UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE MATEMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

FRANCINEI GOMES DE MORAIS

RECONHECIMENTO FACIAL: ESTUDO E PROPOSIÇÃO DE NOVAS ABORDAGENS

RIO DE JANEIRO
2012

FRANCINEI GOMES DE MORAIS

RECONHECIMENTO FACIAL: ESTUDO E PROPOSIÇÃO DE NOVAS ABORDAGENS

Dissertação de Mestrado apresentada ao programa de pós-graduação em informática – Instituto de Matemática da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários a obtenção do título de Mestre em Informática.

Adriano Joaquim de Oliveira Cruz – Ph. D.
Orientador

Antonio Carlos Gay Thomé – Ph. D.
Co-orientador

RIO DE JANEIRO
2012

Gomes Morais, Francinei

Reconhecimento Facial: Estudo e proposição de novas abordagens/ Francinei Gomes de Morais. – Rio de Janeiro, 2012.

135f.: il.;

Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro, Instituto de Matemática, Programa de Pós-Graduação em Informática, Rio de Janeiro, 2012.

Orientador: Adriano Joaquim de Oliveira Cruz; Co-orientador: Antonio Carlos Gay Thomé.

1. Processamento de Imagem. 2. Sistemas Biométricos. 3. Análise de Padrão. 4. Reconhecimento facial – Teses.
I. Oliveira Cruz, Adriano Joaquim (Orient.). II. Universidade Federal do Rio de Janeiro. Instituto de Matemática. Programa de Pós-Graduação em Informática. III. Título

FRANCINEI GOMES DE MORAIS

RECONHECIMENTO FACIAL: ESTUDO E PROPOSIÇÃO DE NOVAS ABORDAGENS

Dissertação de Mestrado apresentada ao programa de pós-graduação em informática – Instituto de Matemática da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários a obtenção do título de Mestre em Informática.

Aprovado por

_____

Adriano Joaquim de Oliveira Cruz – Ph. D. – IM/PPGI/UFRJ - Orientador

_____

Antonio Carlos Gay Thomé – Ph. D. – IM/PPGI/UFRJ – Co-orientador

_____

Josefino Cabral Melo Lima – Ph.D. – IM/PPGI/UFRJ

_____

José Antônio dos Santos Borges – D. SC. – iNCE/UFRJ

_____

Marley Maria Bernardes Rebuzzi Vellasco – Ph.D. – PUC

DEDICATÓRIA

Dedico este trabalho primeiramente aos meus pais, Francisco e Inês por todos os esforços que fizeram para que eu pudesse prolongar meus estudos e tivesse mais oportunidades, pois acredito que a educação é o maior bem que um pai pode dar a um filho. Dedico também à minha irmã por todo orgulho que percebo em seu rosto sempre que fala sobre mim. Por fim, dedico especialmente à minha namorada por ter compreendido os momentos em que fiquei ausente e também por ter acreditado em minha capacidade mesmo nos momentos em que eu mesmo duvidei.

AGRADECIMENTOS

Agradeço a Deus por estar sempre ao meu lado e ter posto em minha vida pessoas maravilhosas que me ajudaram a colocar mais um tijolo no muro do conhecimento. Essas pessoas são meus professores, orientadores, funcionários do PPGI, meus amigos de mestrado e do Laboratório de Projetos. Agradeço também a CAPES pelo apoio financeiro, pois sem este provavelmente não teria conseguido concluir o mestrado, dado a necessidade de trabalhar para complementar a renda.

RESUMO

Muitos são os esforços no desenvolvimento de métodos de reconhecimento facial capazes de fornecer uma modelagem robusta mesmo para conjuntos de imagens obtidas em situações não controladas. Alguns dos grandes desafios encontrados são: a variação de luminosidade, pose, expressão facial, rotação, oclusão de partes da face e conjuntos de treinamentos com poucas amostras por indivíduo. Resultados mais robustos têm sido alcançados combinando-se as técnicas: Transformada *Census*, *Gabor Wavelets*, Análise de Componentes Principais (PCA) e Preservação das Projeções Locais (LPP).

Neste trabalho são apresentados um estudo comparativo e a proposição de novas abordagens para a tarefa de reconhecimento facial. Expansão, CT-Eigenblocos, CT-Eigen, Geométrico-LPP e são os métodos aqui propostos, os quais visam contornar problemas como a não linearidade do espaço de faces, imagens com parte da face encoberta e conjuntos de treinamento com poucas amostras por indivíduo. Os métodos de Correlação, *Eigenfaces*, *Fisherfaces* e *Laplacianfaces* foram tomados como referências de comparação. Os experimentos realizados mostram que as abordagens sugeridas apresentaram taxas de erros próximas e em muitos casos inferiores às técnicas de Correlação, Eigenfaces, Fisherfaces e Laplacianfaces para testes utilizando os bancos de imagem Yale, ORL e PIE.

Palavras chave — Gabor Wavelet; Análise de Componentes Principais; Preservação das Projeções Locais; Reconhecimento facial; Transformada Census.

ABSTRACT

There are many efforts in developing face recognition methods capable to provide a robust model even for those image sets collected in uncontrolled situations. Some of the challenges encountered are: variations of illumination, poses, facial expressions and rotation; occlusion of face parts and sets with few training samples per individual. More robust results have been achieved by combining the techniques: Census Transform, Gabor Wavelet, Principal Components Analysis and Local Preserving Projection.

In this work we presents a comparative study and propose new approaches to face recognition. Expansão, CT-Egenblocos, CT-Eigen and Geométrico-LPP are the proposed methods. They aim to overcome problems such as non-linearity of the face space, face images with covered parts and training sets with few samples per individual. Correlation, Eigenfaces, Fisherfaces and Laplacianfaces were used as reference for comparison. The experiments show that the proposed approaches had next error rates and in many cases lower than the techniques Correlation, Eigenfaces, Fisherfaces and Laplacianfaces for tests using Yale, ORL and PIE face data sets.

Keywords — Census Transform, Gabor Wavelet, Principal Components Analysis and Local Preserving Projection; Face Recognition.

LISTA DE FIGURAS

Figura 2.1: Representação gráfica do campo eletromagnético. 23

Figura 2.2: Campo elétrico. 24

Figura 2.3: O espectro eletromagnético. 25

Figura 2.4: O espectro da luz visível. 25

Figura 2.5: Representação de um pixel de um display colorido. Nota-se que cada pixel é composto por três "micro lâmpadas", nas cores vermelho, verde e azul. O tamanho do pixel diminui quanto maior for a resolução do display. 28

Figura 2.6: Matriz de fósforos. 28

Figura 2.7: O espaço de cor RGB. 30

Figura 2.8: O espaço de cor CMYK. 31

Figura 2.9: O espaço de cor CMY. 31

Figura 2.10: O espaço de cor RYB 32

Figura 2.11: Diagrama de cromaticidade CIExyY. 33

Figura 2.12: Espaço de cores HSL. 35

Figura 2.13 O espaço de cor RGB. 35

Figura 2.14: O espaço de cor HSV (Matiz, Saturação e Valor) 36

Figura 2.15: Imagem digital (GONZALEZ e WOODS, 2000). 36

Figura 2.16: Imagem R, G, B e a decomposição em matrizes apresentada em (MOUTINHO, 2005). 37

Figura 2.17: Imagem em tons de cinza e sua matriz característica apresentada em (MOUTINHO, 2005). 37

Figura 2.18: As imagens de (a)-(d) correspondem a uma imagem com e seus respectivos histogramas. 40

Figura 2.19: Imagem Original 42

Figura 2.20: Imagem Equalizada.....42

Figura 3.1: Diagrama de dispersão.....49

Figura 3.2: Formato dos relacionamentos, linear, exponencial e quadrático.....50

Figura 3.3: Covariância Positiva.....51

Figura 3.4: Covariância Negativa.....51

Figura 3.5: Covariância Nula.....52

Figura 3.6: Tipos de correlações.....53

Figura 3.7: (a) Dilatação, (b) Contração, (c) Mudança de direção. A dilatação ocorre quando ($\lambda > 1$), para contração ($0 < \lambda < 1$) e por fim a mudança de direção é alcançada quando ($\lambda < 0$).....54

Figura 3.8: Visualização no R^2.....55

Figura 3.9: Visualização no R^2 com autovetores. A forma elíptica representa a existência de relação de dependência entre as variáveis. Os comprimentos dos eixos principal e secundário da elipse são dados pelos autovalores calculados na equação 3.14.....56

Figura 3.10: Autovalores Idênticos. A forma circular representa a não existência de dependência linear, em outras palavras o sistema de equações é linearmente independente, sendo assim as observações A e B não carregam informações redundantes.....57

Figura 3.11: Autovalor Nulo. A forma linear representa forte dependência entra as variáveis analisadas, em outras palavras, as observações carregam informações redundantes.....57

Figura 3.12: Diagrama de Taxonomia das técnicas de redução de dimensionalidade. Imagem adaptada de (MAATEN e POSTMA, 2009).....62

Figura 3.13: Parte da imagem de uma impressão digital. A figura à esquerda corresponde a um exemplo de término de linha, já a figura da direita apresenta uma bifurcação (MOUTINO, 2005). 66

Figura 3.14: Construção da matriz de transformação. O diagrama apresentado mostra de maneira simplificada as etapas necessárias para a construção da matriz de transformação dado uma base de imagens para treinamento. 68

Figura 3.15: Construção da base de referência. A figura em questão apresenta os módulos necessários para construção da base de referência, esta base é que será posteriormente utilizada pelos classificadores de padrões. 68

Figura 3.16: Realização do reconhecimento. O diagrama apresentado em tal figura modela todo o processo de reconhecimento/ identificação de um indivíduo. 69

Figura 4.1: Imagem (P) em RGB e sua correspondente (I) em tons de cinza. 72

Figura 4.2: Vetor Linha (Γ^T). 72

Figura 4.3: Medidas geométricas e templates. As regiões retangulares correspondem aos templates comumente extraídos e os segmentos de reta às medidas geométricas. 90

Figura 5.1: Exemplo de imagens originais do banco Yale. 96

Figura 5.2: Exemplo de imagens normalizadas do banco Yale. 96

Figura 5.3: Exemplo de imagens originais do banco ORL. 97

Figura 5.4: Exemplo de imagens normalizadas do banco ORL. 97

Figura 5.5: Exemplo de imagens originais do banco PIE. 98

Figura 5.6: Exemplo de imagens normalizadas do banco PIE. 98

Figura 5.7: Conjunto Yale Original tomado em três resoluções diferentes. 100

Figura 5.8: Taxa de erro versus aumento da dimensionalidade. 100

Figura 5.9: Conjunto Yale Recortado tomado em três resoluções diferentes. 101

Figura 5.10: Taxa de erro versus aumento da dimensionalidade. 101

Figura 5.11: Conjunto ORL Recortado dividido nos subconjuntos contendo 10, 20, 30 ou 40 indivíduos a serem reconhecidos. 103

Figura 5.12: Conjunto ORL Recortado dividido nos subconjuntos contendo 10. 20. 30 ou 40 indivíduos a serem reconhecidos. 103

Figura 5.13: Taxa de erro versus o aumento da dimensionalidade. 104

Figura 5.14: Conjunto PIE Recortado dividido em diferentes conjuntos de treinamento e teste. 105

Figura 5.15: Taxa de erro versus o aumento da dimensionalidade. 105

Figura 5.16: Conjunto PIE Recortado dividido em diferentes conjuntos de treinamento e teste. 106

Figura 5.17: Etapas do CT – Eigen. 107

Figura 5.18: Subconjunto de possíveis Kernels. 109

Figura 5.19: Exemplo de invariância à iluminação. Embora a iluminação varie significativamente nas imagens originais, as imagens resultantes da Transformada Census permanecem bastante similares. 109

Figura 5.20: Etapas do CT-Eigenblocos. 110

Figura 5.21: Etapas para construção da base de referência. 111

Figura 5.22: Etapas do processo de reconhecimento. 112

Figura 5.23: Resposta impulsional do filtro de Gabor bidimensional direcional e real. 113

Figura 5.24: Parte real da resposta impulsional do filtro de Gabor bidimensional direcional e imaginário. 114

Figura 5.25: Parte imaginária da resposta impulsional do filtro de Gabor bidimensional direcional e imaginário. 115

Figura 5.26: Diagrama de bloco de um sistema. 115

Figura 5.27: Interface para rotação e recorte da face. 116

Figura 5.28: Interface para marcação da características geométricas. 117

Figura 5.29: Vetor de características geométricas. 118

Figura 5.30: Etapas do algoritmo Geométrico – LPP. 118

Figura 6.1: Mapa de experimentos. 119

Figura 6.2: Conjunto Yale Recortado tomado em quatro resoluções diferentes. 120

Figura 6.3: Conjunto ORL Recortado dividido nos subconjuntos contendo 10. 20. 30 ou 40 indivíduos a serem reconhecidos. 121

Figura 6.4: Conjunto ORL Recortado dividido nos subconjuntos contendo 10. 20. 30 ou 40 indivíduos a serem reconhecidos. 121

Figura 6.5: Conjunto PIE Recortado dividido em diferentes conjuntos de treinamento e teste. 122

Figura 6.6: Conjunto PIE Recortado dividido em diferentes conjuntos de treinamento e teste. 122

Figura 6.7: Imagem original e seus quadrantes encobertos. 123

Figura 6.8: Conjunto Yale Recortado. Experimento considerando quadrantes encobertos. .. 124

Figura 6.9: Subconjunto ORL Recortado, contendo 20 indivíduos a serem reconhecidos. Experimento considerando quadrantes encobertos. 124

Figura 6.10: Subconjunto ORL Recortado, contendo 30 indivíduos a serem reconhecidos. Experimento considerando quadrantes encobertos. 125

Figura 6.11: Conjunto PIE Recortado, conjunto de treinamento e teste (12,12). Experimento considerando quadrantes encobertos. 125

Figura 6.12: Conjunto PIE Recortado, conjunto de treinamento e teste (12,12). Experimento considerando quadrantes encobertos. 126

Figura 6.13: Conjunto PIE recortado, dividido em diferentes grupos de treinamento e teste. Experimento considerando quadrantes encobertos..... 127

Figura 6.14: Conjunto ORL Recortado, considerando conjuntos contendo 10, 20, 30 ou 40 indivíduos a serem reconhecidos. 128

LISTA DE TABELAS

Tabela 3.1: Tabela representativa de vendas e resultados alcançados. 49

Tabela 3.2: Categorização de técnicas de reconhecimento facial (ZHAO e CHELLAPA, 2003). 70

LISTA DE ABREVIATURAS E SIGLAS

PCA	Principal Components Analysis
LDA	Linear Discriminant Analysis
LFA	Local Feature Analysis
LPP	Local Preserving Projection
Yale	Banco de Imagens Faciais da Universidade de Yale
ORL	Banco de Imagens Faciais da Microsoft
PIE	Banco de Imagens Faciais da Universidade de Carnegie Mellon
Eigenfaces	Método de Reconhecimento Facial baseado em PCA
Fisherfaces	Método de Reconhecimento Facial baseado no discriminante de Fisher
Laplacianfaces	Método de reconhecimento facial baseado em LPP

SUMÁRIO

1 Introdução 20
1.1 Motivações 21
1.2 Proposta 21
1.3 Organização da Dissertação 22
2 Fundamentação Teórica 23
2.1 Ondas eletromagnéticas 23
2.2 A formação da cor 26
2.2.1 As cores primárias e secundárias 27
2.3 Os espaços de cores 29
2.4 Representação de imagens digitais 34
2.4.1 A imagem digital 36
2.5 Equalização de histograma 39
3 Transformações matriciais, análise de dados e reconhecimento facial 43
3.1 Transformações matriciais 43
3.1.1 Translação 44
3.1.2 Mudança de escala 44
3.1.3 Rotação 45
3.2 Técnicas fundamentais de análise dados 47
3.2.1 Covariância 47
3.2.2 Correlação 52
3.2.3 Autovalor e autovetor 54
3.3 Técnicas de redução de dimensionalidade 58
3.3.1 PCA (Principal Components Analysis) 58
3.3.2 Técnicas convexas, não convexas e suas ramificações 62
3.4 Sistemas biométricos 63
3.4.1 Reconhecimento de Iris 64
3.4.2 Reconhecimento de assinaturas 65
3.4.3 Reconhecimento de voz 66
3.4.4 Reconhecimento de Impressões digitais 66
3.5 O reconhecimento Facial 67
4 Os métodos de reconhecimento facial 71
4.1 Os métodos baseados em aparência (métodos holísticos) 71

4.1.1 Eigenfaces 71
4.1.1.1 O algoritmo 72
4.1.1.2 Interpretando as equações 74
4.1.1.3 Construção da base final de referência 76
4.1.1.4 O reconhecimento 77
4.1.2 Fisherfaces 79
4.1.2.1 O algoritmo 79
4.1.2.2 Interpretando as equações 82
4.1.2.3 A base de referência 83
4.1.2.4 O reconhecimento 84
4.1.3 Laplacianfaces 84
4.1.3.1 O algoritmo 86
4.1.3.2 Interpretando as equações 88
4.1.3.3 A base de referência 88
4.1.3.4 O reconhecimento 88
4.2 Métodos baseados em características 89
4.2.1 Características Geométricas 89
4.2.2 Métodos baseados em pontos fiduciais 91
4.2.2.1 O algoritmo 91
4.3 Os métodos baseados em templates 94
4.4 Os métodos Híbridos 94
5 Abordagens propostas nesta dissertação 95
5.1 Bases de imagens utilizadas 95
5.2 Experimentos 99
5.3 Análise comparativa 106
5.3.1 Variantes sugeridas 107
5.4 CT-Eigen 107
5.4.1 Transformada Census 108
5.5 CT-Eigenblocos 110
5.6 Expansão 111
5.6.1 Transformada de Gabor 112
5.7 Geométrico – LPP 116
6 Experimentos e resultados 119
6.1 Experimentos 120

6.2 Análise comparativa 129
7 Conclusões e Trabalhos futuros 130
7.1 Trabalhos Futuros 130
8 REFERÊNCIAS BIBLIOGRÁFICAS 132

1 Introdução

O trabalho a ser apresentado trata do ramo de pesquisa conhecido como Reconhecimento de Padrões Faciais. É uma importante linha de estudos que surge como conseqüência direta da aplicação das técnicas de Análise de Padrões, com o objetivo de criar sistemas capazes de reconhecer automaticamente um indivíduo a partir de uma imagem estática ou de um trecho de vídeo.

Nos últimos anos as pesquisas sobre métodos de reconhecimento facial têm recebido uma maior atenção. Os primeiros sistemas completos datam dos anos 60 com o trabalho de (BLEDSOE, 1966) o qual desenvolveu um método baseado em distância geométrica entre pontos característicos da face. Desde então muito já foi desenvolvido, contudo o assunto permanece desafiador no que se refere à obtenção de sistemas capazes de funcionar com robustez e em tempo real. Diversos algoritmos e suas variantes estão sendo propostos por diversas correntes de pesquisadores e, de uma forma resumida, as abordagens adotadas podem ser classificadas em três categorias: Baseadas em Casamento de Template, baseadas em Aparência e as baseadas em Características Locais.

No casamento de template uma face ou partes da mesma são representadas por um ou mais templates, sendo que o reconhecimento é obtido verificando qual foi a face em que ocorreu o maior número de casamento entre os templates (BRUNELLI e POGGIO, 1993), a desvantagem de tal método é o grande consumo de memória. Na abordagem por aparência, a imagem é transformada em um vetor que por sua vez é visto como um ponto em um espaço de alta dimensionalidade (SEUNG e LEE, 2000). O objetivo é representar o vetor em um espaço de dimensão reduzida, onde as imagens de pessoas diferentes formem grupos distintos, facilitando assim a tarefa de classificação. A técnica de *Principal Components Analysis* (PCA) (TURK e PENTLAND, 1991) e a de *Linear Discriminant Analysis* (LDA) (BELHUMEUR e HESPANHA et al., 1997) são as duas mais populares para tal propósito. Pesquisas mostram que mudanças na iluminação e na expressão facial têm grande impacto sobre a não linearidade do espaço de faces e que diferentes regiões da face possuem diferente nível de importância para o reconhecimento. Em (XIAOFEI e NIYOGI et al., 2005) foi proposto um método eficiente para tratar a não linearidade do espaço de faces, que foi chamado de *Laplacianfaces*, baseado na técnica de *Local Preserving Projection* (LPP).

O trabalho de (WISKOTT et al., 1997), chamado de *Elastic Bunch Graph Matching* (EBGM) , exerceu grande influência sobre a abordagem baseada em características locais. Tal estudo representa faces como sendo um grafo, com os nós posicionados em pontos

característicos (pontos fiduciais), por exemplo, olhos, boca, nariz. As arestas do grafo são rotuladas com a distância entre os nós, sendo cada nó representado por um conjunto de características extraídas via Gabor Wavelet (COHEN e KOVACEVIC, 1996), com diferentes orientações e escalas, as quais são chamadas de "jets". Dado um conjunto de imagens para treinamento é construído um conjunto de grafos. O processo de reconhecimento procura o grafo que melhor se ajuste à imagem a ser reconhecida. A desvantagem desta abordagem está na dificuldade de selecionar automaticamente os pontos característicos.

1.1 Motivações

A demanda da sociedade por soluções nas áreas de segurança, entretenimento e acessibilidade são as principais motivações para o estudo de métodos de reconhecimento facial.

Outros fatores são o crescimento do poder computacional e o barateamento do hardware, que possibilitam submeter os algoritmos de reconhecimento facial a testes mais robustos, como por exemplo, analisar o comportamento dos mesmos frente a imagens com maior resolução, com partes da face encoberta e a banco de dados contendo um grande número de indivíduos a serem reconhecidos. Essa possibilidade de investigação mais profunda abre caminho para a proposição de ajustes ou de novas abordagens.

1.2 Proposta

Este trabalho tem por objetivos realizar um estudo dos métodos holísticos de reconhecimento facial bem como propor novas abordagens. As abordagens aqui propostas visam contornar problemas como a não linearidade do espaço de faces, imagens com parte da face encoberta e conjuntos de treinamento com poucas amostras por indivíduo. Os métodos propostos foram construídos através da combinação das técnicas de Transformada Census, Gabor Wavelets, *Principal Components Analysis* e *Local Preserving Projection*. A primeira minimiza as variações de luminosidade, parâmetro que exerce forte influência na não linearidade do subespaço de faces. A segunda é utilizada para ressaltar características locais da face, sendo aplicada com o enfoque de filtro direcional. A terceira é uma transformação muito utilizada em análise de padrões e visa descobrir um subespaço que maximize a variância do espaço amostral. A última é transformação de dimensionalidade capaz de preservar a estrutura do subespaço de faces, em outras palavras, o subespaço de ordem inferior terá suas distâncias relativas preservadas.

Por fim, pretendemos contribuir para a expansão das pesquisas apresentadas por (MOUTINHO, 2005), a qual trata sobre localização facial e sugere como trabalhos futuros o estudo e a implementação da etapa do reconhecimento facial.

1.3 Organização da Dissertação

O restante do trabalho está estruturado da seguinte maneira. O capítulo 2 apresenta a fundamentação teórica das técnicas de processamento de imagens empregadas ao longo da dissertação. O capítulo 3 apresenta a fundamentação teórica das técnicas de análise de padrões bem como um estudo sobre os métodos biométricos de reconhecimento. O capítulo 4 trata dos métodos de reconhecimento facial tomados como referência. O capítulo 5 é iniciado com uma análise comparativa entre os métodos Correlação, Eigenfaces, Fisherfaces e Laplacianfaces e é finalizado com a proposição de novas abordagens. Os experimentos realizados com as novas abordagens bem com os seus resultados são descritos ao longo do capítulo 6. Por fim, as conclusões e os trabalhos futuros são apresentados no capítulo 7.

2 Fundamentação Teórica

Neste capítulo será apresentada uma breve fundamentação teórica sobre a teoria das cores, o processo de formação da imagem, a representação de imagens digitais, e por fim uma técnica de realce de imagem conhecida como equalização de histograma.

Buscando apresentação do conteúdo de forma didática a seqüência seguida foi ondas eletromagnéticas, a formação da cor, luz visível, os espaços de cores, a imagem digital e equalização de histograma.

2.1 Ondas eletromagnéticas

Fisicamente descrevendo, ondas eletromagnéticas são pulsos de energia que se propagam através do espaço, ou seja, não dependem de meio físico como, líquidos, sólidos e gases. Graficamente são representados pela combinação de um campo elétrico e um campo magnético ortogonais entre si, figura 2.1.

