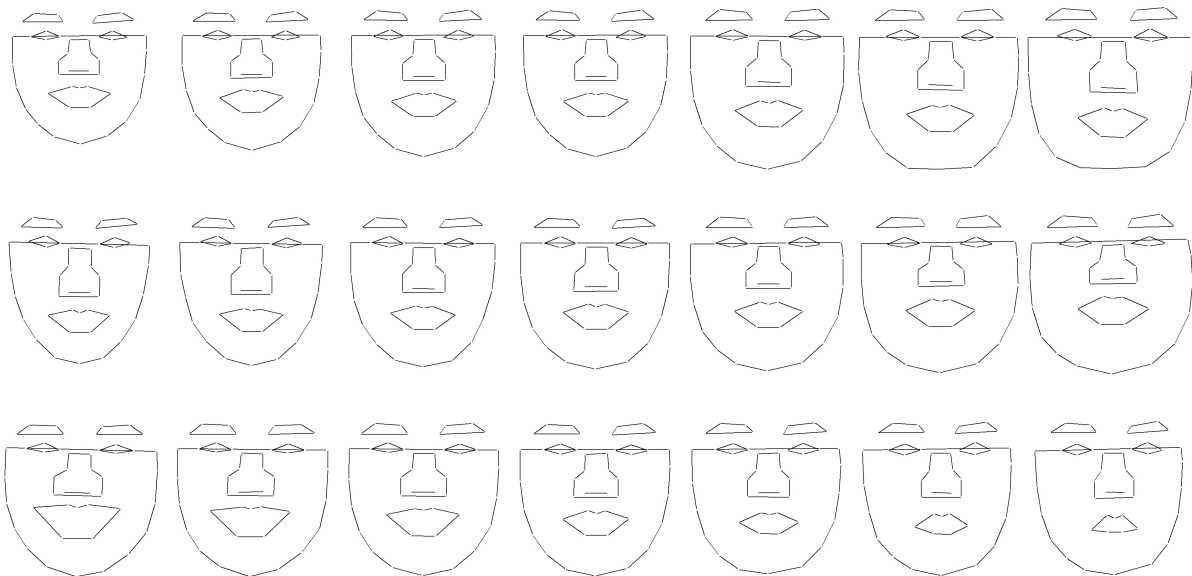


Figura 3: Reconstrução das alterações mais expressivas da base de faces da FEI. Da esquerda para direita:  $[-3\sqrt{\lambda_i}, -2\sqrt{\lambda_i}, -1\sqrt{\lambda_i}, \bar{x}, +1\sqrt{\lambda_i}, +2\sqrt{\lambda_i}, +3\sqrt{\lambda_i}]$ , onde  $i = \{1, 2, 3\}$  de cima para baixo.

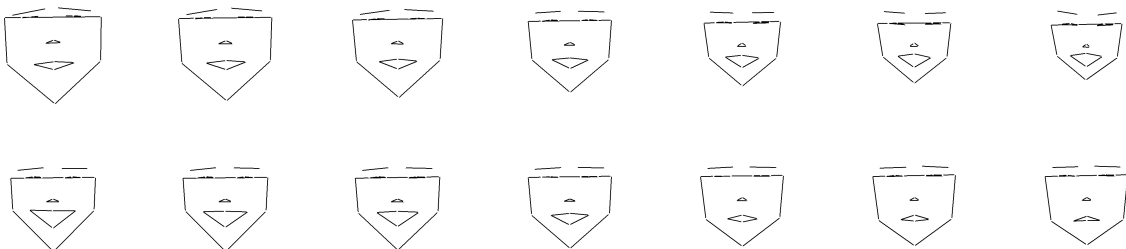


Figura 4: Reconstrução das alterações mais discriminantes relativas a gênero (em cima) e expressão facial (embaixo) da base de faces AR. Da esquerda (classe 1 de exemplos de 'homem' ou 'sorrindo') para direita (classe 2 de exemplos de 'mulher' ou 'não-sorrindo'):  $[-3\sigma_1, -2\sigma_1, -1\sigma_1, \bar{x}, +1\sigma_2, +2\sigma_2, +3\sigma_2]$ .