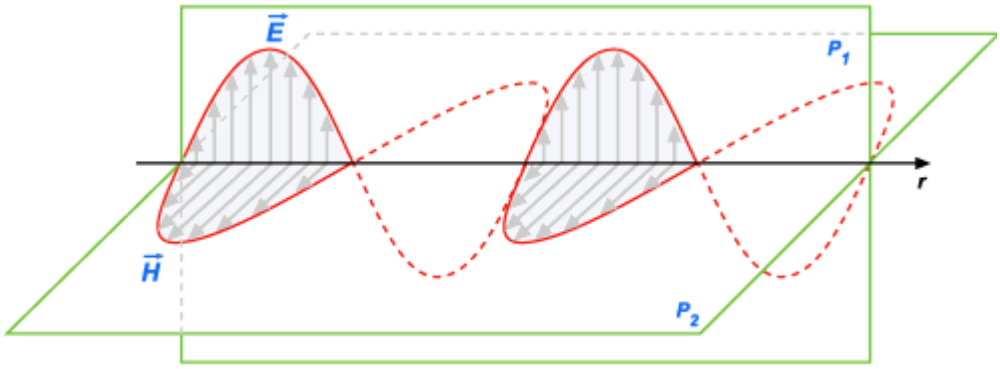


Figura 2.1: Representação gráfica do campo eletromagnético.

As ondas de rádio, celular, Bluetooth, wi-fi e a luz do Sol são exemplos de ondas eletromagnéticas. Os movimentos ondulatórios das ondas eletromagnéticas são descritos matematicamente pela equação 2.1.

$$Y = A \cdot sen\left[2\pi \cdot \left(\frac{t}{T} - \frac{x}{\lambda}\right) + \varphi\right] \tag{2.1}$$

Onde

- A: Representa a amplitude máxima do campo elétrico $\vec{E}$ ou do campo magnético $\vec{M}$;
- t: Representa o tempo;

- T: Representa o período, ou seja, é o tempo de um ciclo completo de uma oscilação de uma onda;
- x: Corresponde à distância em relação ao referencial tomado como origem;
- λ: Comprimento de onda;
- φ: Constante de fase.

A figura 2.2 corresponde à representação gráfica de um campo elétrico de amplitude máxima $E_{\max}$ e comprimento de onda λ. O comprimento de onda corresponde à distância percorrida pela onda elétrica (campo elétrico) a cada período de tempo T.

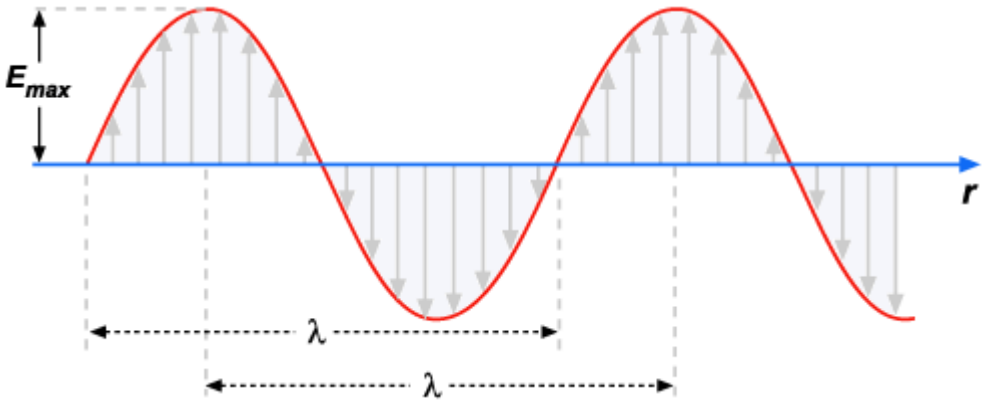


Figura 2.2: Campo elétrico.

Um dos conceitos de maior importância para os conteúdos apresentados nas próximas seções é a freqüência da onda, $f = \frac{1}{T}$.

A figura 2.3 ilustra o que vem a ser conhecido por espectro de freqüência. Em tal figura é possível notar que o comprimento de onda diminui conforme a freqüência aumenta. Em particular as ondas eletromagnéticas cujos comprimentos de ondas estão compreendidos entre 400 e 800 nm são chamadas de Luz Visível ou Espectro Visível.

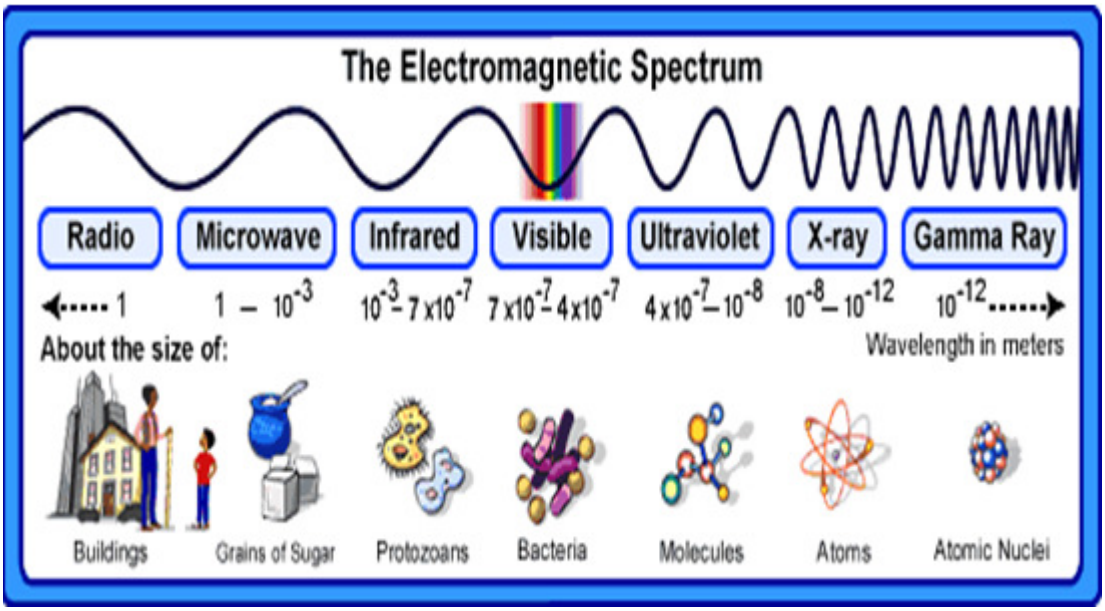


Figura 2.3: O espectro eletromagnético.

A figura 2.4 destaca apenas a faixa visível do espectro. Nota-se que dependendo da freqüência da onda a cor muda; um fato importante é que cada cor não é definida por uma freqüência única, mas sim por uma faixa existindo áreas de interseção entre uma cor e outra, ou seja, as transições entre as cores são suaves.

Cores do espectro visível

Cor	Comprimento de onda	Frequência
vermelho	~ 625-740 nm	~ 480-405 THz
laranja	~ 590-625 nm	~ 510-480 THz
amarelo	~ 565-590 nm	~ 530-510 THz
verde	~ 500-565 nm	~ 600-530 THz
ciano	~ 485-500 nm	~ 620-600 THz
azul	~ 440-485 nm	~ 680-620 THz
violeta	~ 380-440 nm	~ 790-680 THz

Espectro Contínuo

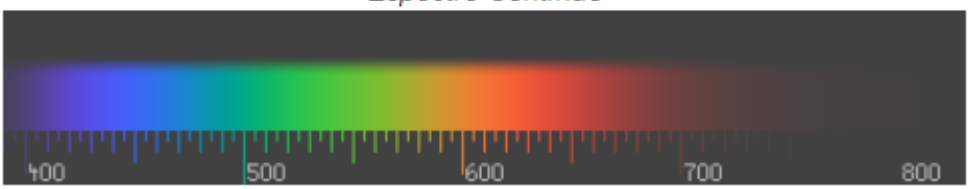


Figura 2.4: O espectro da luz visível.

De maneira resumida pode-se dizer que as ondas eletromagnéticas são luzes que dependendo da freqüência serão visíveis ou não, ou melhor, dependendo da freqüência elas são capazes de provocar a **SENSAÇÃO DE COR**.

2.2 A formação da cor

Cor é a percepção visual provocada pela ação da luz na retina. Tomando como exemplo uma parede com duas cores, metade branca e metade verde, a explicação física para a percepção de tais cores é simples. Dado uma fonte de energia que emita ondas eletromagnéticas na faixa do visível, por exemplo, o Sol ou uma lâmpada incandescente; essa radiação, ou melhor, energia irradiada vai bater na superfície da parede e parte será absorvida e a outra refletida.

Lembrando que cada cor segundo a figura 2.4 é representada por uma faixa de freqüências então é verdadeiro afirmar que corpos que refletem luz e são relativamente **BALANCEADOS** em **TODOS** os comprimentos de onda visíveis provocarão no observador a sensação da **COR BRANCA**. No caso da parede verde, significa que a mesma não é balanceada em **TODOS** os comprimentos de onda visíveis, ela reflete uma **MISTURA** limitada do espectro visível, que por sua vez provocará a sensação da cor verde, ou melhor, de um **TOM** de verde.

Resumindo, dependendo da mistura refletida, basicamente três sensações poderão ser provocadas, são elas:

- **Mudança de cor (Matiz)**;
- **Mudança do brilho da cor:** É o efeito onde a cor parece mais clara, podendo chegar próxima do branco, ou mais escura, podendo chegar próxima do preto, contudo a cor não muda;
- **Mudança da saturação da cor:** É o efeito onde a cor parece mais pura, neste caso é dito que a cor está mais saturada, ou mais misturada com a cor branca, neste caso podendo chegar próxima do branco. Veja também que a cor não muda.

2.2.1 As cores primárias e secundárias

Cores primárias são aquelas que podem ser combinadas para gerar as cores secundárias, logo as cores primárias não podem ser decompostas, mas sim misturadas para gerar as demais.

As cores primárias[1] são dependentes dos espaços de cores utilizados, ou melhor, cada espaço de cor, seção 2.3, possui um conjunto diferente de cores primárias.

Modernamente consideram-se dois casos para a classificação de cores: **cores obtidas por processo aditivo** (objetos que emitem luz – sistema aditivo) e **cores obtidas por processo de reflexão** (objetos que refletem luz – sistema subtrativo).

Para exemplos de objetos que emitem luz tem-se monitores, televisores e lanternas; as cores emitidas por cada pixel de um televisor é resultado da mistura das cores primárias usadas por tal equipamento. Para exemplos de objetos que refletem luz tem-se uma camisa azul, uma bolsa vermelha, um livro de capa verde.

No caso da camisa azul ela é percebida com esta cor, pois se sobre a mesma incidir uma feixe de ondas eletromagnéticas composto por todas as ondas da faixa do visível a maior parte das ondas serão absorvidas e apenas uma mistura de freqüências será refletida. A luz refletida ao chegar aos olhos do observador provocará a percepção visual correspondente a um tom de azul.

Estudos comprovam que o olho humano é mais sensível a uma mistura das faixas de freqüências referentes ao **vermelho**, **verde** e **azul**, contudo a mistura destas cores não é capaz de produzir todas as cores existentes na natureza.

Os dispositivos de visualização de imagem procuram "enganar" o olho humano, usando justamente o fato de o sistema visual ser mais sensível às faixas do vermelho, verde e azul. Cabe ressaltar que atualmente não existem equipamentos capazes de produzir todas as misturas de cores existentes na natureza.

Por exemplo, o que uma TV colorida faz é escolher uma freqüência dentro da faixa do vermelho, outra dentro da do verde e por fim uma freqüência dentro da faixa do azul. Após essa escolha, ela terá apenas o trabalho de emitir as luzes nas freqüências escolhidas.

---

[1] As faixas de freqüência correspondentes ao **vermelho**, **verde** e **azul** são normalmente usadas como cores primárias nos equipamentos que usam **combinações aditivas de cores**, em outras palavras equipamentos que emitem luz. Em áreas que usam **combinações subtrativas de cores**, como na mistura de pigmentos ou corantes, por exemplo, na pintura ou impressão gráfica as cores primárias normalmente são **ciano**, **magenta** e **amarelo**.

Gerando as cores

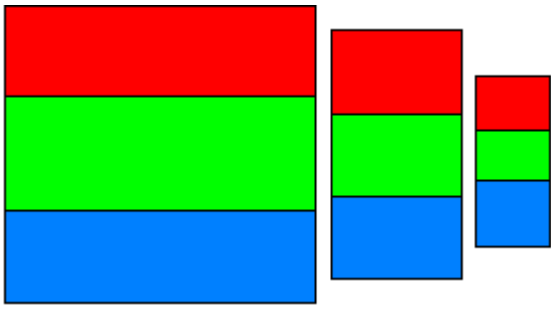


Figura 2.5: Representação de um pixel de um display colorido. Nota-se que cada pixel é composto por três "micro lâmpadas", nas cores vermelho, verde e azul. O tamanho do pixel diminui quanto maior for a resolução do display.

As telas dos dispositivos de visualização de imagens tais como TV, PC e Celulares, podem ser vistas como uma matriz onde cada célula é chamada de pixel, figura 2.5. O número total de pixels corresponde à **resolução** da tela (o termo resolução será posteriormente retomado na explicação sobre imagens digitais). Cada pixel por sua vez é formado por um conjunto de três "pequenas lâmpadas de fósforo", o princípio de funcionamento é simples.

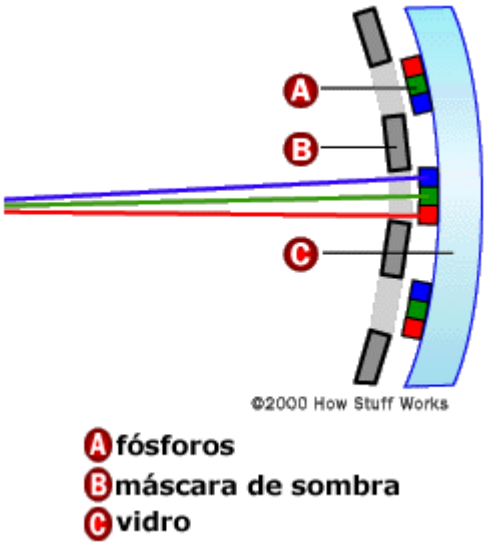


Figura 2.6: Matriz de fósforos.

Um feixe de elétrons é disparado e quando atinge o elemento de fósforo, figura 2.6, este brilha emitindo uma cor; a máscara de sombra ajuda no controle do feixe de elétrons para que esta atinja o fósforo correto. A intensidade do brilho vai depender da intensidade do feixe

de elétrons. Assim como existem lâmpadas com cores diferentes, também existem fósforos que brilham com cores diferentes.

Nos displays coloridos existem três tipos de fósforos: os que brilham na faixa do vermelho, os que brilham na faixa do verde e por fim os que brilham na faixa do azul. É a mistura destas luzes que ao atingirem os olhos do observador provocam as mais diversas sensações de cores. Em displays pretos e brancos os fósforos só brilham em tons de cinza, onde preto corresponde ao fósforo apagado e branco ao brilho máximo.

Para formar uma imagem em preto e branco (tons de cinza) em display colorido, basta que as luzes dos três fósforos que compõem um pixel sejam emitidas com as mesmas intensidades, em outras palavras as ondas eletromagnéticas emitidas possuirão as mesmas amplitudes.

2.3 Os espaços de cores

Os espaços de cores também conhecidos como sistemas de cores ou espetro de cores são modelos abstratos usados para representar as cores de uma maneira uniforme. Esta representação é realizada por meio de fórmulas matemáticas.

A preocupação na criação de tais modelos vem antes da invenção dos monitores coloridos e dos equipamentos de captura e impressão de imagem. O estudo das cores, ou melhor, dos pigmentos começou com os pintores que em tempos antigos dispunham de poucas cores bases (primárias) que tinham que ser misturadas para representar as demais cores existentes na natureza.

Um dos motivos que impulsionou os cientistas, artistas e técnicos a desenvolverem tais modelos foi, por exemplo, o caso em que se deseja pintar algo sobre uma folha de papel e dependendo da qualidade e natureza do pigmento, bem como da cor da folha que servirá como base para a tinta, os resultados que deveriam ficar iguais acabavam com aparências diferentes. Existem espaços de cores criados para representar as cores geradas por processo aditivo, bem como espaços criados para representar as cores geradas por processos subtrativos.

Espaços gerados por processo aditivo: Equipamentos que combinam luz emitida para provocar a sensação de variação de cor.

- **RGB** (Red, Green, Blue - Vermelho, Verde e Azul)

É um modelo amplamente utilizado nos processos aditivos. Nele cada cor é formada por uma tupla (R, G, B), ou melhor, pela mistura de certa quantidade de vermelho, verde e azul. A figura 2.7 corresponde à representação gráfica do espaço de cores RGB.

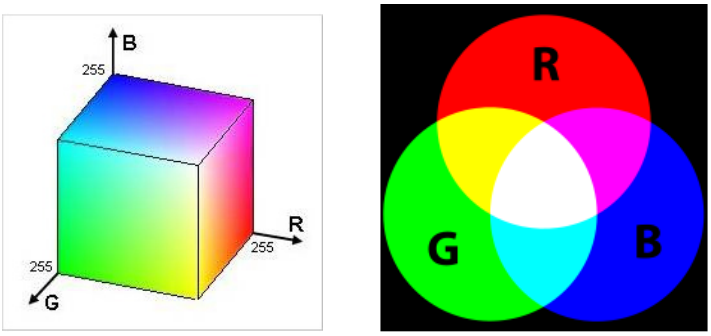


Figura 2.7: O espaço de cor RGB.

- **YIQ**

É um modelo muito utilizado em transmissão de imagens, TV colorida. A componente Y (luminância) transporta a informação de imagens, ou seja, os aparelhos de TV "preto e branco" precisam apenas desta componente para produzir as imagens. Y é proporcional à quantidade de luz percebida pelo olho humano. Os componentes cromáticos I e Q por sua vez são responsáveis por transportarem a informação de cor, eles são chamados respectivamente de em-fase e quadratura.

A vantagem do uso deste modelo em processamento de imagem é que as três componentes são desacopladas o que possibilita aplicar transformações em Y, tal como equalização do histograma (seção 2.5) sem afetar a informação de cor. A transformação do espaço RGB para o espaço YIQ é realizada através da equação 2.2.

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.578 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.2}$$

Equação para conversão de RGB para YIQ.

Espaços gerados por processo subtrativo

- **CMYK** (Cyan, Magneta, Yellow, Key - Ciano, Magenta, Amarelo e Preto).

É um modelo amplamente utilizado em equipamentos de impressão. O uso do preto é por que a mistura de ciano, magenta e amarelo não gera preto e sim um marrom muito escuro, ou seja, é gasto muita tinta e não se atinge o preto, então por questões de economia de produtos adiciona-se um cartucho de preto nestes equipamentos.

A figura 2.8 mostra uma mesma imagem representada nas 4 componentes do espaço de cor $CMYK$.

Figura 2.8: O espaço de cor CMYK.

A figura 2.9 traz o espaço de cor CMY representado via diagrama de Venn.

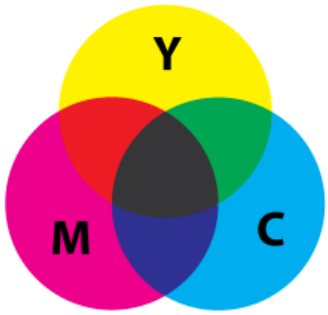


Figura 2.9: O espaço de cor CMY.

A transformação do espaço CMY para o espaço YIQ é realizada através da equação 2.3.

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.3}$$

Equação para conversão de RGB para CMY .

- **RYB** (Red, Yellow, Blue – Vermelho, Amarelo e Azul).

É um modelo muito usado nas artes, principalmente em pintura. Seu uso vem de antes da teoria científica moderna sobre cores. Conforme a figura 2.10 as cores primárias bem como as secundárias (violeta, laranja e verde) são dispostas em um círculo cromático de modo que formem uma tríade de cores, em outras palavras, formem um triângulo eqüilátero onde cada cor fica em um dos vértices do triângulo.

Figura 2.10: O espaço de cor RYB

Outros espaços de cores

CIE1931 XYZ: É um espaço de cor tridimensional que funciona como base para todos os sistemas de gerenciamento de cores. Este espaço é capaz de representar todas as cores perceptíveis pelo olho humano. Cabe lembrar que atualmente nem todas essas cores são reproduzidas pelos monitores e impressoras.

A projeção do espaço CIE1931XYZ em um espaço bidimensional é chamado de diagrama de cromaticidade CIExyY. Muitas interpretações podem ser obtidas usando o diagrama CIExyY, mas outras só são possíveis com base no CIE1931XYZ ou em outra variante conhecida por CIE-Lab.

As variáveis X, Y e Z correspondem aos valores de vermelho, verde e azul necessários para formarem uma cor. O conjunto formado por tais elementos é chamado de triestímulo (HOFFMANN, 2003). Através da equação de normalização, equação 2.4, são calculados os coeficientes tricromáticos.

$$x = \frac{X}{X+Y+Z}\ ,\ y = \frac{Y}{X+Y+Z},\ z = \frac{Z}{X+Y+Z},\ x+y+z=1 \qquad (2.4)$$

Equações para o cálculo dos triestímulos.

Onde

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0\,49000 & 0.31000 & 0.20000 \\ 017697 & 0.81240 & 0.01063 \\ 0.00000 & 0.01000 & 0.99000 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.5}$$

Equação para conversão de RGB para CIE1931 XYZ.

A figura 2.11 corresponde ao **Diagrama de Cromaticidade CIExyY**.

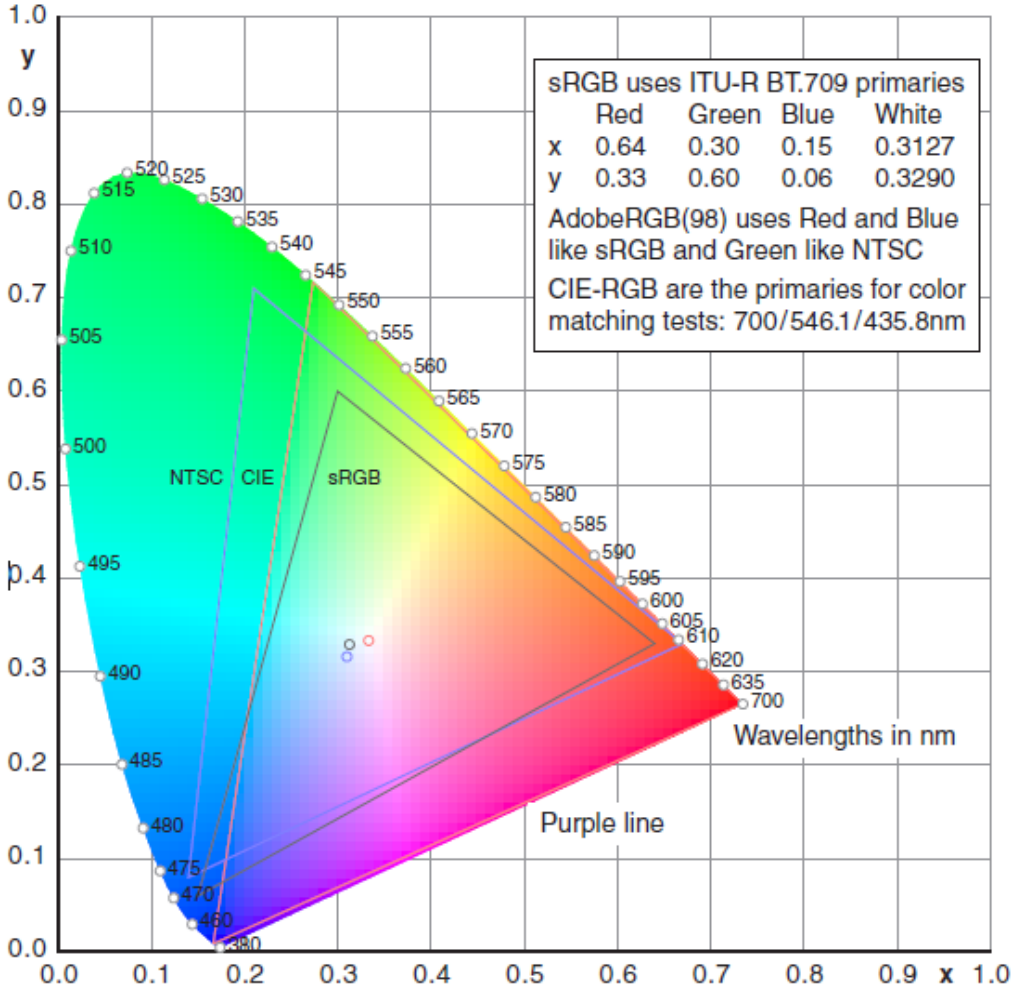


Figura 2.11: Diagrama de cromaticidade CIExyY.

Para que seja gerado um novo espaço de cores basta escolher três pontos, cores primárias, e ligá-los por segmentos de retas formando assim um triângulo. As cores dentro do triângulo corresponderão a todas as variações possíveis de serem geradas pelo espaço de cores criado. Estendendo esta noção de dentro e fora se conclui que a região dentro do diagrama de cromaticidade corresponde a todas as cores possíveis de serem derivadas da mistura de cores

na faixa do visível, e a região externa ao diagrama corresponde a "cores" formadas pelas misturas de ondas eletromagnéticas fora da faixa do visível.

2.4 Representação de imagens digitais

Para que a imagem e a cor de um objeto possam ser vistas é necessária que uma quantidade mínima de luz seja refletida e chegue aos olhos do observador. Quando a luz é pouca dá a impressão de que os objetos estão mais escuros e quando a luz é muito intensa dá a impressão de que as cores dos objetos ficaram mais vivas e no caso limite tende para branco $R = G = B = 255$, sistema de cor RGB.

Esta sensação de variação de luz é obtida com a mudança nos valores $0 \leq R \leq 255$, $0 \leq G \leq 255$ e $0 \leq B \leq 255$, conforme os valores tendem à 255 indica que a intensidade da luz em cada uma destas freqüências está aumentando, é como girar um potenciômetro ligado a uma lâmpada para fornecer mais corrente elétrica, onde a amplitude máxima significa $R = G = B = 255$.

Para fazer com que a cor de um pixel varie do preto ao branco (escala do cinza), dado que se esteja usando o modelo de cor RGB, basta que os valores (R, G, B) do pixel sejam sempre iguais, em outras palavras, é equivalente a escolher valores ao longo da diagonal do cubo que representa o espaço de cores RGB, figura 2.7.

Uma forma eficiente de alterar a sensação de luz emitida por um pixel (o pixel fica mais claro ou mais escuro) sem mudar a cor do mesmo é transformar os valores dos pixels para outro sistema de cores onde a componente de luminância fique desacoplada das componentes de cor. Isso é possível nos modelos HSV, HSL, HSB, YC_RC_B, YIQ entre outros. Nestes modelos é possível aplicar transformações matriciais na componente de luminância[2] tal como Expansão do Histograma (seção 2.5) e em seguida aplicar a transformação inversa para retornar ao espaço RGB sem que as componentes de cor sejam afetadas, em outras palavras, a cor predominante do pixel não vai mudar.

Por exemplo, imagens geradas por equipamentos tais como: televisores, celulares e monitores de computador são sempre vistas no espaço RGB, pois, conforme já explicado, os displays destes equipamentos são construídos com elementos de fósforo que emitem luz na freqüência do vermelho, verde e azul, ou seja, trabalham com o formato RGB. De fato a cor vista na figura 2.12 não é a do "pixel no espaço HSL" e sim a cor que o pixel terá no espaço

[2] A componente de luminância é a responsável por modelar a sensação de intensidade luminosa. Tomando o modelo HSL como referência, sua componente de luminância é representada pela variável L.

RGB, pois os monitores emitem em RGB, para uma dada tupla (matiz, saturação, brilho) escolhida no modelo HSL e convertida para RGB.

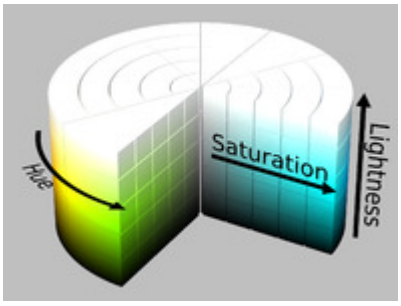


Figura 2.12: Espaço de cores HSL.

Um erro muito comum é encontrar artigos mencionando que saturação é a quantidade de branco misturada no matiz. Essa definição depende do modelo, pois no caso do sistema HSB a saturação é a quantidade de cinza misturado no matiz, e no caso de saturação mínima o pixel fica cinza e não branco.

Por fim a última observação importante antes das explicações sobre imagens digitais é que o espaço de cor RGB é dito **não perceptualmente uniforme**. Por exemplo, dado um pixel na cor vermelha $(50,0,0)$ se suas coordenadas G e B sofrerem alterações as cores resultantes apresentadas pelo mesmo provavelmente não serão uma variação de tons de vermelho. Já o HSV $(H = Matiz, S = Saturação, V = Brilho)$ é dito **perceptualmente uniforme**, pois, para mudar o brilho de um pixel basta variar V, para tornar a cor mais viva ou mais misturada com o branco é só variar S e por fim a mudança de cor é feita variando o valor de H, em outras palavras, as coordenadas são desacopladas e basta fixar o valor de duas e variar livremente a terceira que o resultado é **exatamente o esperado**. A figura 2.13 e a figura 2.14 correspondem respectivamente aos espaços de cor RGB e HSV.

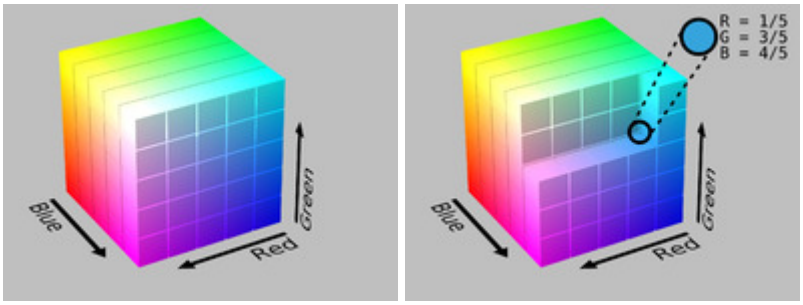


Figura 2.13 O espaço de cor RGB.

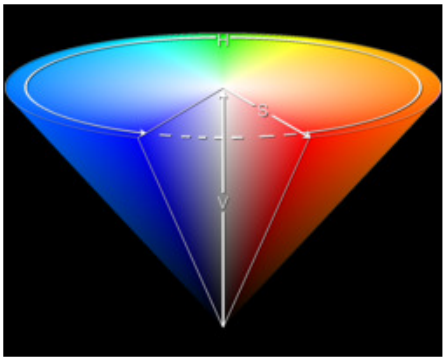


Figura 2.14: O espaço de cor HSV (Matiz, Saturação e Valor)

2.4.1 A imagem digital

Uma imagem digital nada mais é do que uma matriz, onde cada célula desta matriz é chamada de pixel. A quantidade de pixel vai determinar a resolução da imagem. A figura 2.15 mostra a representação matricial de uma imagem digital e a convenção de eixos utilizada.

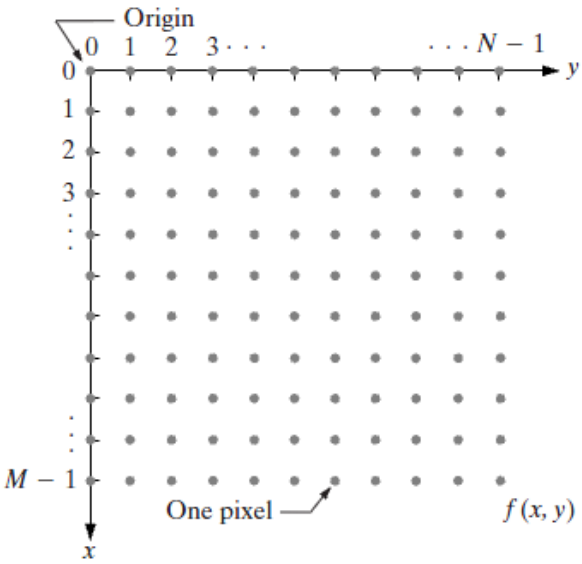


Figura 2.15: Imagem digital (GONZALEZ e WOODS, 2000).

Basicamente as imagens digitais se dividem em imagens coloridas e imagens em tons de cinza. Cada pixel em uma imagem colorida guarda três valores numéricos compreendidos entre 0 e 255, estes são os valores (R, G, B) ou de outro espaço de cores do pixel. Outra forma de imaginar a imagem colorida é como sendo uma matriz em três camadas, são elas camadas R, G e B, cada qual correspondendo a uma dimensão de cor. Esta característica é muito utilizada no processamento de imagens.

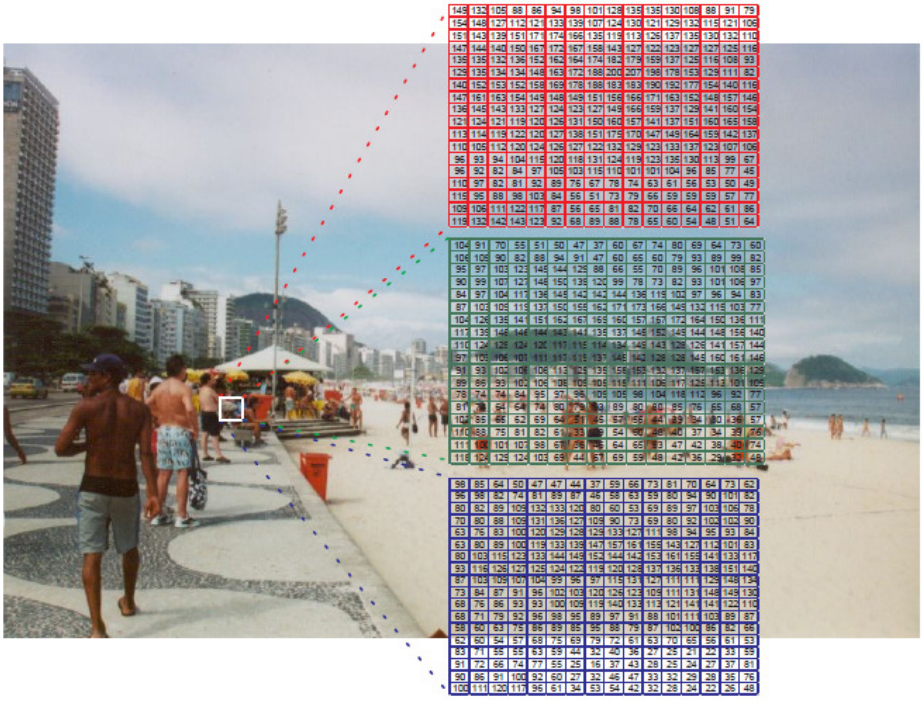


Figura 2.16: Imagem R, G, B e a decomposição em matrizes apresentada em (MOUTINHO, 2005).

A imagem em tons de cinza é representada por uma matriz com uma única camada, em que cada pixel por sua vez guarda apenas um valor numérico compreendido entre 0 e 255. Este número corresponde a uma intensidade de tom de cinza.

Figura 2.17: Imagem em tons de cinza e sua matriz característica apresentada em (MOUTINHO, 2005).

Para obter uma imagem em tons de cinza a partir de sua imagem colorida basta convertê-la do espaço de cor RGB, sendo que agora R, G e B são matrizes, para espaços de cor que desacoplem as dimensões, por exemplo, HSV, HSL, HSB, YC_RC_B, YIQ. Nestes a dimensão referente à luminância corresponderá à matriz em tons de cinza, no caso de espaço YC_RC_B a componente Y será uma imagem em tons de cinza, figura 2.17.

Através da equação 2.6 é possível realizar a conversão do espaço de cores RGB para YC_RC_B, onde a matriz Y quando visualizada na forma de imagem, por meio de um display colorido ou preto e branco, vai representar a imagem original só que agora em tons de cinza.

$$\begin{aligned} Y &= 0{,}30R + 0{,}59G + 0{,}11B \\ C_R &= R - Y \\ C_B &= B - Y \end{aligned} \tag{2.6}$$

Por fim existem duas dúvidas muito comuns, a primeira é quando se tenta visualizar as matrizes R, G e B em separado esperando obter imagens em tons de vermelho, verde e azul, mas o que se vê são imagens em tons de cinza, estas por sua vez não correspondem exatamente à imagem original passada para tom de cinza. Ao longo do capítulo 2 quando foi falado que a imagem estava em tons de cinza querer-se-ia dizer que ela sofreu a conversão de RGB para espaços que desacoplam as dimensões, tal como YC_RC_B onde a componente Y correspondente à luminância, em outras palavras corresponde à verdadeira imagem em tons de cinza.

Para se conseguir uma imagem em tons de vermelho a partir de uma imagem colorida, basta atribuir zeros para todos os valores nas matrizes G e B mantendo os valores de R. Tal procedimento é válido para se gerar imagens em tons de G ou B, bastando sempre atribuir zeros aos valores das outras duas matrizes.

A segunda é como são geradas as imagens pretas e brancas que possuem três camadas assim como as imagens coloridas. A característica que faz com que estas imagens sejam vistas em tons de cinza é que suas três camadas são iguais, ou seja, o pixel da posição UM é igual nas três matrizes, o da posição DOIS é igual nas três e assim por diante. Retomando a explicação dada sobre como um display colorido consegue mostrar uma imagem preta e branca, naquela explicação foi dito que bastava que os três elementos de fósforo que formam um pixel estivessem emitindo as cores (R, G, B) em igual intensidade, que a sensação provocada por tal mistura de luzes ao atingir os olhos do observador seria a da **cor branca**, ou

melhor, de um tom de branco que poderia ir do branco total passando por várias tonalidades de cinza até atingir o preto conforme os valores atribuídos a trupla (R,G,B). Falando em espaço de cores isso corresponde a percorrer a diagonal principal do cubo, origem em $(0,0,0)$, que modela o espaço RGB .

2.5 Equalização de histograma

O histograma nada mais é do que a distribuição de freqüências dos valores dos pixels em uma imagem monocromática. Quando se calcula o histograma de alguma imagem é comum que esta primeiramente seja representada em níveis de cinza.

Tomando o espaço de cores YIQ como exemplo, no qual a componente Y(luminância) corresponde à imagem em tons de cinza, é possível equalizá-la sem interferir nas componentes de cor, em outras palavras, a cor predominante de cada pixel na imagem não muda.

Um erro comum é pegar uma imagem em RGB, separá-la em suas componentes R, G e B, e equalizar cada uma das imagens monocromáticas formadas, por fim juntar as três camadas formando a imagem final; pois bem o histograma ficará equalizado em cada componente, mas se estará mexendo na informação de cor, pois a imagem resultante terá suas cores completamente modificadas. O correto seria converter, por exemplo, para o espaço YIQ equalizar a componente Y e por fim aplicar a transformação inversa retornando para o espaço de cores RGB.

Para a construção do histograma considera-se uma imagem[3] em níveis de cinza cujos pixels assumem valores no intervalo $r_k \in [0, L-1]$ e uma função discreta $p(r_k) = {n_k}/{n}$, onde r_k é o k -ésimo nível de cinza, n_k é o número de pixels com nível de cinza igual a r_k e n é o total de pixels na imagem. O histograma nada mais é do que a curva descrita por $p(r_k) = {n_k}/{n}$ ao longo do intervalo $[0, L-1]$.

[3] Supondo que é dada uma imagem de resolução NxM , quantidade de pixels; se for dito, por exemplo, que a mesma possui 8 bits isto nada mais é do que dizer que cada pixel da imagem só poderá receber um valor, tom de cinza r_k , sendo $\{r_k \in N \mid 0 \le r_k \le (2^8 - 1)\}$.

Abaixo seguem exemplos de imagens e seus histogramas

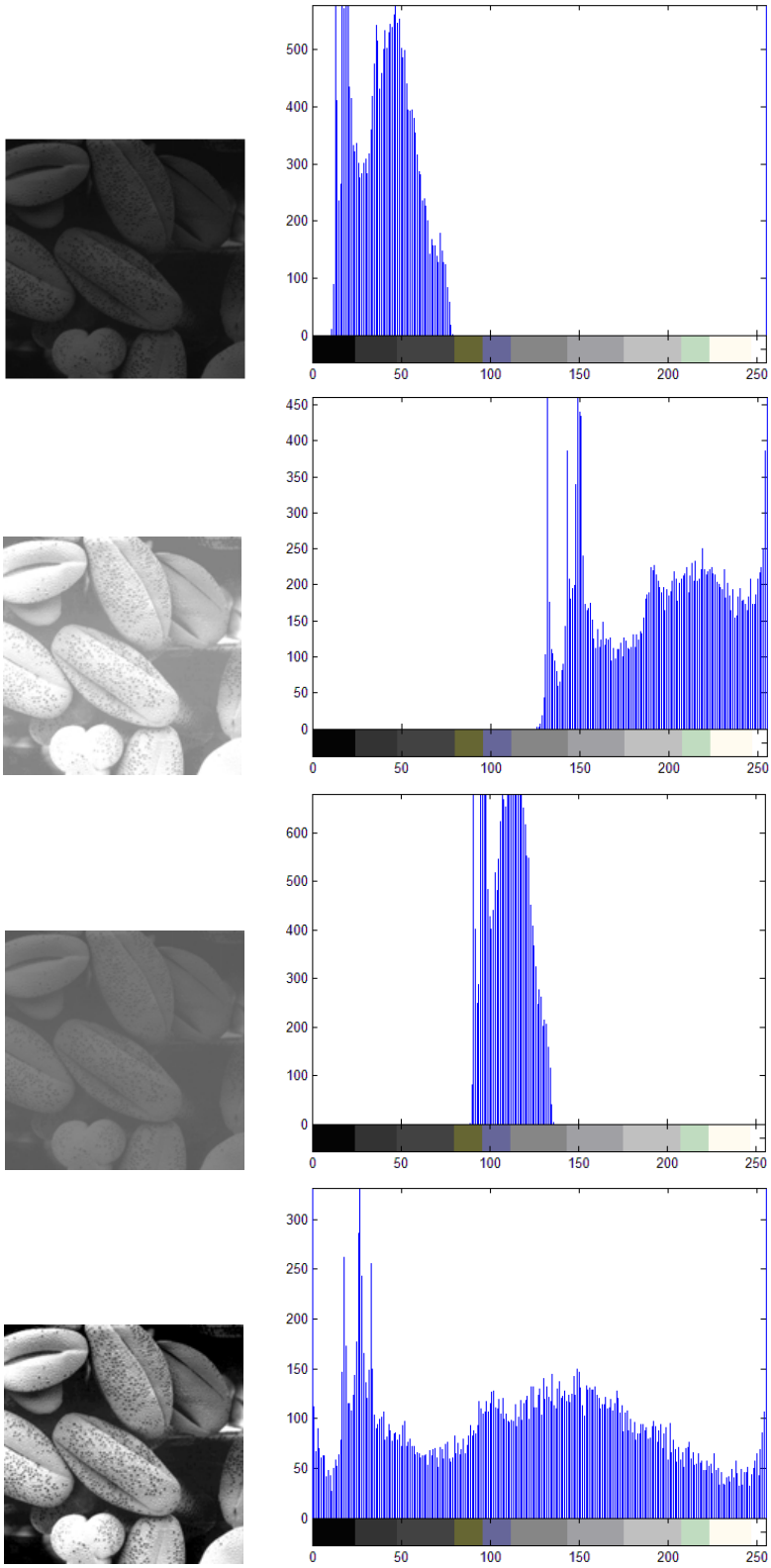


Figura 2.18: As imagens de (a)-(d) correspondem a uma imagem com e seus respectivos histogramas.

Na figura 2.18 é apresentada uma seqüência de quatro imagens e seus respectivos histogramas. A saber, tem-se uma imagem escura, uma clara, uma com baixo nível de contraste e a última com alto nível de contraste. Como pode ser visto o histograma em si traz apenas informação global sobre a imagem, ou seja, não diz nada sobre o conteúdo da mesma; já a forma do histograma indica se é válido ou não aplicar técnicas de realce do contraste como, por exemplo, a equalização do histograma.

A técnica de equalização do histograma visa aumentar o contraste dos pixels, em outras palavras ela tenta transformar as imagens, estando estas em tons de cinza, tal que seus histogramas fiquem uniformemente distribuídos ao longo do intervalo $[0,255]$. Aumento de contraste significa ampliar a diferença média entre os maiores e menores valores de luminância contidos na imagem em níveis de cinza (GONZALEZ e WOODS, 1992).

O procedimento de equalização

O objetivo da equalização do histograma é definir uma função de transformação $T(r_k)$ que mapeie cada valor r_k, antigo nível de cinza, em outro valor $s_k = T(r_k)$, novo nível de cinza. Em outras palavras deseja-se percorrer uma imagem, pixel a pixel, verificar qual é o seu nível de cinza, r, e substituí-lo por um novo valor, s. Por fim ao se calcular o histograma, $p(s_k) = {n_k}/{n}$, da nova imagem é esperado que este esteja uniformemente distribuído ao longo do intervalo $[0, L-1]$. Em (GONZALEZ e WOODS, 1992) é demonstrado que a equação de mapeamento que melhor atende a esta finalidade é dada por:

$$s_k = T(r_k) = \sum_{j=0}^{k} \frac{n_j}{n} = \sum_{j=0}^{k} p_r(r_j) \tag{2.7}$$

Probabilidade acumulada de $p(r_k)$.

Onde

$0 \le r_k \le 1$: Nível de cinza;

$k = 0, \quad 1, \quad \cdots, \quad L-1$: Índice para percorrer os diferentes níveis de cinza.

As figuras 2.19 e 2.20 ilustram o resultado da equalização do histograma sobre uma imagem.

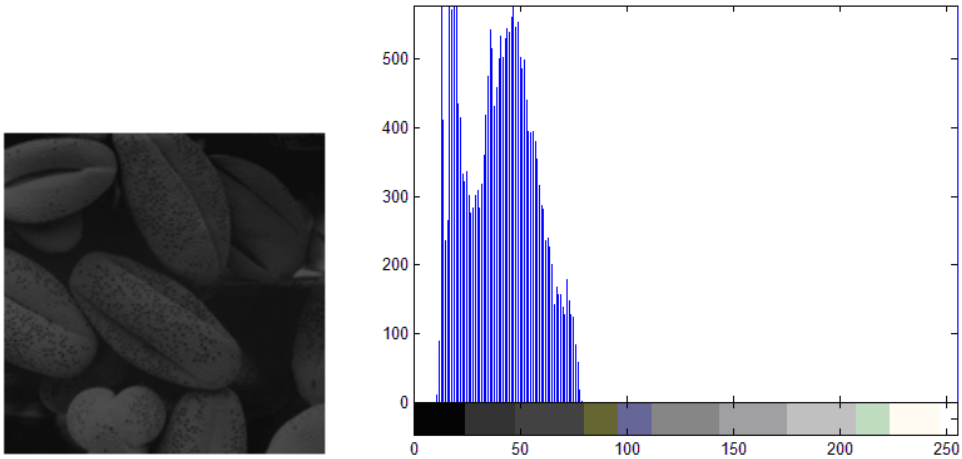


Figura 2.19: Imagem Original

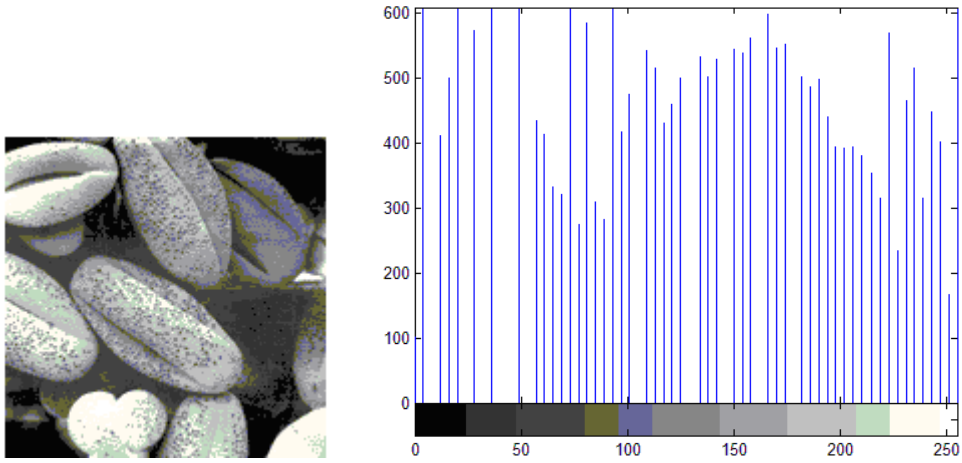


Figura 2.20: Imagem Equalizada

3 Transformações matriciais, análise de dados e reconhecimento facial

A ordem de apresentação e o conteúdo abordado ao longo do capítulo 3 foram escolhidos de modo a proporcionarem um entendimento de maneira gradual e linear. Os tópicos a seguir discorrerão sobre técnicas de transformação matricial, técnicas fundamentais de análise de padrões, técnicas de redução de dimensionalidade e por fim uma breve introdução aos métodos biométricos e aos sistemas de reconhecimento facial.