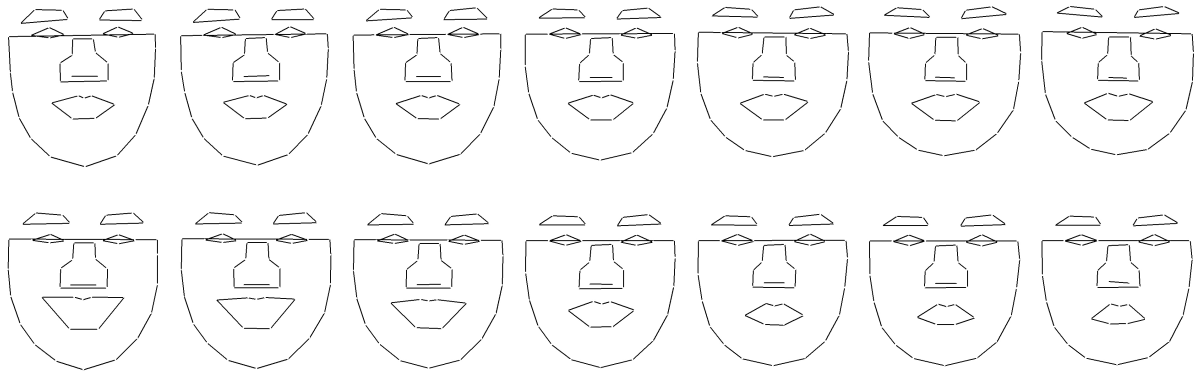


Figura 5: Reconstrução das alterações mais discriminantes relativas a gênero (em cima) e expressão facial (embaixo) da base de faces da FEI. Da esquerda (classe 1 de exemplos de 'homem' ou 'sorrindo') para direita (classe 2 de exemplos de 'mulher' ou 'não-sorrindo'):  $[-3\sigma_1, -2\sigma_1, -1\sigma_1, \bar{x}, +1\sigma_2, +2\sigma_2, +3\sigma_2]$ .

- Devijver, P. and Kittler, J. (1982). *Pattern Classification: A Statistical Approach*, Prentice-Hall.
- Edwards, G. J., Cootes, T. F. and Taylor, C. J. (1998). Face recognition using active appearance models, *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 581–698.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems, *Annals of Eugenics* **7**: 179–188.
- Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*, Morgan Kaufmann.
- Jain, A. K. and Chandrasekaran, B. (1982). Dimensionality and sample size considerations in pattern recognition practice, *Handbook of Statistics* **2**: 835–855.
- Kitani, E. C. and Thomaz, C. E. (2007). Análise de discriminantes lineares para modelagem e reconstrução de imagens de face, *Proceedings of the Encontro Nacional de Inteligência Artificial (ENIA)*, pp. 962–971.
- Kitani, E. C., Thomaz, C. E. and Gillies, D. F. (2006). A statistical discriminant model for face interpretation and reconstruction, *Proceedings of the Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, pp. 247–254.
- Lanitis, A., Taylor, C. J. and Cootes, T. F. (2005). An automatic identification system using flexible appearance models, *Image and Vision Computing* **13**: 393–401.
- Martinez, A. and Benavente, R. (1998). The ar face database. CVC Technical Report no. 24.
- Pearson, K. (1901). On lines and planes of closest fit to system of point in space, *Philosophical Magazine* **2**: 559–572.
- Sirovich, L. and Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces, *Journal of Optical Society of America* **4**: 519–524.
- Swets, D. and Weng, J. (1996). Using discriminants eigenfeatures for image retrieval, *IEEE Trans. Pattern Anal. Mach. Intell.* **18(8)**: 831–836.
- Thomaz, C. E., Amaral, V., Giraldo, G. A., Kitani, E. C., Sato, J. R. and Gillies, D. F. (2009). A multi-linear discriminant analysis of 2d frontal face images, *Proceedings of the Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, pp. 216–223.
- Thomaz, C. E. and Giraldo, G. A. (2010). A new ranking method for principal components analysis and its application to face image analysis, *Image and Vision Computing* **28(6)**: 902–913.
- Thomaz, C., Kitani, E. and Gillies, D. (2006). A maximum uncertainty lda-based approach for limited sample size problems - with application to face recognition, *Journal of the Brazilian Computer Society* **12(2)**: 7–18.
- Turk, M. and Pentland, A. (1991). Eigenfaces for recognition, *Journal of Cognitive Neuroscience* **3**: 71–86.
- Vukadinovic, D. and Pantic, M. (2005). Fully automatic facial feature point detection using gabor feature based boosted classifiers, *Proceedings of the International Conference on Systems, Man and Cybernetics*, pp. 1692–1698.