3.1 Transformações matriciais

Em processamento de imagem é muito comum se fazer o uso das técnicas de translação, mudança de escala e rotação, sendo que estas nada mais são do que transformações lineares do tipo $T : R^n \rightarrow R^n$; no caso de processamento de imagens $2D$ tem-se $T : R^2 \rightarrow R^2$.

Usando a notação matricial

$$P_T = T \cdot P \tag{3.1}$$

Onde

- T : Matriz de transformação de $R^2 \rightarrow R^2$.
- $P = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$, coordenadas de ponto (pixel) no R^3; não é o valor do pixel na imagem.

  Para o caso R^2 basta fazer $z = 0$;
- $P_T = \begin{bmatrix} x_T \\ y_T \\ z_T \end{bmatrix}$, coordenadas do ponto após transformação.

Com o objetivo de simplificar o cálculo das transformações e construir uma matriz genérica, ou seja, que trate de transformações também no R^3, é comum representar T como uma matriz quadrada, (GONZALEZ e WOODS, 1992); Porém quando se tratarem de imagens em $2D$ basta fazer $z_T = z_0 = 0$ nos sistemas de equações de translação, mudança de escala e rotação.

3.1.1 Translação

A translação de um ponto ao longo dos eixos X, Y e Z é dada pela transformação

$$\begin{bmatrix} x_T \\ y_T \\ z_T \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & x_0 \\ 0 & 1 & 0 & y_0 \\ 0 & 0 & 1 & z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{3.2}$$

Onde

- $T = \begin{bmatrix} 1 & 0 & 0 & x_0 \\ 0 & 1 & 0 & y_0 \\ 0 & 0 & 1 & z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$: Matriz de transformação;
- (x_0, y_0, z_0): Quantidades transladadas ao longo dos eixos X, Y e Z.

3.1.2 Mudança de escala

A mudança de escala de um objeto ao longo dos eixos X, Y e Z é realizada através da aplicação de fatores multiplicativos ao longo de tais eixos. A transformação que realiza tal tarefa é dada por:

$$\begin{bmatrix} x_S \\ y_S \\ z_S \\ 1 \end{bmatrix} = \begin{bmatrix} S_x & 0 & 0 & 0 \\ 0 & S_y & 0 & 0 \\ 0 & 0 & S_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{3.3}$$

Onde

- $S = \begin{bmatrix} S_x & 0 & 0 & 0 \\ 0 & S_y & 0 & 0 \\ 0 & 0 & S_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$: Matriz de transformação;
- (S_x, S_y, S_z): Fatores multiplicativos ao longo dos eixos X, Y e Z.

Para garantir que o objeto final esteja na posição correta após aplicar S, são necessários alguns passos.

1. Deve-se transladar o objeto para que o centro do mesmo coincida com a origem dos eixos coordenados;
2. Aplicar a mudança de escala;
3. Aplicar a translação inversa.

3.1.3 Rotação

O caso geral de rotação é quando se deseja rotacionar um ponto P_a em torno de um ponto P_b. Para isso é necessário transladar todo o sistema, ou seja, P_a e P_b de modo que o ponto P_b caia sobre a origem dos eixos de coordenadas, aplica-se então a rotação de P_a em torno de P_b, por fim aplica-se a translação inversa do conjunto P_a e P_b de modo que P_b retorne a sua posição original antes das transformações. Como resultado final P_a não coincidirá com sua posição inicial visto que ele sofreu uma rotação.

Para fins de simplificação serão considerados três casos em separado, rotação em torno do eixo X, em torne do eixo Y e por fim em torno do eixo Z.

Rotação no sentido horário em relação ao eixo X:

$$\begin{bmatrix} x_{R_\alpha} \\ y_{R_\alpha} \\ z_{R_\alpha} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & sen\alpha & 0 \\ 0 & -sen\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{3.4}$$

Onde

- $R_\alpha = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & sen\alpha & 0 \\ 0 & -sen\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$: Matriz de transformação;
- α: Corresponde ao ângulo de rotação do ponto (x, y, z) em torno do eixo X.

Rotação no sentido horário em relação ao eixo Y:

$$\begin{bmatrix} x_{R_\beta} \\ y_{R_\beta} \\ z_{R_\beta} \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\beta & 0 & -sen\beta & 0 \\ 0 & 1 & 0 & 0 \\ sen\beta & 0 & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{3.5}$$

Onde

- $R_\beta = \begin{bmatrix} \cos\beta & 0 & -sen\beta & 0 \\ 0 & 1 & 0 & 0 \\ sen\beta & 0 & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$: Matriz de transformação;
- β: Corresponde ao ângulo de rotação do ponto (x, y, z) em relação ao eixo Y.

Rotação no sentido horário em relação ao eixo Z:

$$\begin{bmatrix} x_{R_\theta} \\ y_{R_\theta} \\ z_{R_\theta} \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & sen\theta & 0 & 0 \\ -sen\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{3.6}$$

Onde

- $R_\theta = \begin{bmatrix} \cos\theta & sen\theta & 0 & 0 \\ -sen\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$: Matriz de transformação;
- θ: Corresponde ao ângulo de rotação do ponto (x, y, z) em torno do eixo Z.

3.2 Técnicas fundamentais de análise dados

A análise de dado, também conhecida como análise de padrões, é uma subárea da Estatística. De um modo geral ela tem por objetivos buscar tendências, padrões, relações e inferências através da manipulação matemática de uma massa de dados. No caso do reconhecimento facial, essa massa corresponde a imagens contendo faces.

Na presente seção serão apresentadas algumas técnicas fundamentais que serão utilizadas nos métodos de reconhecimento facial, apresentados no capítulo 4.

3.2.1 Covariância

Na teoria de análise de padrões é comum a presença do termo "**VARIÁVEL**"; uma forma simples de exemplificação é supor a existência de um grupo de pessoas onde se deseja tomar nota das alturas e pesos e colocar em uma tabela com uma coluna para alturas e outra para pesos. A coluna com as alturas corresponderá à variável **ALTURA**, já a coluna com os pesos será a variável **PESO**.

A covariância e a correlação são medidas matemáticas que buscam expressar o grau de relacionamento entre duas variáveis (MONTGOMERY e GEORGE, 2003), por exemplo, a ALTURA e o PESO, em outras palavras elas medem como duas variáveis variam conjuntamente.

Antes da apresentação das equações que definem a covariância e a correlação é necessária a definição de algumas medidas auxiliares, sendo elas:

- **Média**

Dado o conjunto de valores $A = \{x \mid x_i \in R; i = 1, \cdots, M\}$ a média dos elementos do conjunto A é definida por:

$$\mu = \frac{1}{M} \sum_{i=1}^{M} x_i \tag{3.7}$$

Estendendo o conceito para vetores tem-se que dado o conjunto de vetores definidos por $B = \{\Phi \mid \Phi_i \in R^n; i = 1, \cdots, M\}$ o vetor médio é definido por:

$$\mu = \frac{1}{M}\sum_{i=1}^{M}\Phi_i \tag{3.8}$$

Onde

- $\Phi = [x_1, x_2, \cdots, x_n]$ e x_i : Corresponde a uma variável de estudo

- **Desvio Padrão:** É uma medida de espalhamento dos dados em torno do ponto médio do conjunto. Matematicamente é definido por:

$$\sigma_X = \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}(x_i - \mu)^2} \tag{3.9}$$

- **Variância:** É outra medida de espalhamento da amostra, matematicamente é muito semelhante ao desvio padrão, ou melhor, corresponde ao quadrado do desvio padrão.

$$\sigma_X^2 = \frac{1}{M-1}\sum_{i=1}^{M}(x_i - \mu)^2 \tag{3.10}$$

Definições

Pelo fato da covariância e da correlação serem capazes de modelar dependência entre duas variáveis, são necessárias algumas definições que auxiliarão num melhor entendimento do conteúdo abordado na presente seção.

Dependência entre Variáveis

Duas variáveis quaisquer X e Y são consideradas independentes se os resultados obtidos por X não influenciarem os resultados obtidos por Y, e vice-versa. Se os resultados de X influenciarem os resultados de Y, então X e Y serão dependentes.

Variáveis emparelhadas

Duas variáveis serão classificadas como emparelhadas sempre que na observação de um evento forem "anotados" dois valores: um para a variável X e outro para a variável Y, em outras palavras, quando uma é observada a outra também o é.

A Tabela 3.1 apresenta dados fictícios dos investimentos em propaganda e os resultados alcançados nas vendas. É possível notar a existência de um relacionamento entre as variáveis, pois quando o valor de uma muda o da outra também muda.

Tabela 3.1: Tabela representativa de vendas e resultados alcançados.

Investimentos em propaganda (X) – R$	Resultado das vendas (Y) – R$
30	430
21	335
35	520
42	490
37	470
2	210
8	195
17	270
35	400
25	480

Outra forma de analisar o relacionamento entre as variáveis é através do diagrama de dispersão, figura 3.1.

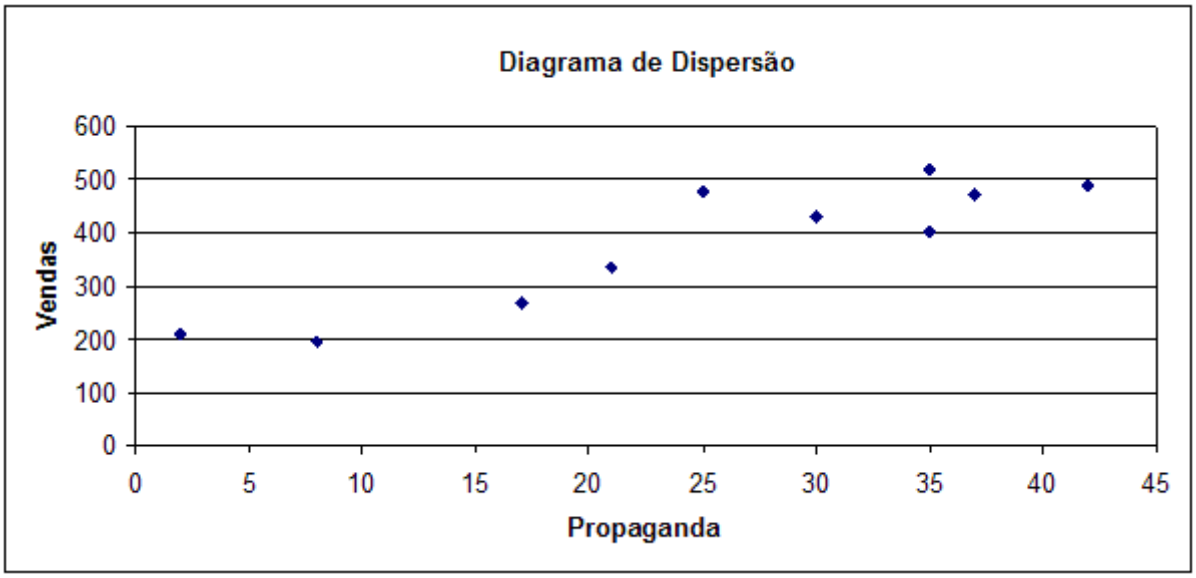


Figura 3.1: Diagrama de dispersão.

O diagrama de dispersão anteriormente apresentado permite não somente verificar a existência de relacionamento entre as variáveis mais também qual é o tipo e formato do mesmo.

Tipos de relacionamento

Os tipos de relacionamento podem ser divididos em positivos e negativos.

- **Positivos**: Este tipo de relacionamento ocorre quando as variáveis seguem a mesma tendência, ou seja, quando X cresce Y também cresce, quando X diminui Y também diminui.
- **Negativo:** Este tipo de relacionamento ocorre quando as variáveis seguem tendências opostas, ou seja, quando uma cresce a outra decresce.

O formato do relacionamento

Existem diversos formatos de relacionamento tais como linear, exponencial e quadrático.

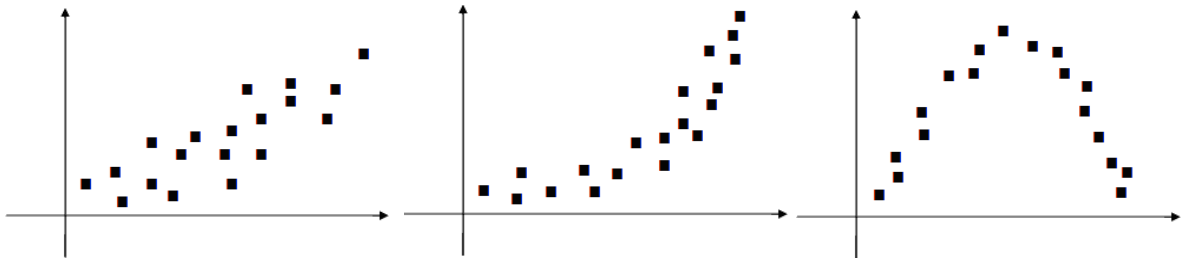


Figura 3.2: Formato dos relacionamentos, linear, exponencial e quadrático.

Equação da covariância

$$\sigma_{X,Y} = \frac{\sum_{i=1}^{M} (X_i - \mu_X)(Y_i - \mu_Y)}{M} \tag{3.11}$$

Onde

- μ_X : Média dos valores de X .
- μ_Y : Média dos valores de Y .
- M : Números de elementos de X e Y .

Interpretando a covariância

- $\sigma_{X,Y} > 0$: Quando X e Y aumentam juntos ou X e Y diminuem juntos, então as variáveis são ditas dependentes com relacionamento positivo.
- $\sigma_{X,Y} < 0$: Quando uma variável aumenta e a outra diminui, logo elas são ditas dependentes com relacionamento negativo.
- $\sigma_{X,Y} = 0$: Condição necessária, mas não suficiente para que duas variáveis sejam independentes. Em outras palavras, quando a covariância entre duas variáveis for nula isso não quer dizer que elas sejam independentes, porém necessariamente variáveis independentes possuem covariância nula.

Representações gráficas

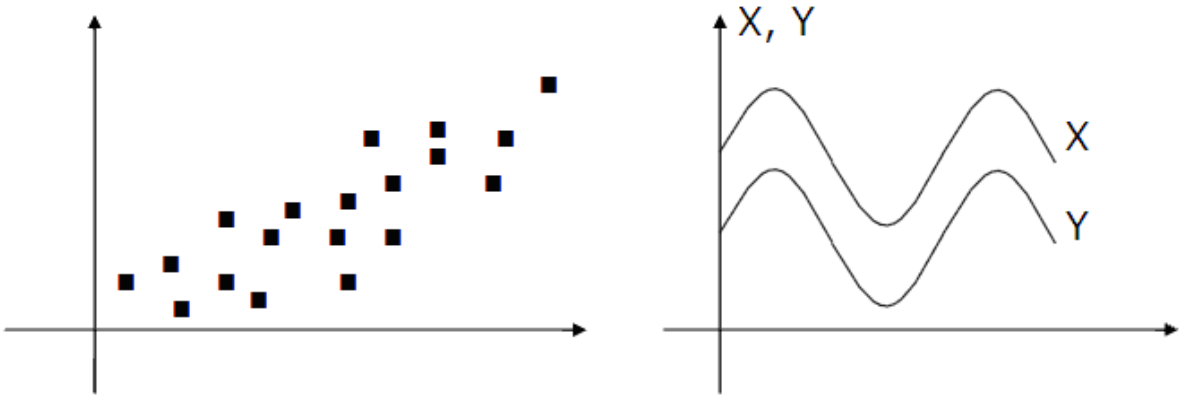


Figura 3.3: Covariância Positiva.

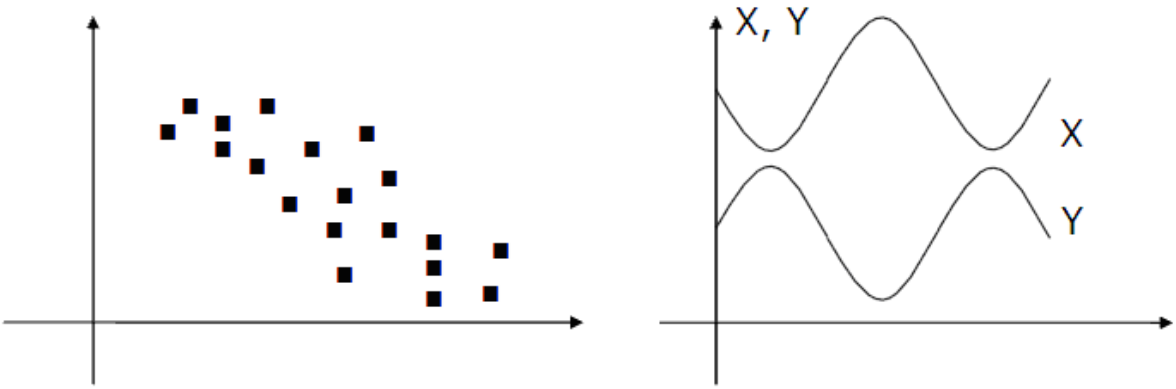


Figura 3.4: Covariância Negativa.

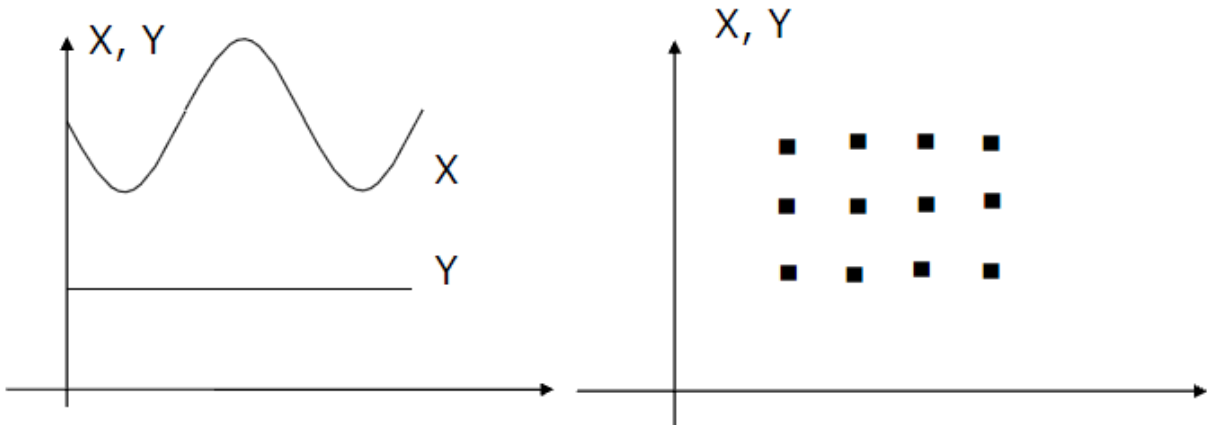


Figura 3.5: Covariância Nula.

Desvantagens no uso da Covariância

- Apresenta uma unidade de medida difícil de entender. Por exemplo, se as variáveis forem PESO (kg) e ALTURA (m) a unidade será kg x m.
- Pode resultar também em grandes valores de saída. O ideal é que todos os possíveis resultados estejam dentro de uma faixa de valores, situação esta em que a saída está normalizada.

3.2.2 Correlação

A correlação, criada por Karl Pearson em 1901, é um conceito semelhante ao da covariância e tem como um de seus objetivos contornar algumas das desvantagens apresentadas para o uso da covariância.

Equação da correlação

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \tag{3.12}$$

Onde

- $\sigma_{X,Y}$: Covariância entre as variáveis X e Y ;
- σ_X : Desvio padrão da variável X ;
- σ_Y : Desvio padrão da variável Y .

Propriedades

- Possui unidade de medida adimensional, ou seja, não possui unidade;
- Apresenta saída normalizada, $-1 < \rho_{X,Y} < 1$.
- Quanto mais próximo da unidade for o valor calculado para $\rho_{X,Y}$ significa que as variáveis X, Y são mais dependentes;
- Para o $\rho_{x,y}$ devemos primeiramente calcular o desvio padrão de ambas as variáveis.

Interpretando a correlação

- $\rho_{X,Y} = 1$: Quando X e Y aumentam juntos ou X e Y diminuem juntos; logo as variáveis são ditas dependentes com relacionamento positivo.
- $\rho_{X,Y} = -1$: Quando uma variável aumenta e a outra diminui; logo elas são ditas dependentes com relacionamento negativo.
- $\rho_{X,Y} = 0$: Condição necessária, mas não suficiente para que duas variáveis sejam independentes. Em outras palavras, elas são linearmente independentes, mas não exclui a existência de dependência não linear.

Representações gráficas

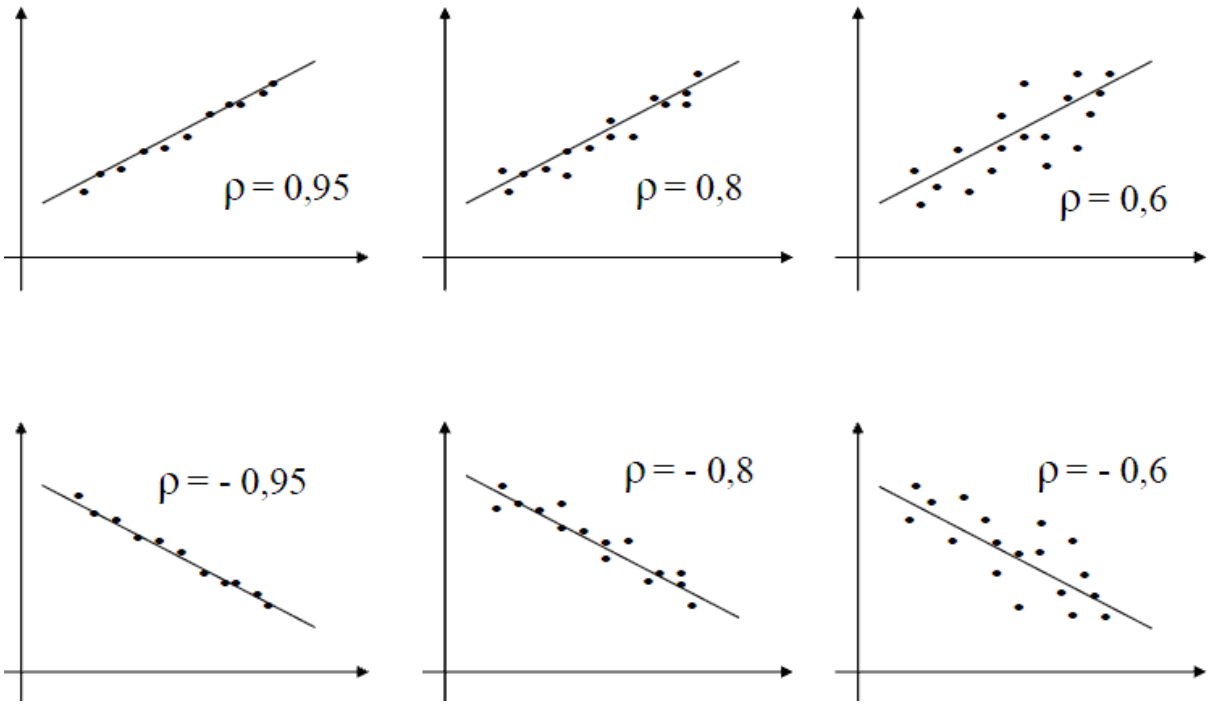


Figura 3.6: Tipos de correlações.

3.2.3 Autovalor e autovetor

Em áreas da matemática tal como Álgebra Linear é muito comum a presença de sistemas lineares com N variáveis por N incógnitas, os quais são representados de forma matricial pela equação $AX = \lambda X$, onde λ é um escalar e X representa uma direção (**vetor**) que é preservada após a transformação $Y = AX$. Na presente seção será apresentada uma técnica de solução destes sistemas, bem como as devidas interpretações a serem tomadas.

Definição: Se A é uma matriz quadrada de ordem N x N, então o vetor não nulo X em R^n é chamado de **autovetor** (*eigenvector*) de A se o vetor $Y = AX$ for um múltiplo escalar do vetor X, ou seja, $AX = \lambda X$. O escalar λ é chamado de **autovalor** (*eigenvalue*) de A associado ao autovetor X.

A figura 3.7 mostra a interpretação geométrica para autovalores e autovetores.

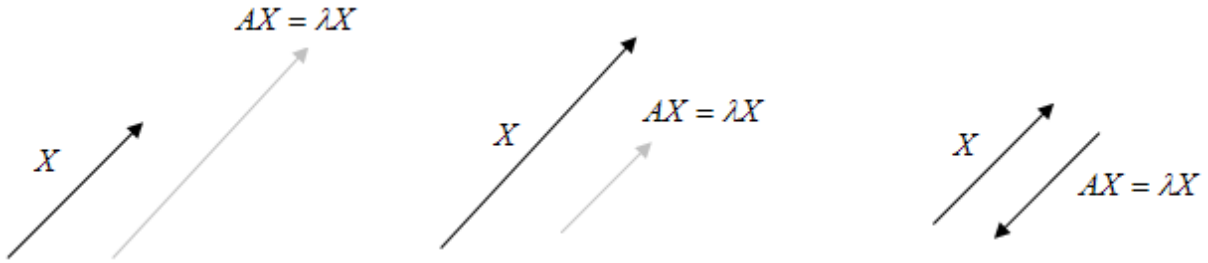


Figura 3.7: (a) Dilatação, (b) Contração, (c) Mudança de direção. A dilatação ocorre quando ($\lambda > 1$), para contração ($0 < \lambda < 1$) e por fim a mudança de direção é alcançada quando ($\lambda < 0$).

O cálculo dos autovalores e autovetores seguem os passos abaixo:

- Montar o sistema $AX = \lambda X$;
- Expandir para $AX = \lambda IX$, onde I é a matriz identidade;
- Subtrair λIX de ambos os lados de $AX = \lambda IX$, obtendo $(A - \lambda I)X = 0$;
- Calcular agora uma solução não trivial $X \neq 0$ para $(A - \lambda I)X = 0$. Da teoria de determinantes tem-se que $\det(A - \lambda I) = 0$ gera uma solução não trivial que satisfaz a equação $(A - \lambda I)X = 0$. A equação que envolve o determinante é chamada de **equação característica** de A.

- Calcular as raízes do polinômio característico;

$$\det(A - \lambda \mathrm{I}) = \lambda^N + c_1 \lambda^{N-1} + \cdots + c_N = 0 \tag{3.13}$$

Onde

- Os autovalores λ são as raízes do polinômio característico $\det(A - \lambda I) = 0$, com coeficientes, c, iguais a 1.

Abaixo segue um exemplo para um melhor esclarecimento

Considerando a matriz simétrica $A = \begin{bmatrix} 4 & 8 \\ 8 & 4 \end{bmatrix}$ e supondo que linhas (observações) diferentes correspondam a pessoas diferentes, onde a primeira coluna é referente à variável peso e a segunda coluna à variável altura, visualizando no R^2 tem-se:

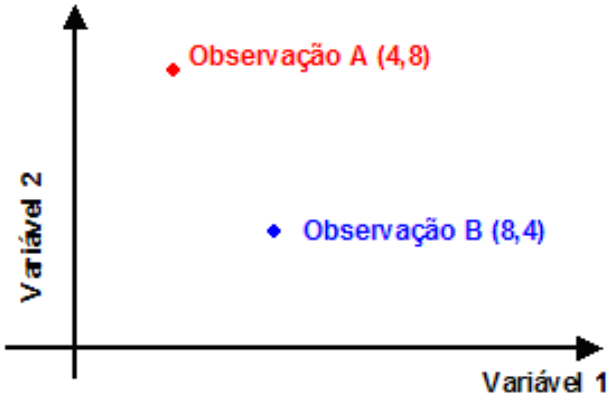


Figura 3.8: Visualização no R².

Os autovetores e autovalores da matriz A serão obtidos pela solução de equação:

$$\det\begin{bmatrix} 4-\lambda & 8 \\ 8 & 4-\lambda \end{bmatrix} = (4-\lambda)^2 - 64 = 0 \qquad (3.14)$$

Para a equação 3.14 as soluções são $\lambda_1 = 12$ e $\lambda_2 = -4$. Substituindo estes valores respectivamente em $(A - \lambda I)X = 0$ obtém-se os sistemas:

$$\begin{bmatrix} 4-\lambda_1 & 8 \\ 8 & 4-\lambda_1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{ cuja solução é } X_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ (autovetor).}$$

$$\begin{bmatrix} 4-\lambda_2 & 8 \\ 8 & 4-\lambda_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{ cuja solução é } X_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \text{ (autovetor).}$$

A figura 3.9 corresponde à inclusão dos autovetores $X_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ e $X_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ no gráfico da figura 3.8.

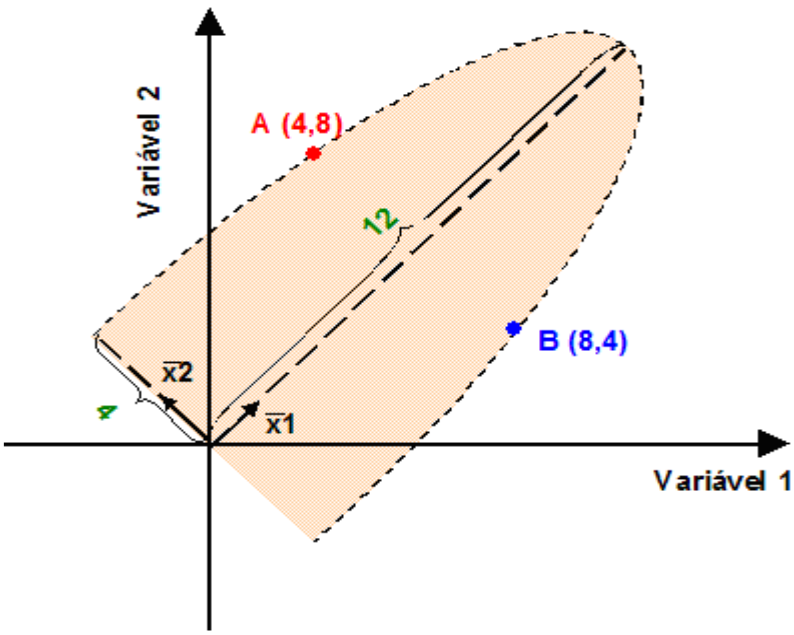


Figura 3.9: Visualização no R² com autovetores. A forma elíptica representa a existência de relação de dependência entre as variáveis. Os comprimentos dos eixos principal e secundário da elipse são dados pelos autovalores calculados na equação 3.14.

Existem dois casos gerais de grande importância, a saber:

Autovalores idênticos

Considere $A = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ cujos autovalores são $\lambda_1 = 4$ e $\lambda_2 = 4$. Os autovetores associados aos autovalores serão indeterminados, ou seja, podem ser quaisquer dois autovetores que satisfaçam a equação $(\lambda I - A)X = 0$, graficamente tem-se:

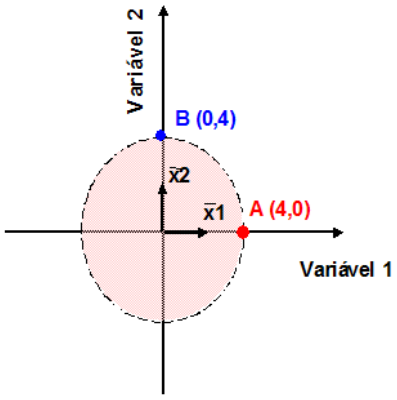


Figura 3.10: Autovalores Idênticos. A forma circular representa a não existência de dependência linear, em outras palavras o sistema de equações é linearmente independente, sendo assim as observações A e B não carregam informações redundantes.

Autovalor nulo, ou muito próximo de zero

Considere $A = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$, onde os autovalores serão $\lambda_1 = 8$ e $\lambda_2 = 0$. Os autovetores associados serão $X_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ e $X_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ respectivamente, graficamente tem-se:

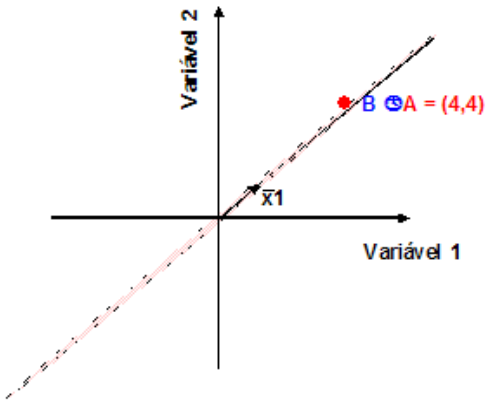


Figura 3.11: Autovalor Nulo. A forma linear representa forte dependência entra as variáveis analisadas, em outras palavras, as observações carregam informações redundantes.

3.3 Técnicas de redução de dimensionalidade

No mundo moderno é cada vez maior o número de pessoas que estão conectadas na internet enviando e consumindo conteúdos dos mais diversos tipos tais como textos, áudios e vídeos. Toda esta gama de dados quando processada em nível de hardware ou software é comumente representada via as estruturas de dados conhecida como matrizes e vetores, ambos explicados no capítulo 2.

Um problema recorrente ao se representar um dado (amostra) por meio de um vetor é que em muitas situações o tamanho do vetor fica tão grande que torna inviável o processamento matemático do mesmo a fim de se extrair quaisquer informações da amostra.

As técnicas de redução de dimensionalidade foram desenvolvidas justamente para servirem como ferramentas para sobrepujar a limitação anteriormente apresentada. Por meio de transformações de redução de dimensionalidade um vetor pode ser representado com seu comprimento reduzido, isto não quer dizer que o mesmo foi em parte cortado. O que o processo de transformação faz é, por exemplo, pegar um vetor no R^n aplicar uma transformação (ANTON e RORRES, 2005) do tipo $T: R^n \rightarrow R^m$, onde $m < n$, chegando assim a um vetor de comprimento menor.

O objetivo é que o vetor estando representado em uma dimensão reduzida possa ser processado tal que as informações intrínsecas a esta dimensionalidade possam ser extraídas, em outras palavras, informações que não são visíveis em uma dimensão superior tornam-se visíveis em dimensões reduzidas. Como conseqüência as tarefas de classificação, visualização e compressão de dados tornam-se mais fáceis em dimensionalidade reduzidas.

3.3.1 PCA (Principal Components Analysis)

A Análise de Componentes Principais também conhecida como transformada de Karhunen-Loève (KIRB e SIROVICH, 1990) é uma técnica oriunda da estatística e tem encontrado aplicações em áreas como reconhecimento de face e compressão de imagens. Ela tem se mostrado uma ferramenta muito poderosa na procura por padrões em dados de alta dimensão, em outras palavras, ela é muito utilizada para realização de **redução de dimensionalidade.**

Análise dimensional

Com o objetivo de apresentar o conteúdo de forma didática optou-se por iniciar as explicações por meio do exemplo a seguir.

Dado um questionário com 7 perguntas, cujas respostas serão numéricas apenas para facilitar as explicações, este por sua vez é submetido para um grupo de 4 pessoas. A representação matricial para o enunciado é:

$$A = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & x_{1,5} & x_{1,6} & x_{1,7} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & x_{2,5} & x_{2,6} & x_{2,7} \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & x_{3,5} & x_{3,6} & x_{3,7} \\ x_{4,1} & x_{4,2} & x_{4,3} & x_{4,4} & x_{4,5} & x_{4,6} & x_{4,7} \end{bmatrix}_{4x7} \tag{3.15}$$

Onde cada linha corresponde uma pessoa (observação), ou melhor, às respostas dadas por uma pessoa a todas as perguntas do questionário, e cada coluna representa as respostas dadas para uma mesma pergunta (variável), ou seja, respostas que diferentes indivíduos deram a uma mesma pergunta.

O objetivo agora é construir uma matriz S_{Nx7} tal que $R = S_{Nx7} \cdot A_{7x4}^{T}$. Considerando $N < 7$, por exemplo, $N < 5$ a transformação que estará sendo realizada é chamada de **redução de dimensionalidade**, pois resultará em R_{5x4}.

Tomando por $B = R^{T} = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & x_{1,5} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & x_{2,5} \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & x_{3,5} \\ x_{4,1} & x_{4,2} & x_{4,3} & x_{4,4} & x_{4,5} \end{bmatrix}_{4x5}$, verifica-se que a matriz resultante B possui menos colunas que A denotando assim a redução de dimensionalidade.

O problema inverso

Pensando no problema anterior de maneira inversa, a matriz B representa que 4 pessoas responderam um questionário com 5 perguntas, ou seja, um questionário **MENOR**, sendo assim, uma observação (respostas dadas pela mesma pessoa) representada em A possui mais variáveis do que quando esta mesma observação é representada em B, logo as observações em A estão em uma dimensão maior do que quando representadas em B.

A função da matriz S é mapear (transformar) as observações de A para suas correspondentes em B. Cabe aqui ressaltar que o "**questionário**" em B não é gerado pegando o "questionário" em A e jogando duas perguntas fora.

Uma linha na matriz B representa um novo questionário, menor, com novas perguntas e que não são as mesmas que estão em A, mas que procuram carregar em si a mesma quantidade de informação, em outras palavras, busca-se através de 5 novas perguntas obter o mesma quantidade de informação do que com as 7 perguntas antigas. Outra característica importante é que agora as variáveis (perguntas) em B serão menos correlacionadas do que em A.

O exposto acima corresponde à teoria que está por traz não só do PCA, mas também de muitas outras técnicas de redução de dimensionalidade (MAATEN e POSTMA, 2009) que por sua vez são amplamente utilizadas em reconhecimento facial.

PCA e suas equações

O objetivo do PCA nada mais é do que encontrar uma matriz de transformação que não só mapeie os elementos de A em B com o mínimo de perda de informação mais também que as variáveis resultantes sejam menos correlacionadas.

Abaixo seguem os passos necessários para o cálculo do PCA.

- Montar a matriz D_{NxM}, onde N corresponde ao número de pessoas submetidas ao "questionário", M é número de perguntas no questionário e cada coluna representa uma variável.

$$D = \begin{bmatrix} x_{1,1} & x_{1,2} & - - & x_{1,M} \\ x_{2,1} & - & - - & - \\ - & - & - - & x_{(N-1),M} \\ x_{N,1} & - & - - & x_{N,M} \end{bmatrix}_{\mathrm{NxM}} \tag{3.16}$$

- Calcular as médias de cada coluna, onde $\bar{x}_1$ corresponde ao valor médio dos elementos da primeira coluna e $\bar{x}_M$ ao da m-ésima coluna.
- Subtrair o valor médio da coluna $\bar{x}$ em todos os elementos da respectiva coluna; a este processo dá se o nome de **normalização**, gerando a matriz E cujas variáveis possuirão valor médio igual a zero.
- Calcular a matriz de covariância da matriz E;

$$C = \begin{bmatrix} \mathrm{cov}(x,x) & \mathrm{cov}(x,y) & \mathrm{cov}(x,z) \\ \mathrm{cov}(y,x) & \mathrm{cov}(y,y) & \mathrm{cov}(y,z) \\ \mathrm{cov}(z,x) & \mathrm{cov}(z,y) & \mathrm{cov}(z,z) \end{bmatrix} \tag{3.17}$$

Onde

- $\mathrm{cov}(y,z)$: Covariância entre as variáveis y e z da matriz E. Note que a matriz C é quadrada.

Pelo fato da matriz C ser quadrada torna-se possível calcular seus autovalores e autovetores associados. Os autovetores serão unitários e perpendiculares dois a dois conforme figuras 3.9 e 3.10.

- Calcular os autovetores e autovalores da matriz de covariância $\mathrm{C}_{\mathrm{MxM}}$, esta etapa vai gerar M autovetores, eig_n, e M autovalores associados, λ_n;
- Montar a matriz de autovetores (matriz de transformação);

$$Eig = \left[eig_1 \; eig_2 \; eig_3 \cdots eig_n \right] \tag{3.18}$$

Ordenando-os segundo a ordem decrescente de autovalores. Esta matriz é que será responsável pelo mapeamento entre dimensões. Para aplicar a redução de dimensionalidade basta "jogar fora" alguns autovetores associados aos menores autovalores; a quantidade depende do quanto se deseja de redução de dimensionalidade.

- Aplicar transformação a seguir;

$$FinalEig = Eig^T \cdot E^T \tag{3.19}$$

Onde

- E^T : Base de treinamento na qual as colunas são as observações e as linhas são as variáveis;
- $FinalEig$: Matriz resultante ou matriz de referência cujas colunas representam os dados mapeados com dimensões reduzidas.

3.3.2 Técnicas convexas, não convexas e suas ramificações

A figura 3.12 mostra a taxonomia das técnicas de redução de dimensionalidade proposta por (MAATEN e POSTMA, 2009).

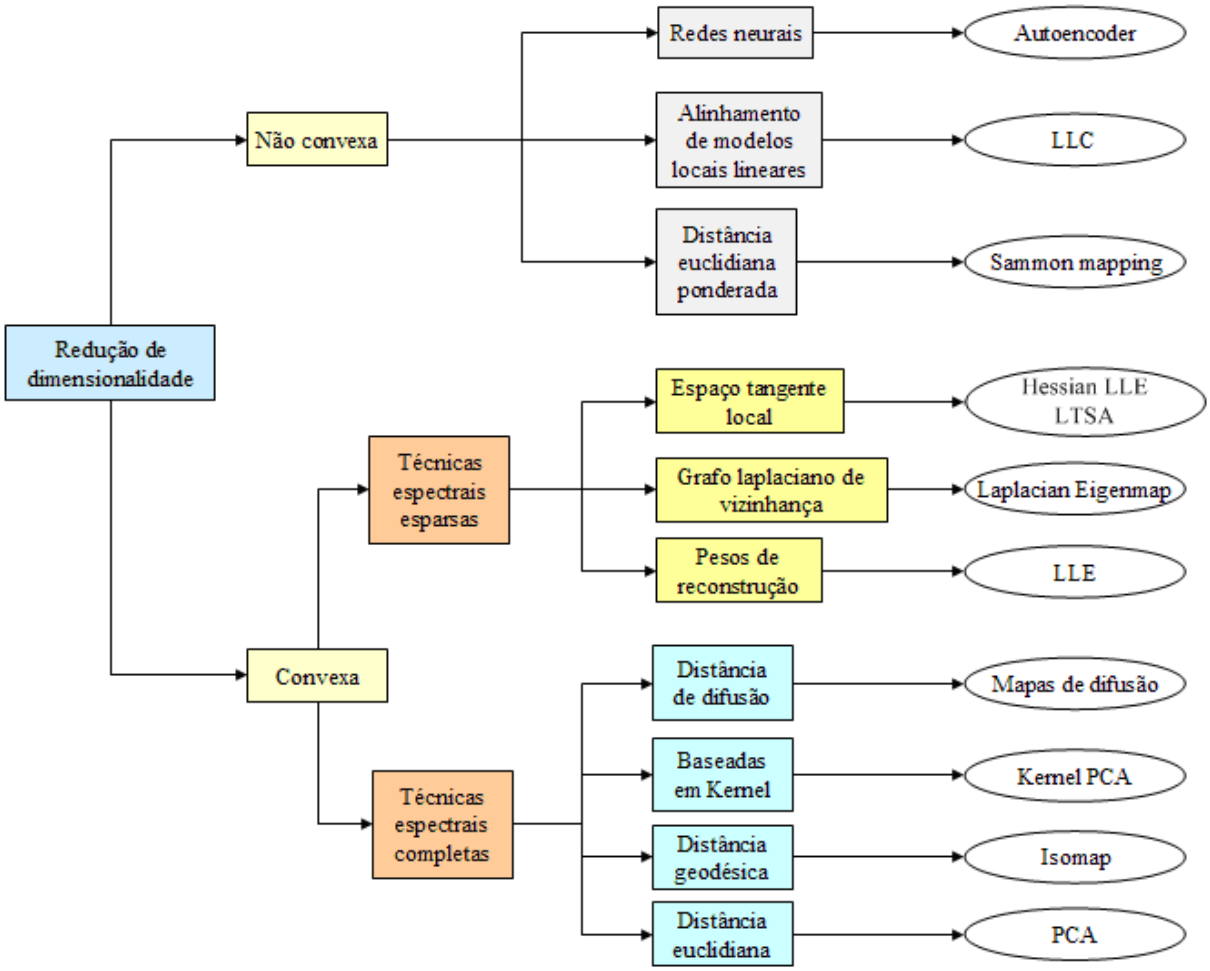


Figura 3.12: Diagrama de Taxonomia das técnicas de redução de dimensionalidade. Imagem adaptada de (MAATEN e POSTMA, 2009)

Para melhor compreensão das classes apresentadas na figura 3.12 abaixo segue a lista com a descrição das principais classes representadas na figura em questão.

- **Convexa:** São técnicas cuja função a ser otimizada é do tipo $\Phi(Y) = \frac{Y^T AY}{Y^T BY}$, que por sua vez não possui um ótimo local, em outras palavras, seu espaço solução é convexo (BOYD e VANDENBERGHE, 2004) e podem ser solucionadas via problema dos autovetores generalizados, capítulo 4. A maioria das técnicas de redução de dimensionalidade pertence a essa classe.

- **Não convexa:** São aquelas cuja função a ser otimizada é não convexa, bem como o espaço solução da mesma;
- **Técnicas espectrais completas:** São técnicas que constroem sua matriz de transformação com base na decomposição em autovalores e autovetores da matriz de covariância extraída a partir da massa de dados de treinamento. Como exemplo tem-se PCA, Isomap, Kernel PCA e os Mapas de Difusão (MAATEN e POSTMA, 2009).
- **Técnicas espectrais esparsas:** São técnicas cuja construção da matriz de transformação necessita da solução do problema dos autovetores generalizados. Como exemplo tem-se *Linear Discriminant Analysis* (LDA), *Local Linear Embedding* (LLE), *Laplacian Eigenmap*, Hessian LLE e *Local Tangent Space Analysis* (LTSA), (MAATEN e POSTMA, 2009).
- **Distância Euclidiana**, **Distância Geodésica** e **Distância Geométrica** são exemplos de medidas de similaridade entre dois padrões, ou seja, através das mesmas é possível medir o grau de similaridade entre dois vetores.

Nos capítulos 4 e 5 serão apresentados os métodos de reconhecimento facial a serem comparados no presente trabalho. É importante ressaltar que tais métodos tiveram origem nas seguintes técnicas de redução de dimensionalidade, a saber: *Principal Components Analysis* (PCA) deu origem ao método Eigenface, Kernel PCA deu origem à *Local Feature Analysis* (LFA), *Linearity Preserving Projection* (LPP) deu origem à Laplacianfaces e por fim *Linear Discriminant Analysis* (LDA) deu origem ao método Fisherfaces.

3.4 Sistemas biométricos

Um sistema biométrico é aquele que é capaz de extrair padrões biométricos de uma pessoa visando à identificação ou autenticação da mesma; tais padrões são medidas fisiológicas ou comportamentais de um indivíduo.

Segundo (JAIN e ROSS et al., 2004) para que um padrão biométrico possa ser utilizado em sistemas de reconhecimento é necessário que o mesmo atenda aos requisitos de universalidade, exclusividade, permanência, facilidade de coleta, resistência a fraudes, desempenho e aceitação.

- **Universalidade:** Corresponde a uma determinada característica pertencente a todas as pessoas;

- **Exclusividade:** Significa que a característica é diferente em cada indivíduo, logo pode ser utilizada para identificá-lo;
- **Permanência:** É quando o padrão biométrico não muda com o tempo, sua variação ocorra em um tempo muito longo;
- **Facilidade de coleta:** Quando não precisa de aparatos muitos rebuscados para ser obtido, por exemplo, para se obter a imagem de uma pessoa basta se ter uma câmera fotográfica não havendo a necessidade de ligar nenhum equipamento no indivíduo;
- **Resistência a fraudes:** São características de difícil alteração proposital;
- **Desempenho:** São características de fácil processamento computacional, permitindo assim que o sistema seja mais rápido e preciso em suas respostas;
- **Aceitação:** É referente a dados biométricos onde as pessoas não se sentem incomodadas em fornecê-los.

No mundo moderno devido ao grande crescimento populacional e ao aumento da renda familiar, está surgindo uma nova classe consumidora, ávida por novos serviços e tecnologias. É cada vez maior o número de pessoas viajando de ônibus, trem, avião; comprando celulares e computadores; fazendo compras com cartões de crédito e online; indo a teatro, shows e estádios de futebol.

Todos estes novos hábitos e tendências trazem consigo novas preocupações como, por exemplo, o desenvolvimento de sistemas mais robustos de segurança, controle e monitoramento. Estes ambientes são, na sua maioria, favoráveis ao uso de soluções biométricas. Nas próximas seções serão descritos alguns sistemas biométricos dando ênfase às situações onde a aplicação dos mesmos traz grandes benefícios à população.

3.4.1 Reconhecimento de Iris

Quando é dito que alguém possui olhos verdes, azuis e castanhos está sendo feito referência a um músculo do olho conhecido como íris. O uso de padrões biométricos extraídos do íris data do século XIX (BERTILON, 1885), contudo um dos primeiros sistemas automáticos de reconhecimento de pessoas através de padrões extraídos das imagens da íris e que obteve grande aceitação foi desenvolvido por (DAUGMAN, 1993).

Um dos grandes obstáculos para o correto funcionamento deste sistema é a obtenção de imagens de qualidade, pois para consegui-las é necessário que a pessoa a ser reconhecida

ou identificada aproxime seus olhos dos dispositivos de captura de imagem, isto mostra a grande dependência da colaboração do indivíduo.

Em (DAUGMAN, 1993) após a captura da imagem do olho, a íris é localizada via processamento de imagem, aplicando a técnica de extração de bordas buscando por bordas circulares. Depois de localizar a forma circular correspondente à íris, é realizada uma transformação matemática utilizando coordenadas polares; o objetivo é associar o modelo circular a um sistema de coordenadas retangulares, minimizando com isso as possíveis mudanças de escala na imagem da íris.

Estando a íris representada por um modelo retangular (MOUTINHO, 2005) este será codificado através do filtro de Gabor, o qual apresentará como resultado um vetor numérico correspondente que representa as características extraídas da imagem da íris de um indivíduo. O processo final de reconhecimento é feito pela comparação de tal vetor com todos os outros vetores de características previamente armazenados em um banco de dados. Para comparar dois vetores são utilizadas diversas medidas de similaridade, como por exemplo, distância euclidiana, Mahalanobis, Nearest-Neighbor e Hamming, (THEODORIDIS, 2009).

3.4.2 Reconhecimento de assinaturas

Os sistemas de reconhecimento automático de assinaturas ou padrões de escrita podem ser utilizados em situações como verificação de assinaturas em cheques, contratos e busca por padrões de escrita em cartas manualmente redigidas.

Como exemplo de aplicação bem sucedida de tal tecnologia tem-se os dispositivos com telas sensíveis ao toque, onde um indivíduo por meio de uma caneta com ponta de borracha escreve sua assinatura sobre a tela do dispositivo e esta é capturada, processada e comparada com as demais previamente armazenadas em um banco de dados. Esta solução é usada por muitos bancos para verificar se uma dada pessoa está de fato autorizada a executar transações financeiras na correspondente agência.

As características comumente procuradas nas assinaturas são:

- Proporção de ocupação das palavras;
- Espaços e regularidade na escrita;
- Pressão sobre o papel, ângulo da caneta e tempo de escrita (no caso de dispositivos de captura online da escrita) (TAPPERT e WAKAHARA, 1990).

3.4.3 Reconhecimento de voz

Os sistemas de reconhecimento de voz são divididos em duas categorias: os baseados em textos e os independentes do texto. No primeiro caso é necessário que um indivíduo repita o texto várias vezes para que dados como freqüência, potência e tempo médio de sinal de voz possam ser extraídos e cadastrados para posteriormente serem utilizados durante o processo de identificação de uma dada pessoa. Os independentes de textos ainda representam um grande desafio. Em (RUDASI, 1992) é apresentado uma abordagem para este tipo de sistema.

O reconhecimento por voz ainda apresenta grandes obstáculos, como a mudança da voz da pessoa com o passar dos anos, rouquidão temporária e variação no tempo de pronúncia de uma palavra, ou melhor, de um fonema são desafios a serem resolvido.

As técnicas mais utilizadas em sistemas de reconhecimento de voz são: *Dinamic Time Warping* (SHARIN, 1998), Modelos Escondidos de Markov - HMM (RABINER, 1989), Modelos de Mistura de Gaussianas - GMM (SCHMIDIT et al., 1997) e por fim as Redes Neurais - NN (FRANZINI e WITBROK et al., 1989).

3.4.4 Reconhecimento de Impressões digitais

As impressões digitais são características formadas desde a fase embrionária e se mantém praticamente constante ao longo de toda a vida. A partir de uma imagem da impressão digital os sistemas automáticos de identificação extraem suas características, também conhecidas como minúcias.

Os pontos de minúcias correspondem aos pontos onde ocorrem descontinuidades nas linhas da impressão digital, linhas papilares. Existem dois tipos de minúcias normalmente encontradas, a de término de linha e a de bifurcação de linhas, figura 3.13.

Figura 3.13: Parte da imagem de uma impressão digital. A figura à esquerda corresponde a um exemplo de término de linha, já a figura da direita apresenta uma bifurcação (MOUTINO, 2005).

Existem diversos métodos de extração de minúcias, em (MARQUES, 2004) é proposto um sistema automático de extração de minúcias baseado em redes neurais. Nele as

etapas de pré-processamento são: obtenção da versão em tons de cinza da imagem da impressão digital, binarização e por fim o afinamento.

O processo de binarização é simples, pois a partir de uma imagem em tons de cinza, ou seja, que os valores de seus pixels estejam compreendidos entre 0 e 255, é tomado um valor com referência (threshold). De posse deste valor percorrem-se todos os pixels da imagem fazendo a comparação destes com o threshold, se o valor do pixel for inferior ao do threshold o pixel receberá 0 com sendo seu novo valor, caso contrário receberá 255. Como resultado a imagem final terá pixels apenas com valores 0 ou 255.

A etapa final de pré-processamento é aplicar o afinamento, que faz com que as linhas possuam largura de 1 pixel. Nesta etapa os pontos de término ou de bifurcação ficarão mais fáceis de serem encontrados. O vetor de características será construído com as coordenadas x e y das minúcias. Quanto maior o número de minúcias encontradas maior será o vetor de características facilitando assim a etapa de identificação.

3.5 O reconhecimento Facial

O crescimento do poder computacional dos equipamentos de captura e processamento de imagem, bem como o barateamento dos mesmos, têm possibilitado que mais grupos desenvolvam seus projetos de pesquisas na área do reconhecimento facial. As pesquisas não são impulsionadas apenas pelos desafios fundamentais do problema de reconhecimento (variação da pose, iluminação, expressão facial, oclusão e uso de acessórios como óculos), mas também pela demanda da sociedade por soluções como: sistemas automáticos de autenticação e sistemas de interação homem máquina. Aplicar o reconhecimento facial traz como diferencial a possibilidade de se desenvolver sistemas menos intrusivos pois para a captura da imagem não há a necessidade de contato físico entre o usuário o sistema.

Entenda autenticação como comparação um a um, onde a imagem de entrada é comparada com outra imagem para verificar se determinado indivíduo é quem ele diz ser. Já no caso da identificação é quando uma imagem é comparada com todas as imagens de um banco de dados para encontrar com quem aquele indivíduo mais se parece. Em ambos os casos o algoritmo precisar receber um threshold (valor de referência) indicando o máximo erro aceitável, pois somente assim ele é capaz de tomar a decisão durante o processo de autenticação ou identificação.

A performance dos sistemas de reconhecimento facial tem melhorado desde os primeiros trabalhos desenvolvidos por (KANADE, 1973), principalmente em situações com ambiente controlado.

Etapas do processo de reconhecimento facial

As figuras 3.14 a 3.16 descrevem as etapas do processo de reconhecimento facial.

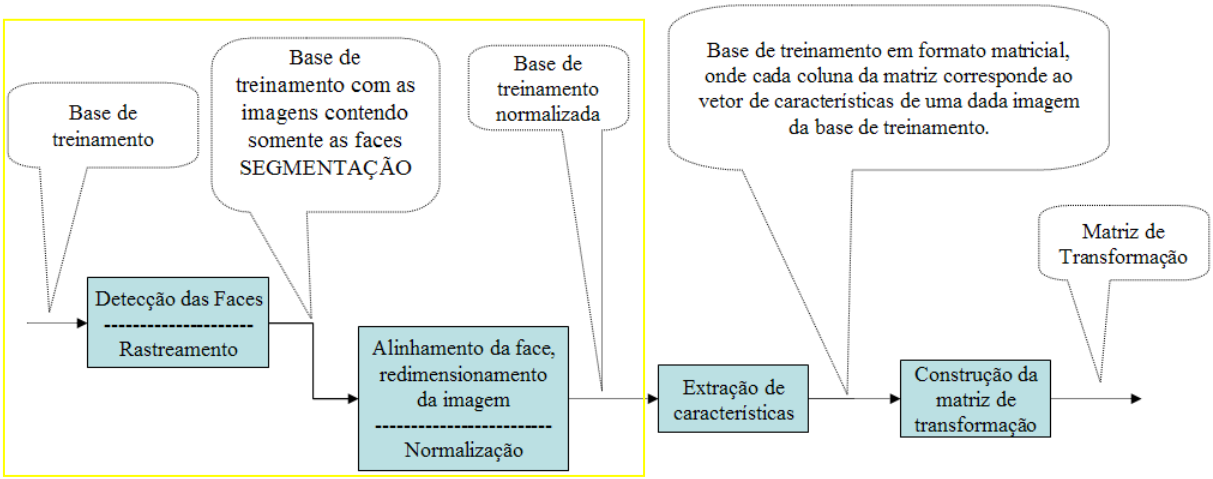


Figura 3.14: Construção da matriz de transformação. O diagrama apresentado mostra de maneira simplificada as etapas necessárias para a construção da matriz de transformação dado uma base de imagens para treinamento.

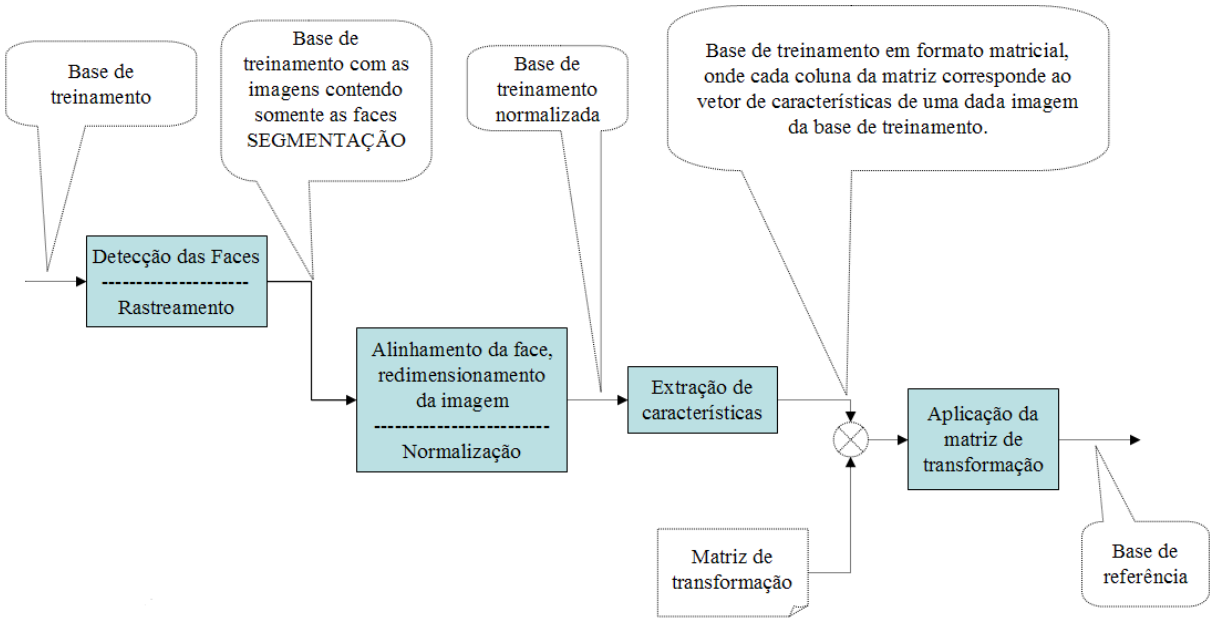


Figura 3.15: Construção da base de referência. A figura em questão apresenta os módulos necessários para construção da base de referência, esta base é que será posteriormente utilizada pelos classificadores de padrões.

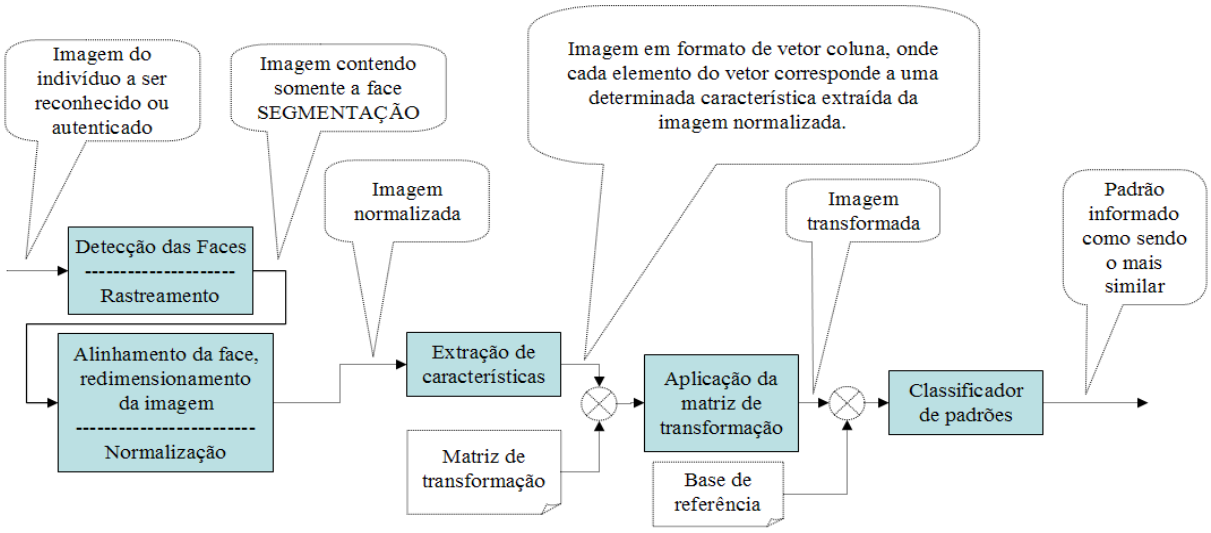


Figura 3.16: Realização do reconhecimento. O diagrama apresentado em tal figura modela todo o processo de reconhecimento/ identificação de um indivíduo.

Pré-processamento:

Esta etapa é compreendida por: Aquisição da imagem, localização da face, extração do fundo, recorte (segmentação) apenas da região de interesse, rotação da imagem, aumento ou redução de escala e por fim aplicação de técnicas variadas de processamento de imagem.

Como exemplo de técnicas de processamento de imagem tem-se: aplicação de filtros no domínio do espaço ou da freqüência para minimizar a dependência da variação da iluminação, filtros para realçar bordas e filtros para eliminar ruídos que podem ter sido introduzidos pelos equipamentos de aquisição. Conforme a figura 3.14 saída da etapa de pré-processamento (bloco em amarelo) são as imagens normalizadas da base de treinamento.

Transformação:

É a etapa em que a matriz imagem é transformada em um vetor coluna e sobre a mesma aplicam-se as matrizes de transformação, como exemplos transformações têm-se as matrizes de redução de dimensionalidade.

A etapa de transformação compreende a parte central dos métodos de reconhecimento facial, os quais serão apresentados ao longo dos capítulos 4 e 5. A Tabela 3.2 traz uma visão resumida dos métodos de reconhecimento facial divididos nas abordagens holística, baseada em característica e por fim na abordagem híbrida.

Pós-processamento e reconhecimento

Corresponde à etapa em que a imagem transformada, imagem a ser reconhecida, e mais a base de referência são submetidos aos classificadores de padrões para que estes, através de um processo de medição de similaridade, possam informar qual “imagem” da base

de referência é mais similar à imagem a ser reconhecida. Existem diversos tipos de classificadores, como exemplo, o Bayesiano, o Euclidiano, Nearest-Neighbor e as Redes Neurais.

De um modo geral um classificador precisa receber um threshold e uma medida de similaridade para que com base nestes possa tomar a decisão no momento de informar se existe ou não na base de referência algum padrão similar à imagem a ser reconhecida. A medida de similaridade mais utilizada é a distância euclidiana.

Tabela 3.2: Categorização de técnicas de reconhecimento facial (ZHAO e CHELLAPA, 2003).

Abordagens	Trabalhos representativos
Métodos holísticos	
Principal component analysis (PCA)	
Eigenfaces	Direct application of PCA (TURK e PENTLAND, 1991)
Probabilistic eigenfaces	Two-class problem with prob. measure (MOGHADDAM e PENTLAND, 1997)
Fisherfaces/subspace LDA	FLD on eigenspace (BELHUMEUR et al., 1997)
SVM	Two-class problem based on SVM (PHILLIPS, 1998)
Evolution pursuit	Enhanced GA learning (LIU e WECHSLER 2000)
Feature lines	Point-to-line distance based (LI e LU, 1999)
ICA	ICA-based feature analysis (BARTLETT et al., 1998)
Outras representações	
LDA/FLD	LDA/FLD on raw image (ETEMAD e CHELLAPPA, 1997)
PDBNN	Probabilistic decision based NN (LIN et al., 1997)
Métodos baseados em características	
Métodos puramente geométricos	Earlier methods (Kanade, 1973) Recent methods (COX et al. 1996)
Arquitetura de link dinâmico	Graph matching methods (OKADA et al., 1998)
Modelo de Markov	HMM methods (NEFIAN e HAYES, 1998)
Redes Neurais	SOM learning based CNN methods (LAWRENCE et al., 1997)
Métodos híbridos	
Modular eigenfaces	Eigenfaces and eigenmodules (PENTLAND et al., 1994)
LFA híbrido	Local feature method (PENEV e ATICK, 1996)
Forma normalizada	Flexible appearance models (LANITIS et al., 1995)
Baseados em componentes	Face region and components (HUANG et al. 2003)

4 Os métodos de reconhecimento facial

Conforme apresentado na Tabela 3.2 os métodos de reconhecimento facial podem ser divididos em quatro abordagens. Optamos por seguir esta mesma estrutura, sendo assim, ao longo do capítulo 4 apresentaremos uma breve descrição de cada abordagem.

4.1 Os métodos baseados em aparência (métodos holísticos)

Os métodos holísticos (ZHAO e CHELLAPPA et al., 2003) são aqueles ditos baseados na informação da imagem, ou seja, não precisam tomar medidas geométricas como distância olho-olho, olho-boca, nariz-boca etc. Eles usam a imagem como um todo e aplicam transformações sobre ela. As etapas de organização da informação para posterior transformação serão detalhadas ao longo desta seção.

A maioria dos métodos são variações do PCA, o qual foi inicialmente proposto por (KIRB e SIROVICH, 1990) com o nome de Karhunen-Loève com o foco de servir como uma técnica geral de redução de dimensionalidade, não tendo sido desenvolvida para uso exclusivo em reconhecimento facial. Posteriormente (TURK e PENTLAND, 1991) estenderam a técnica aplicando-a em reconhecimento de face.

Ao longo desta seção serão descritos alguns destes métodos, os quais foram escolhidos por serem os fundamentais e que deram origem a uma série de variações.

4.1.1 Eigenfaces

Eigenfaces é um método holístico com origens no PCA que foi inicialmente proposto por (TURK e PENTLAND, 1991) e seu objetivo era o de fornecer um algoritmo rápido, de baixa complexidade computacional, capaz de realizar a tarefa de reconhecimento dado um ambiente controlado, ou seja, imagens contendo predominantemente a face de uma pessoa com pequenas variações de pose, iluminação e rotação.

Matematicamente falando o processo para o cálculo dos eigenfaces consiste em pegar um conjunto de imagens para treinamento, passá-las para tons de cinza e transformá-las em vetores colunas. Estes vetores são vistos como pontos em um espaço de dimensão n, onde n é o comprimento do vetor. Assim como um ponto no R^2 é representado por duas coordenadas, por exemplo, (x_1, x_2), no R^n um ponto é representado por n coordenadas $(x_1, x_2, \cdots, x_n)$.

Os vetores de imagem serão organizados lado a lado em uma grande matriz de treinamento. Da matriz de treinamento é calculada sua matriz de covariância que, por sua vez,

é usada no cálculo do PCA, que ao final do processo da origem a uma matriz de transformação responsável por implementar a redução de dimensionalidade. As colunas da matriz de transformação são também conhecidas como eigenfaces.

4.1.1.1 O algoritmo

Dado uma imagem de entrada P de dimensões FxJ são aplicadas sobre a mesma algumas operações a título de pré-processamento, tais como rotação, equalização do histograma, redimensionamento e por fim transformação para tons de cinza, gerando assim uma nova imagem I de mesmas dimensões ou não.

A imagem, já em tons de cinza e com dimensões FxJ, é transformada no vetor coluna Γ de comprimento (FxJ). Tal vetor é construído concatenando pela direita as linhas da matriz I, formando assim um grande vetor linha, a transposta deste dará origem à Γ. Com a finalidade de simplificar a notação ao longo do trabalho a matriz P será considerada quadrada e de ordem NxN gerando com isso Γ de comprimento N^2.

Figura 4.1: Imagem (P) em RGB e sua correspondente (I) em tons de cinza.

Figura 4.2: Vetor Linha (Γ^T).

O procedimento descrito acima é aplicado em todas as imagens do conjunto de treinamento. Os vetores coluna gerados são agrupados lado a lado formando a matriz de treinamento B. Cabe ressaltar que a matriz B não possui imagens de apenas uma pessoa, mas de todas que comporão a BASE DE REFERÊNCIA.

$$B = [\Gamma_1 \quad \Gamma_2 \quad . \quad . \quad . \quad \Gamma_M] \tag{4.1}$$

Onde

- M : Número de imagens em B.

A matriz B passará agora por um processo de normalização que consiste em calcular seus valores médios por linha, $\Psi = [\Psi_1, \Psi_2, \cdots, \Psi_{N^2}]^T$. Estes valores serão subtraídos de todos os elementos das respectivas linhas, gerando a matriz A, já normalizada tendo média zero por linha.

$$\Phi_i = \Gamma_i - \Psi \text{, (imagem normalizada)} \tag{4.2}$$

$$A = [\Phi_1 \quad \Phi_2 \quad . \quad . \quad . \quad \Phi_M] \text{, onde } A_{N^2 xM} \tag{4.3}$$

De posse de A calcula-se sua matriz de covariância $C(x, y) = \frac{1}{M} A \cdot A^T$, onde $C_{N^2 xN^2}$ e para esta são encontrados os autovalores e autovetores associados.

Os autovetores são concatenados lado a lado, cada um representando uma coluna de comprimento N^2, formando a matriz de autovetores $X = \lfloor u_1 \quad u_2 \quad \cdots \quad u_{N^2} \rfloor$; as colunas de X são chamadas de eigenfaces. A concatenação dos eigenfaces se dá seguindo a ordem decrescente dos autovalores, ou seja, a primeira coluna de X corresponde ao autovetor associado ao maior autovalor.

A transposta de X, X^T, é a matriz de transformação responsável por aplicar a redução de dimensionalidade das imagens contidas em A, onde tal redução é alcançada pelo produto $\Omega_{ref} = X^T \cdot A$.

A quantidade de linhas mantidas em X^T corresponde ao número de dimensões do novo espaço sobre o qual as imagens em A serão projetadas. Cada linha em X^T pode ser interpretada como sendo um eixo de coordenadas neste novo espaço dimensional.

Em eigenfaces não existe uma regra que determine quantas nem quais colunas de X serão eliminadas para então formar X^T; é de costume escolher as colunas associadas aos menores autovalores. Da teoria de autovalores e autovetores é sabido que quanto menor for um autovalor mais próximos ficarão os pontos, imagens, projetados na direção do autovetor

associado a tal autovalor. Em outras palavras, o autovetor em questão não será uma boa direção de projeção para problemas cujo objetivo final seja a classificação de dados. Quando uma linha de X^T é multiplicada por uma coluna (imagem) de A obtém-se um número, também chamado de peso ω_i, que corresponde ao valor da projeção da imagem, Φ, sobre o referido eixo.

Quando todas as linhas de X^T são multiplicadas por uma única coluna de A, o que se está fazendo é o cálculo das projeções desta coluna sobre todos os eixos representados por X^T. O vetor coluna, Ω, formado por todas estas projeções nada mais é do que a imagem representada na nova dimensão.

A redução de dimensionalidade ocorre se o número de linhas de X^T for menor que o número de elementos em um vetor coluna de A, pois, após a multiplicação deste por todas as linhas de X^T, a nova imagem terá menos elementos do que tinha quando representada em A.

Após o cálculo de X^T, a base de referência $\Omega_{ref} = [\Omega_1 \quad \Omega_2 \quad \cdots \quad \Omega_M]$ é encontrada a partir do produto $\Omega_{ref} = X^T \cdot A$.

4.1.1.2 Interpretando as equações

Para melhor compreensão da abordagem por Eigenfaces segue abaixo uma lista de equações com as devidas descrições.

- I : Matriz bidimensional (imagem) NxN de 8 bits de intensidade, ou seja, I é uma imagem em tons de cinza, cujos tons podem variar de 0 à 255.
- Γ_n: Imagem representada como um vetor coluna de comprimento N^2.
- $B = [\Gamma_1 \quad \Gamma_2 \quad . \quad . \quad . \quad \Gamma_M]$: Matriz com imagens de treinamento com a média ainda não normalizada.
- $\Psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n$: Vetor com valores médios por linha;
- $\Phi_i = \Gamma_i - \Psi$: Vetor imagem com média normalizada;
- $A = [\Phi_1 \quad \Phi_2 \quad . \quad . \quad . \quad \Phi_M]$: Matriz com imagens de treinamento com média zero;
- $C(x,y) = \frac{1}{M}\sum_{n=1}^{M}\Phi_n(x)\cdot\Phi_n^T(y) = \frac{1}{M}A\cdot A^T$: Matriz de covariância de A, onde $C_{N^2xN^2}$.

- λ_n: Autovalor de C;
- u_n: Autovetor de C associado ao autovalor λ_n. Os autovetores calculados são os que melhor descrevem os dados em A. Eles são ortonormais, logo:

$$u_l^T \cdot u_k = \delta_{lk} = \begin{cases} 1, & if \quad l = k \\ 0, & otherwise \end{cases} \tag{4.4}$$

Para o cálculo do autovalor λ_n tem-se:

$$\lambda_n = \frac{1}{M} \sum_{n=1}^{M} (u_k^T \cdot \Phi_n)^2 \tag{4.5}$$

O PROBLEMA DOS AUTOVETORES GENERALIZADOS

Se o número, M, de imagens, ou melhor, a quantidade de amostras em A for menor do que o comprimento, N^2, do vetor imagem Γ_n, $(M < N^2)$, então após o cálculo dos N^2 autovalores de C apenas M destes serão significativamente diferentes de zero.

Para contornar o problema do cálculo desnecessário de autovalores e autovetores associados é construída a matriz $L = A^T \cdot A$ de ordem MxM, onde $L_{mn} = \Phi_m^T \cdot \Phi_n$ e então, são calculados os M autovetores, v_l, de L.

Através de uma combinação linear entre os autovetores v e as imagens Φ chega-se aos M autovetores, u, associados aos M maiores autovalores de C.

A combinação linear é dada por:

$$u_l = \sum_{k=1}^{M} v_{lk} \cdot \Phi_k \ , \ l = 1, \cdots, M \tag{4.6}$$

O procedimento acima reduz a computação dos autovalores e autovetores da ordem de N^2 para M; na prática normalmente ocorre que $(M << N^2)$, justificando a abordagem via construção de L.

Após o cálculo dos autovetores u é montada a matriz $X = [u_1 \quad u_2 \quad \cdots \quad u_M]$, onde o número de colunas corresponde ao total de imagens do conjunto de treinamento.

Normalmente uma boa redução de dimensionalidade é atingida deixando-se apenas M' autovetores associados aos M' maiores autovalores de $L = A^T \cdot A$, sendo assim $X = \lfloor u_1 \quad u_2 \quad \cdots \quad u_{M'} \rfloor$. O passo de projeção das imagens contidas em A para o domínio representado por X^T é realizado por:

$$\Omega_{ref} = X^T \cdot A \tag{4.7}$$

Onde

- M : Número de imagens em $A = [\Phi_1 \quad \Phi_2 \quad . \quad . \quad . \quad \Phi_M]$;
- M' : Número de autovetores, u, de $L = A^T \cdot A$, mantidos em X ;
- $\Omega_l^T = [\omega_1 \quad \omega_2 \quad \cdots \quad \omega_{M'}]$: Imagem resultante da projeção da l- ésima coluna de A;

Fazendo

- $\omega_k = u_k^T \cdot \Phi_l$;
- $k = 1, \cdots, M'$ e $l = 1, \cdots, M$, $(M' \leq M)$.
- ω_k : É o peso resultante da projeção da l- ésima coluna de A sobre o eixo representado por u_k^T.
- Ω_{ref} : É a matriz resultante cujas colunas representam as respectivas imagens de A agora com a dimensão reduzida, pois $(M' << N^2)$.

4.1.1.3 Construção da base final de referência

Em (TURK e PENTLAND, 1991) são definidos os *thresholds*, $\theta_{original}$ e $\theta_{transformado}$, os vetores de distâncias $\in_{original}$ e $\in_{transformado}$, e uma versão modificada de Ω_{ref}, ou melhor, $\Omega_{ref-FINAL}$. Tal modificação é detalhada na explicação sobre o cálculo de Ω_k, equação 4.9.

$$\in^2_{original,f} = \left\| \Phi - \Phi_f \right\|^2 \text{ e } f = 1, \cdots, M \tag{4.8}$$

Quadrado da distância Euclidiana no domínio original.

Onde M é a quantidade de imagens na matriz A.

$$\in^2_{transformado,k} = \left\|(\Omega - \Omega_k)\right\|^2;\ k = 1,\cdots,G \text{ e } G < M \tag{4.9}$$

Quadrado da distância Euclidiana no domínio transformado.

Onde G é a quantidade de indivíduos (pessoas) diferentes na matriz A e Ω_k, neste caso, representa uma **IMAGEM MÉDIA** e não a projeção de uma única imagem. Para gerar Ω_k são realizados os seguintes passos:

- Formação do conjunto de treinamento, A, com mais de uma imagem por pessoa;
- Sobre o conjunto de treinamento são executadas todas as etapas até chegar em Ω_{ref};
- De Ω_{ref} são extraídas todas as colunas, que tenham sido geradas pelas imagens de uma mesma pessoa, e este grupo de colunas é substituído por uma única coluna Ω_k, a qual representa o vetor médio (padrão) das colunas anteriormente extraídas. O padrão Ω_k é representativo da k-ésima pessoa; este é o vetor que será guardado para posterior comparação com a imagem, Ω, esta já estando no espaço de dimensionalidade reduzida. A base final de referência resultante da manipulação das colunas de Ω_{ref} é denotada por $\Omega_{ref-FINAL}$.

O vetor Φ_f corresponde a f- ésima imagem de Ω_{ref} reconstruída para o domínio original através da equação $\Phi_f = \sum_{i=1}^{M'} \omega_i \cdot u_i$, em outras palavras, é tomada a f- ésima imagem de Ω_{ref} e sobre esta é aplica uma transformação que a levará para a dimensão original, resultando assim em Φ_f.

4.1.1.4 O reconhecimento

Dada uma imagem de entrada, indivíduo a ser reconhecido, são realizadas as operações de pré-processamento descritas na seção 4.1.1.1 chegando-se em $S_{original}$, imagem em tons de cinza no formato coluna com comprimento N^2. O próximo passo é dado pela equação $S_{reduzida} = X^T \cdot (S_{original} - \Psi)$, imagem normalizada e com dimensão reduzida. A etapa final é a medição da similaridade de $S_{reduzida}$ em relação a todas as imagens de Ω_{ref}. Este processo de comparação pode ser feito por classificadores convencionais tipo Pattern Matching Bayesianos e os do tipo adaptativos (Redes Neurais e SVM), (DUDA e HART et al., 2001).

Em (TURK e PENTLAND, 1991) a etapa de reconhecimento é uma extensão do processo apresentado no parágrafo anterior. Na seção 4.1.1.3 foram apresentados os vetores de distâncias $\in_{original}$ e $\in_{transformado}$, os *thresholds* $\theta_{original}$ e $\theta_{transformado}$, e por fim as matrizes Ω_{ref} e $\Omega_{ref-FINAL}$.

Onde

- $\theta_{original}$: É a máxima distância permitida para $\in_{original,f}$;
- $\theta_{transformado}$: É a máxima distância permitida para $\in_{transformado,k}$.

A definição de tais parâmetros possibilitou avaliar o problema de reconhecimento em quatro casos:

1. Se $\in_{transformado,k} < \theta_{transformado}$ e $\in_{original,f} < \theta_{original}$ a imagem de entrada é classificada como possivelmente contendo face e que esta, por sua vez, possivelmente pertence à base de referência. Contudo a chance de ocorrer um **falso positivo** não está descartada.
2. Se $\in_{transformado,k} > \theta_{transformado}$ e $\in_{original,f} < \theta_{original}$ a imagem de entrada é classificada como possivelmente contendo face e que esta, por sua vez, possivelmente não pertence à base de referência, ou seja, é uma pessoa desconhecida.
3. Se $\in_{transfomado,k} < \theta_{transformado}$ e $\in_{original,f} > \theta_{original}$ a imagem de entrada é classificada como possivelmente não contendo face; contudo a imagem no domínio transformado indica a existência de face e que esta pertence a um indivíduo contido na base de referência. Neste caso pode ocorrer um **falso positivo**, onde o sistema acha que reconheceu corretamente, mas na verdade a imagem original ou não contém face ou em caso contrário esta não ocupa espaço predominante na imagem, o que é um pré-requisito do trabalho original.
4. Se $\in_k > \theta_{\in_k}$ e $\in > \theta_{\in}$ é o caso onde sistema informa ou que não existe face ou que esta não ocupa espaço predominante da imagem.

4.1.2 Fisherfaces

Derivado do *Fisher's Linear Discriminant* (FLD) também conhecida por LDA: *Linear Discriminant Analysis* (LDA) (DUDA, HART e STORK, 2001). Fisherfaces é um método holístico baseado em PCA e LDA/FLD, inicialmente proposto por (BELHUMEUR, HESPANHA e KRIEGMAN, 1997). Seu objetivo, assim como no caso de Eigenfaces, era o de fornecer um algoritmo rápido e de baixa complexidade computacional, capaz de realizar a tarefa de reconhecimento dado um ambiente controlado, ou seja, imagens contendo predominantemente a face de uma pessoa, porém com maiores variações de iluminação[4], expressão e rotação do que as suportadas pela abordagem por Eigenfaces.

Em (BELHUMEUR, HESPANHA e KRIEGMAN, 1997) não foram contemplados testes com relação à variação de pose, ou seja, as imagens utilizadas eram frontais. A variação de luminosidade foi obtida variando a intensidade, a direção e a quantidade de fontes de luz; considerou-se também que as faces já estivessem alinhadas dentro da imagem.

Em Eigenfaces os clusters, grupos contendo imagens de uma mesma pessoa, formados após projeção das imagens não estão na sua forma mais compacta. Tal fato é herança da técnica do PCA, pois os autovetores utilizados como bases de projeções indicam as direções de maior distribuição das amostras e não as direções onde os grupos fiquem os mais compactos possíveis. Já em Fisherfaces buscou-se superar esta limitação, ou seja, o objetivo fora construir uma matriz de redução de dimensionalidade capaz de projetar as imagens de forma tal que os grupos resultantes ficassem compactos e bem separados, facilitando assim a etapa do reconhecimento.

4.1.2.1 O algoritmo

Dado o conjunto de treinamento $A = [x_1 \quad x_2 \quad \cdots \quad x_M]$, x_i é uma imagem[5] que já passou pelas etapas de pré-processamento mencionadas na seção 4.1.1.1 e está representada como um vetor coluna de comprimento N^2. Considerando que o conjunto A possua um número qualquer de imagens por pessoa é possível dividi-lo em classes X_i, onde cada classe corresponde a uma partição de A formada apenas pelas imagens de uma mesma pessoa, logo o conjunto de classes é representado por $[X_1 \; X_2 \cdots X_c]$.

---

[4] As variações, entre as imagens de uma mesma pessoa, devido à iluminação e direção de visão são quase sempre maiores do que as variações devido à mudança de indivíduo (MOSES, ADINI e ULLMAN, 1994).

[5] Nas seções 4.1.1 e 4.1.2, para fins de simplificação na notação, as imagens foram consideradas quadradas com dimensões NxN.

Também são construídas duas matrizes:

$$S_B = \sum_{i=1}^{c} N_i (\mu_i - \mu) \cdot (\mu_i - \mu)^T \text{ , onde } S_B \in R^{nxn} \tag{4.10}$$

Que representa a Matriz de dispersão entreclasses no domínio original.

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i) \cdot (x_k - \mu_i)^T \text{ , onde } S_W \in R^{nxn} \tag{4.11}$$

Que representa a Matriz de dispersão intraclasse no domínio original.

- x_k: É uma imagem em tons de cinza pré-processada conforme seção 4.1.1.1 e representada como um vetor coluna de comprimento N^2;
- μ_i: É a média das imagens da classe X_i;
- μ: É a média das imagens de todo o conjunto A;
- N_i: É o número de imagens na classe X_i;
- c: É o número de classes.

Segundo (DUDA, HART e STORK, 2001) se S_W for não singular, ou seja, possuir inversa, então a matriz de transformação W_{opt} é escolhida de forma a maximizar a razão entre os determinantes:

$$W_{opt} = \arg\max_{W} \frac{\left| W_{fld}^T S_B W_{fld} \right|}{\left| W_{fld}^T S_W W_{fld} \right|}$$
$$W_{opt} = W_{fld} = [w_1 \; w_2 \cdots w_m] \tag{4.12}$$

Onde

- $W_{fld}^T S_B W_{fld}$: É a Matriz de dispersão entre classes no domínio gerado após transformação via W_{fld}.
- $W_{fld}^T S_W W_{fld}$: É a Matriz de dispersão intraclasse no domínio gerado após transformação via W_{fld}.

- $\{w_i \mid i = 1, 2, \cdots, m\}$: É o conjunto de autovetores generalizados de S_B e S_W correspondentes aos m maiores autovalores generalizados $\{\lambda_i \mid i = 1, 2, \cdots, m\}$ isto é:

$$S_B w_i = \lambda_i S_W w_i \text{, onde } i = 1, 2, \cdots, m \tag{4.13}$$

O PROBLEMA DOS AUTOVETORES GENERALIZADOS

Conforme explicado em Eigenfaces, se o número de imagens M em A for muito menor que o tamanho da imagem, N^2, em outras palavras, for menor que o comprimento de um vetor coluna x_i, então haverá no máximo M autovalores significativamente diferentes de zero. A equação 4.13 tem como conjunto solução no máximo $c-1$ autovalores generalizados significativamente diferentes de zero, onde c corresponde ao número de classes.

Para contornar o problema dos autovetores generalizados de modo diferente ao apresentado em (DUDA, HART e STORK, 2001), o método Fisherfaces aplica a redução de dimensionalidade em duas etapas:

1. Primeiro aplica PCA sobre a matriz A para reduzir da dimensão N^2 para a dimensão $N^2 - c$, através da transformação $Y = W_{pca}^T \cdot A$, ou seja, W_{pca} possui N^2 linhas por $N^2 - c$ colunas.
2. O segundo passo é usar as imagens em Y como sendo um novo conjunto de treinamento para calcular a transformação $\Omega_{ref} = W_{fld}^T \cdot Y$; o cálculo de W_{fld} é dado pelas equações 4.12 e 4.13, ou seja, $W_{fld} = W_{opt}$. O que ocorre neste último passo é a redução de dimensionalidade de $N^2 - c$ para $c-1$, ou seja, W_{fld} possui $N^2 - c$ linhas por $c-1$ colunas.

Concluindo, a matriz de transformação total é dada por $W_{otimo}^{T} = W_{fld}^{T} \cdot W_{pca}^{T}$, onde as colunas de W_{otimo} são chamadas de Fisherfaces. Para o cálculo da matriz de referência basta fazer $\Omega_{ref} = W_{otimo}^{T} \cdot A$.

4.1.2.2 Interpretando as equações

Para melhor compreensão da abordagem por Fisherfaces segue baixo uma lista de equações com as devidas descrições.

- $A = [x_1 \quad x_2 \quad \cdots \quad x_M]$: Conjunto de treinamento;
- $S_B = \sum_{i=1}^{c} N_i (\mu_i - \mu) \cdot (\mu_i - \mu)^T$, onde $S_B \in R^{nxn}$: Matriz de dispersão entreclasses no domínio original;
- $S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i) \cdot (x_k - \mu_i)^T$, onde $S_W \in R^{nxn}$: Matriz de dispersão intraclasse no domínio original;
- $S_B w_i = \lambda_i S_W w_i$, onde $i = 1, 2, \cdots, m$: Equação dos autovetores generalizados os quais darão origem à W_{fld};
- $W_{fld}^{T} S_B W_{fld}$: É a Matriz de dispersão entreclasses no domínio gerado após uma transformação via W_{fld}.
- $W_{fld}^{T} S_W W_{fld}$: É a Matriz de dispersão intraclasse no domínio gerado após uma transformação via W_{fld}.
- $Y = W_{pca}^{T} \cdot A$: Equação de redução de N^2 para $N^2 - c$;
- $\Omega_{ref} = W_{fld}^{T} \cdot Y$: Equação de redução de $N^2 - c$ para $c - 1$;
- $W_{otimo}^{T} = W_{fld}^{T} \cdot W_{pca}^{T}$: Matriz de transformação total (solução ótima);
- $\Omega_{ref} = W_{otimo}^{T} \cdot A$: Matriz de referência utilizada na etapa de classificação;

A solução ótima $W_{otimo}^T = W_{fld}^T \cdot W_{pca}^T$ é tal que maximiza

$$W_{pca} \quad = \quad \arg\max_W \left| W_{pca}^T S_T W_{pca} \right| \quad \text{e} \quad W_{fld} \quad = \quad \arg\max_W \frac{\left| W_{fld}^T W_{pca}^T S_B W_{pca} W_{fld} \right|}{\left| W_{fld}^T W_{pca}^T S_W W_{pca} W_{fld} \right|} \qquad (4.14)$$

- $S_T = \sum_{k=1}^{N} (x_k - \mu) \cdot (x_k - \mu)^T$: É a Matriz de dispersão total calculada em relação ao domínio original, onde $S_T = S_W + S_B$;
- $W_{pca}^T S_T W_{pca}$: É a Matriz de dispersão total no domínio gerado após a primeira redução de dimensionalidade via W_{pca};
- $W_{pca}^T S_B W_{pca}$: É a Matriz de dispersão entre classes no domínio gerado após a primeira redução de dimensionalidade via W_{pca}.
- $W_{pca}^T S_W W_{pca}$: É a Matriz de dispersão intraclasse no domínio gerado após a primeira redução de dimensionalidade via W_{pca}.
- $W_{fld}^T W_{pca}^T S_B W_{pca} W_{fld}$: É a Matriz de dispersão entre classes no domínio gerado após a segunda transformação, esta realizada via W_{fld}.
- $W_{fld}^T W_{pca}^T S_W W_{pca} W_{fld}$: É a Matriz de dispersão intraclasse no domínio gerado após a segunda transformação., esta realizada via W_{fld}.

Cabe lembrar que tanto W_{otimo} como W_{pca} e W_{fld} são matrizes com colunas ortonormais e que a redução de dimensionalidade por meio de Fisherfaces leva a grupos mais compactos e separados, diferente de Eigenfaces que foca em projetar os dados nas direções de maior distribuição em detrimento do compromisso de gerar grupos compactos.

4.1.2.3 A base de referência

A base de imagens utilizadas no processo de reconhecimento é formada pela projeção de todas as imagens do conjunto de treinamento, ou seja, $\Omega_{ref} = W_{otimo}^T \cdot A$ que por sua vez corresponde às imagens de treinamento com a dimensionalidade reduzida de N^2 para $c-1$.

Em (BELHUMEUR, HESPANHA e KRIEGMAN, 1997) foi usado um banco de imagens do Harvard Robotics Laboratory e outro da Universidade de Yale, este criado pelos próprios pesquisadores.

O banco do Harvard Laboratory foi utilizado para testes de variação nas condições de iluminação, já o da Universidade de Yale foi utilizado para testes de variações de expressões (feliz, triste, sono, surpreso), iluminação e se a pessoa estava ou não usando óculos.

4.1.2.4 O reconhecimento

Em todos os experimentos em (BELHUMEUR, HESPANHA e KRIEGMAN, 1997) a etapa do reconhecimento foi realizada usando o classificador Nearest-Neighbor, não sendo informada a métrica utilizada para o cálculo das distâncias.

4.1.3 Laplacianfaces

Conforme explicado Eigenfaces usa o PCA para reduzir a dimensionalidade do vetor imagem através de uma matriz de transformação, na qual suas colunas são autovetores ortonormais entre si. Cada autovetor será capaz de capturar uma das regiões de maior variância da amostra de entrada, no caso o conjunto de imagens.

Em Fisherfaces é aplicado PCA e posteriormente LDA/FLD (*Fisher Linear Discriminant*), este por sua vez é responsável por gerar uma matriz de transformação na qual suas colunas representam autovetores capazes de projetar as imagens de entrada para uma dimensão reduzida. Em FLD as imagens de uma mesma pessoa irão formar um cluster o mais compacto possível, logo pessoas diferentes geram grupos diferentes, que estarão os mais separados possíveis, facilitando assim o processo de classificação.

Muitos autores advogam que algoritmos baseados em LDA sejam superiores aos baseados em PCA, porém em (MARTINEZ e KAK, 2001) fora mostrado que quando o conjunto de treinamento é pequeno PCA supera LDA e também que nestas condições PCA é menos sensível a variações nos conjuntos de treinamentos.

Seguindo o raciocínio de que uma imagem pode ser vista como um ponto em um espaço de dimensão equivalente ao seu tamanho, um conjunto de imagens nada mais é do que um conjunto de pontos em alta dimensão, espaço de dimensão N^2. Trabalhos como (CHANG, HU e TURK, 2003), (ROWIES e SAUL, 2000) e (SEUNG e LEE, 2000) visaram mostrar que se fosse possível visualizar a superfície descrita pelos pontos no espaço N^2 dimensional, ver-se-ia que estes descrevem uma superfície não linear, sinalizando que as técnicas lineares de

redução de dimensionalidade tais como PCA e LDA não seriam as mais eficientes para solucionar problemas de redução de dimensionalidade quando o conjunto de entrada é formado por imagens complexas como as de uma face.

Algumas técnicas não lineares de redução de dimensionalidade foram propostas, Isomap (TENENBAUM, SILVA e LANGFORD, 2000) e LEE (*Locally Linear Embedding*) (ROWEIS e SAUL, 2000), para a tarefa de reconhecimento facial, contudo segundo (XIAOFEI, ZHANG et al., 2005) elas não produziram bons resultados.

Outra abordagem para o problema foi o uso de técnicas baseadas em kernel tais como em (LIU, HUANG et al., 2002) e (YANG, 2000). Esses métodos são capazes de modelar a não linearidade da superfície, ESPAÇO DE FACES, descrita pelas diferentes imagens de faces usadas para a construção da base de referência. Todavia eles ainda são computacionalmente custosos e não são capazes de modelar a ESTRUTURA do espaço de faces.

A estrutura do espaço de faces

Seguindo a idéia de que uma imagem pode ser vista como um ponto em um espaço de alta dimensão, ou seja, o espaço de faces; se as imagens (pontos) de uma mesma pessoa forem ligadas através de arestas, dada uma lei de formação, a estrutura final será a de uma malha, ou melhor, um grafo também chamado de **ESTRUTURA DO ESPAÇO DE FACES**.

O objetivo do Laplacianfaces é definir uma matriz de transformação, W_{lpp}, capaz de reduzir a dimensionalidade das imagens preservando ao máximo a forma do grafo. Em outras palavras, o grafo na dimensão reduzida terá que conservar suas proporções; é como se o mesmo fosse uma bola que pode diminuir ou aumentar, mas nunca ficar deformada ou ser rompida. No caso da redução de dimensão deseja-se diminuir proporcionalmente a bola, gerando com isso grupos mais compactos, logo as imagens de um mesmo indivíduo ficarão mais similares na nova dimensão.

Em (XIAOFEI e SHUICHENG et al., 2005) mostrou-se também que usando a noção de grafos aplicados a massa de dados de entrada é possível desenvolver as teorias do PCA, do LDA e LPP. O que muda entre as teorias é apenas a lei de formação do grafo.

Por fim, Laplacianfaces é um método holístico baseado em PCA (*Principal Components Analysis*) e LPP (*Locally Preserving Projections*), inicialmente proposto por (NIYOGI e ZHANG et al., 2005), com o objetivo de fornecer um método capaz de reduzir a

dimensionalidade das imagens de faces, modelar as possíveis não linearidades do espaço de faces e preservar sua estrutura.

4.1.3.1 O algoritmo

Dado o conjunto de treinamento $A = [x_1 \; x_2 \; \cdots \; x_M]$, com M elementos onde x_i é uma imagem que já passou pelas etapas de pré-processamento mencionadas na seção 4.1.1.1 e está representada como um vetor coluna de comprimento N^2, é comum surgir o problema dos autovalores generalizados.

Para evitar cálculos desnecessários de autovalores que posteriormente serão descartados, o método do Laplacianfaces, assim como em Fisherfaces, aplica a redução de dimensionalidade em duas etapas:

1. Primeiro aplica PCA sobre a matriz A para reduzir da dimensão N^2 para M, através da transformação $Y = W_{pca}^T \cdot A$, ou seja, W_{pca} possui N^2 linha e M colunas.
2. O segundo passo é usar as imagens em Y como sendo um novo conjunto de treinamento para calcular a transformação $\Omega_{ref} = W_{lpp}^T \cdot Y$. A matriz W_{lpp}^T é construída apenas para realizar uma transformação em Y e não uma redução de dimensionalidade, algo que já fora efetuado na etapa 1, ou seja, W_{lpp} possui M linhas por M colunas.

Concluindo, a matriz de transformação total é dada por $W_{otimo}^T = W_{lpp}^T \cdot W_{pca}^T$, onde as colunas de W_{otimo} são chamadas de Laplacianfaces e para o cálculo da matriz de referência basta $\Omega_{ref} = W_{otimo}^T \cdot A$.

O processo de construção de W_{lpp}

Para se construir a matriz W_{lpp} três etapas são necessárias:

1. **Construção do grafo dos p vizinhos mais próximos, Nearest-Neighbor:** Este grafo vai ligar as imagens, já no domínio reduzido, para gerar uma malha que modele o espaço de faces e carregue consigo a informação de localidade. Durante o processo de

construção dois pontos, ou melhor, duas imagens Φ_i e Φ_j só serão ligadas por uma aresta se uma pertencer ao conjunto dos p vizinhos da outra[6].

2. **Escolha de pesos para as arestas:** O valor do peso da aresta que liga os pontos Φ_i e Φ_j é dado pela equação:

$$S_{ij} = e^{-\frac{\|\Phi_i - \Phi_j\|^2}{t}} \tag{4.15}$$

Onde $t = cte$ e $S_{ij} = 0$ quando não existir aresta ligando Φ_i e Φ_j e S: Matriz que modela a estrutura do espaço de faces.

3. **Cálculo do Eigenmap** (HE e NIYOGI, 2003)**:** Esta etapa implica na solução do problema dos autovetores generalizados, dado por:

$$\begin{aligned} YLY^T w_i &= \lambda YDY^T w_i \\ Y &= [\Phi_1 \ \Phi_2 \ \cdots \ \Phi_M] \end{aligned} \tag{4.16}$$

Onde

- $D_{ii} = \sum_j S_{ji}$: Matriz diagonal, cujos valores são dados pela soma das linhas ou colunas de S, pois esta é simétrica;
- $L = D - S$: Matriz Laplaciana.

Os k autovetores, $W_{lpp} = [w_0 \ w_1 \cdots w_{k-1}]$, encontrados serão ordenados na ordem **CRESCENTE**[7] de autovalores, $0 \le \lambda_0 \le \lambda_1 \le \cdots \le \lambda_{k-1}$.

---

[6] Para a construção do grafo dos p vizinhos mais próximos foi usado *Nearest-Neighbor* com métrica de distância Euclidiana. Contudo existem diversas outras métricas de distância tal como Mahalanobis.

[7] No processo de construção da matriz $W_{lpp} = [w_0 \ w_1 \cdots w_{k-1}]$ todos os autovetores serão preservados; e diferentemente de PCA, Eigenfaces e Fisherfaces eles serão organizados segundo a ordem **CRESCENTE** de autovalores.

4.1.3.2 Interpretando as equações

Para melhor compreensão do Laplacianfaces abaixo segue a lista de equações com as devidas descrições.

- $A = [x_1 \; x_2 \; \cdots \; x_M]$: Conjunto de treinamento;
- $Y = W_{pca}^T \cdot A$: Equação de redução de N^2 para M;
- $\Omega_{ref} = W_{lpp}^T \cdot Y$: Equação de transformação de M em M;
- $W_{otimo}^T = W_{lpp}^T \cdot W_{pca}^T$: Matriz de transformação total (solução ótima);
- $\Omega_{ref} = W_{otimo}^T \cdot A$: Matriz de referência utilizada na etapa de classificação;
- W_{lpp}^T: É a matriz responsável pela transformação que preserva as informações de localidade, ou seja, mantém a estrutura do espaço de faces.

Cabe ressaltar que tanto W_{otimo} como W_{pca} e W_{lpp} são matrizes com colunas ortonormais e que o método de Laplacianfaces tenta preservar a estrutura local do espaço de faces; diferente de Eigenfaces que tenta preservar a estrutura global e Fisherfaces que tenta preservar a informação discriminante.

4.1.3.3 A base de referência

A base de imagens utilizadas no processo de reconhecimento é formada pela projeção de todas as imagens do conjunto de treinamento, em outras palavras, $\Omega_{ref} = W_{otimo}^T \cdot A$ corresponde às imagens de treinamento com a dimensionalidade reduzida de n para N.

4.1.3.4 O reconhecimento

Em todos os experimentos em (NIYOGI e ZHANG et al., 2005) a etapa do reconhecimento foi realizada com o classificador Nearest-Neighbor configurado para 4 ou 7 vizinhos, dependendo do banco de dados, e com métrica Euclidiana para o cálculo das distâncias.

Foi mencionada a realização de testes com classificadores Bayesianos (MOGHADDAM e PENTLAND, 1997) e com o classificador Support Vector Machine (ROWEIS, SAUL e HINTON, 2001), mas os resultados não foram apresentados. Por fim foram sugeridos testes do Nearest-Neighbor com outras métricas, tal como Distância Normalizada pela Variância.

4.2 Métodos baseados em características

Os métodos baseados em características são em geral divididos em três grupos: os que são baseados em medidas geométricas da face tais como largura dos olhos, boca, nariz, face, distâncias olho-olho, olho-boca, nariz-boca; os que visam encontrar pontos na imagem que serão utilizados como característica da face (pontos fiduciais); e por fim os métodos que constróem modelos deformáveis (WISKOTT, FELLOUS et al., 1997), onde uma face é representada por um grafo.

Um dos primeiros trabalhos a utilizar a abordagem por medidas geométricas foi (KANADE, 1973), tendo atingido 75% de acerto no reconhecimento para um banco com 20 pessoas sendo duas imagens por pessoa, onde uma imagem foi utilizada para a construção da base de referência e a outra para posterior comparação.

Sistemas automáticos de extração de características como (PENEV e ATICK, 1996) fazem uso da abordagem que visa encontrar pontos fiduciais. Pelo fato da localização automática destes pontos serem um dos grandes desafios de tal método, muitos trabalhos tiveram como principal foco apenas a localização destas características sem se preocuparem com a posterior etapa de reconhecimento.

As técnicas apresentadas nas próximas seções foram escolhidas por terem sido as pioneiras e já tendo apresentado resultados satisfatórios; os demais trabalhos encontrados dentro destas linhas são em sua maioria pequenas variações das mesmas como no caso de (CHEN e YUAN, 2008) que aplica *Local Features Analysis* seguido de Laplacianfaces ou em (EHSAN e TSOTSOS, 2009) que aplica *Local Features Analysis* seguido de Transformada de Gabor.

4.2.1 Características Geométricas

O desenvolvimento de sistemas automáticos de reconhecimento de objetos através de processamento de imagem ainda hoje é um grande desafio para a área da visão computacional. Vários métodos foram propostos e re-propostos ao longo dos últimos anos, sendo (KANADE, 1973) um dos pioneiros na abordagem por características geométricas.

As medidas geométricas nada mais são que as distâncias entre pontos marcados na imagem da face de um indivíduo, por exemplo, as distâncias entre olhos, entre olho e boca, entre boca e nariz. No trabalho proposto por (BRUNELI e POGGI, 1993) foi realizada a comparação entre um método baseado em características geométricas e outro baseado em templates. Na abordagem por medidas geométricas o vetor construído possuía 35 valores, mas

não foi informado como obter todos eles. Na abordagem por templates cada pessoa foi representada por um vetor construído a partir de quatro partes da imagem original. A primeira parte é a imagem toda convertida em um vetor coluna, a segunda é uma região menor que incluí os olhos também convertida em vetor coluna, a terceira continha somente o nariz e a quarta apenas a boca; ambas foram transformadas em vetores colunas. O vetor final foi composto pela concatenação dos quatro vetores anteriores; não foi informada a seqüência de concatenação.

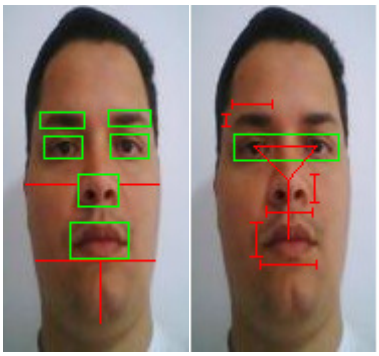


Figura 4.3: Medidas geométricas e templates. As regiões retangulares correspondem aos templates comumente extraídos e os segmentos de reta às medidas geométricas.

A normalização da imagem foi igual para ambos os métodos, a saber: a partir da imagem em tons de cinza primeiro foi realizada a rotação, nos casos necessários, a fim de que os olhos ficassem alinhados, em seguida ajustou-se o tamanho da imagem até que a distância entre olhos atingisse um valor pré determinado de pixels, por fim foi recortado um retângulo de tamanho NxM contendo a face e este foi usado como sendo a nova imagem, ou seja, na fase final todas as imagens possuem o mesmo tamanho.

Para a classificação foi usada a distância euclidiana e também o classificador bayesiano. O objetivo do experimento foi tentar responder questões como:

- Quão discriminantes foram as características utilizadas;
- O quanto o desempenho do sistema depende da resolução da imagem;
- As informações de baixa freqüência são menos ou mais efetivas do que as de alta freqüência;
- Qual das duas estratégias é melhor.

A fim de avaliar melhor o comportamento do sistema (BRUNELI e POGGI, 1993) expandiram o conjunto de treinamento e teste em quatro casos:

1. Conjunto de treinamento e teste sem qualquer alteração;
2. Conjunto de treinamento e teste após passarem por um filtro de média;

3. Conjunto de treinamento e teste após o cálculo do gradiente das imagens;
4. Conjunto de treinamento e teste após o cálculo do Laplaciano da imagem.

4.2.2 Métodos baseados em pontos fiduciais

Local Features Analysis é um método baseado em características (pontos fiduciais) capaz de fornecer uma representação compacta do espaço de imagens. Diferente do PCA, o qual produz representações globais não topográficas, em LFA é possível extrair representações locais topográficas para qualquer classe de objetos.

LFA foi inicialmente proposto por (PENEV e ATICK, 1996) com o foco em aplicações de reconhecimento facial. Variações para este método foram apresentadas em (CHEN e YUAN, 2008) e (EHSAN e TSOTSOS, 2009) mostrando a robustez da técnica frente a pequenas variações de pose, iluminação, rotação e oclusão da face. Com relação à complexidade do algoritmo, ele exige mais tempo de processamento que nos casos do Eigenfaces, Fisherfaces e Laplacianfaces.

Assim como grande parte das técnicas de reconhecimento facial, LFA tem base em PCA. *Local Features Analysis* é mais uma técnica capaz de promover a redução de dimensionalidade da amostra de entrada levando-a para uma dimensão que facilita o processo de classificação.

4.2.2.1 O algoritmo

A partir de um conjunto de imagens utilizadas para treinamento do sistema é construída a matriz $A = [\Phi_1\ \Phi_2 \ldots \Phi_M]_{N^2 xM}$, conforme equação 4.3. As colunas da matriz A correspondem às imagens de treinamento já pré-processadas e representadas com vetores coluna, Φ_i de ordem N^2. De posse da matriz A é calculada a matriz de covariância $C(x, y) = \frac{1}{M} A \cdot A^T$ e para esta são obtidos os autovalores e autovetores associados, representados respectivamente por $\Psi_r(x)$ e $\lambda_r(x)$.

A matriz de transformação do método LFA é computada através da equação 4.17.

$$K(x, y) = \sum_{r=1}^{N} \Psi_r(x) \cdot \left(1 \Big/ \sqrt{\lambda_r} \right) \cdot \Psi_r(y) \tag{4.17}$$

A projeção das imagens do conjunto de treinamento em um novo subespaço vetorial é dada pela equação 4.18.

$$O(x) = K(x, y)\Phi(x) = \begin{bmatrix} O(x_1) \\ O(x_2) \\ \vdots \\ O(x_V) \end{bmatrix} \tag{4.18}$$

Contudo neste novo subespaço os vetores imagem $O(x)$ ainda não possuem dimensão reduzida. A redução de dimensionalidade é dada pelo processo chamado de esparcificação.

A etapa de esparcificação

Dado o vetor de saída $O(x) = K(x, y)\Phi(x) = \begin{bmatrix} O(x_1) \\ O(x_2) \\ \vdots \\ O(x_V) \end{bmatrix}$ o algoritmo escolherá um subconjunto de $O(x)$, de modo que o mesmo represente uma saída de dimensionalidade reduzida. A seguir são descritos os critérios de escolha dos pontos que formam o subconjunto de $O(x)$.

O algoritmo inicia com um conjunto vazio representado por $M^{(0)} = \emptyset$ e a cada passo adiciona um ponto à M. No m-ésimo passo duas coisas são feitas:

1. A partir dos elementos do conjunto $M^{(m)}$ tenta-se reconstruir a saída $O(x)$ através de:

$$O^{rec}(x) = \sum_{m=1}^{|M|} a_m(x) O(x_m) \tag{4.19}$$

Onde

- $a_m(x)$: Representa os coeficientes lineares de predição ótimos, escolhidos de modo a minimizar o erro médio quadrático de reconstrução de $O(x)$, dado pela equação 4.20.

$$E = \left\langle \left\| O^{err}(x) \right\|^2 \right\rangle \equiv \left\langle \left\| O(x) - O^{err}(x) \right\|^2 \right\rangle \tag{4.20}$$

$$a_m(x) = \sum_{t=1}^{|M|} P(x, x_l)\left(P^{'-1}\right)_{lm} \tag{4.21}$$

- $P(x_l, x_m) \equiv P'_{lm}$: É um número correspondente ao elemento da linha l e coluna m da matriz P, matriz de correlação residual.
- $P(x, x_l)$: Corresponde à coluna l da matriz P, independente da linha.
- $(P'_{lm}) \equiv P^{-1}(x_l, x_m)$: É um número correspondente ao elemento da linha l e coluna m da inversa da matriz P, ou seja, P^{-1}.

2. No passo $(m+1)$- ésimo é feita a escolha do ponto que gera o maior erro de reconstrução E, em outras palavras, quer dizer que os pontos vão sendo escolhidos de modo que $O^{rec}(x)$ se aproxime suavemente de $O(x)$, o ponto que satisfaz tal condição será adicionado ao conjunto M. O processo de adição de pontos continua até que o erro fique abaixo de um valor aceitável, ou até que se atinja um total de N iterações. A quantidade N representa o comprimento final do vetor de saída, ou seja, é o número da dimensão para a qual $O(x)$ foi reduzido[8].

Ao final do processo de esparcificação obtém-se um vetor coluna $O(x)$ para cada imagem de entrada $\Phi(x)$ sendo que o comprimento de $O(x)$ será no máximo N, limite pré-definido de iterações, indicando que o mesmo sofreu a redução de dimensionalidade.

A base de referência é construída submetendo o conjunto $A = [\Phi_1\ \Phi_2 \ldots \Phi_M]_{N^2 xM}$ à transformação $O(x) = K(x, y) \cdot A$ e posteriormente ao processo de esparcificação, gerando a base de referência $O_{ref}(x)$ com dimensões reduzidas. Já na etapa de reconhecimento o vetor imagem, $\Phi(x)$, a ser reconhecido passa pelos processos de transformação $O(x) = K(x, y)\Phi(x)$ e esparcificação, gerando o vetor de saída $O_{output}(x)$ com dimensões reduzidas. Por fim a classificação pode ser realizada via classificador Euclidiano, Nearest-Neighbor, Bayesiano etc.

---

[8] Para cada entrada $\Phi(y)$ está associado um vetor de saída diferente $O(x)$, porém todos com o mesmo comprimento. O processo de esparcificação reduz a dimensionalidade vetores do tipo $O(x)$, através da seleção de subconjuntos de variáveis que sejam as mais descorrelacionadas possíveis. Contudo cabe lembrar que os subconjuntos escolhidos são diferentes de imagem para imagem, a única coisa que será igual é o comprimento final dos vetores $O(x)$ após a esparcificação.

4.3 Os métodos baseados em templates

Os métodos chamados de baseados em templates não constroem matrizes de transformação a partir das imagens de treinamento. Estes métodos apenas pegam as imagens, aplicam as etapas de pré-processamentos tais como: rotação, variação na resolução, equalização do histograma e conversão para tons de cinza; e por fim extraem partes da imagem. Tais regiões são chamadas de templates, que são usadas para construir o vetor de características, posteriormente utilizado na fase de classificação.

O reconhecimento é dado pegando a imagem de entrada, aplicando todas as etapas descritas acima e comparando seu template (vetor de características) com os templates das imagens da base de referência. Uma forma simples de realizar a comparação pode ser feita pelo cálculo da correlação, a imagem da base de referência mais similar à imagem a ser reconhecida será aquela que apresentar maior correlação. A etapa de comparação também é chamada de casamento de template.

4.4 Os métodos Híbridos

Os métodos híbridos são aqueles que usam a combinação de dois ou mais métodos de categorias diferentes. Um exemplo simples de método híbrido é aquele que usa medidas geométricas e templates na construção dos vetores de treinamento e teste.

Em (KAYA e KOBAYASHI, 1972) é apresentada uma abordagem híbrida, ou seja, o vetor de características representativo de uma pessoa foi composto em parte por valores obtidos via medidas geométricas e em parte por templates que foram extraídos, convertidos em vetores coluna e então concatenados ao vetor de características geométricas formando assim um grande vetor representativo. Em tal estudo foi dito que o conjunto de características extraídas da imagem deve preencher os seguintes requisitos:

- As características devem ser de fácil estimação;
- A dependência em relação à luminosidade deve ser a menor possível;
- A dependência em relação à variação de expressão facial deve ser a menor possível;
- O conjunto deve carregar o máximo de informação, ou seja, o conjunto de características extraídas deve ser dotado de grande poder de discriminação.

5 Abordagens propostas nesta dissertação

No presente capítulo apresentamos um estudo dos métodos holísticos Correlação, Eigenfaces, Fisherfaces e Laplacianfaces, para tal são realizados os experimentos 1, 2 e 3 visando a avaliação dos mesmos nas condições de:

- Redução de dimensionalidade;
- Variação na resolução das imagens;
- Tamanho da base de dados (Indivíduos a serem reconhecidos);
- Variação na iluminação;
- Diferentes conjuntos de treinamento e teste;
- Imagem com e sem fundo.

Plataforma de desenvolvimento utilizada

- Sistema Operacional: Microsoft Windows XP Professional, Versão 2002, Service Pack 2;
- Computador: AMD Turion 64 X2 Mobile Technology TL-58; 1,90 GHZ e 2 GB de RAM;
- Interface de desenvolvimento: Matlab 2009b.

Com base na análise dos resultados alcançados foi feita a proposição de novas abordagens. As variantes sugeridas são descrita nas Seções 5.4 à 5.7 e os experimentos de validação são apresentados ao longo do capitulo 6.

5.1 Bases de imagens utilizadas

Para avaliar os métodos holísticos selecionados os experimentos 1, 2 e 3 foram realizados com o emprego de três bancos de imagens publicamente disponíveis e amplamente utilizados, o ORL (http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html), o CMU PIE (SIM e BAKER, 2002) e o Yale (http://cvc.yale.edu/projects/yalefaces/ yalefaces.html).

O banco de imagens de Yale foi construído pelo Centro de Controle e Visão Computacional da Universidade de Yale e contém 165 imagens de 15 pessoas, ou seja, 11 imagens por pessoa. As imagens são frontais e apresentam variações de iluminação (fonte de luz pela esquerda, direita e frontal), expressão facial (normal, feliz, triste, dormindo surpreso e piscando) e uso ou não de óculos. O conjunto de treinamento utilizado no experimento foi formado com 7 imagens por pessoa, escolhidas de maneira randômica, o conjunto de teste ficou com as 4 imagens restantes por indivíduo. As imagens já em níveis de cinza foram normalizadas em escala e orientação para que os olhos ficassem alinhados, e recortadas para

conter apenas a parte interna da face. Por fim elas foram redimensionadas para as dimensões 128x128, 64x64, 32x32 e 16x16. A figura 5.1 apresenta imagens do banco Yale em seu formato original e a figura 5.2 apresenta as mesmas imagens após algumas etapas de pré-processamento.

Figura 5.1: Exemplo de imagens originais do banco Yale.

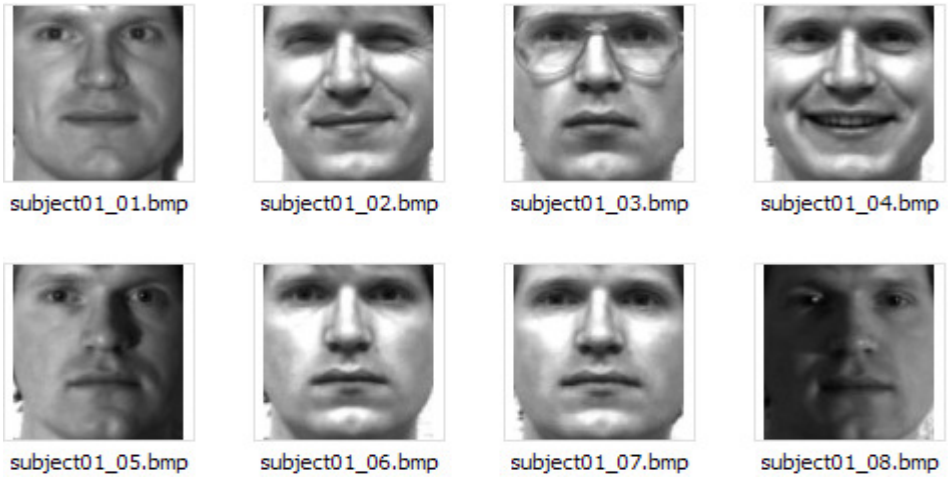


Figura 5.2: Exemplo de imagens normalizadas do banco Yale.

O bando de imagens de ORL contém imagens de 40 indivíduos, 10 imagens por pessoa, tendo sido tomadas em diferentes momentos. As variações de expressão facial são olhos abertos, olhos fechados, sorrindo e normal; os detalhes faciais são uso ou não de óculos. Todas as imagens possuem fundo preto, as faces são frontais e com pequenas variações de pose e orientação, no máximo 20 graus. A resolução original é 92X112 com 256 níveis de cinza.

Todo o processo de normalização é análogo ao aplicado às imagens do banco Yale, a diferença é que tanto o conjunto de treinamento quanto o de teste são formados com 5 imagens por indivíduo.

A figura 5.3 apresenta imagens do banco ORL em seu formato original e a figura 5.4 apresenta as mesmas imagens após algumas etapas de pré-processamento.

Figura 5.3: Exemplo de imagens originais do banco ORL.

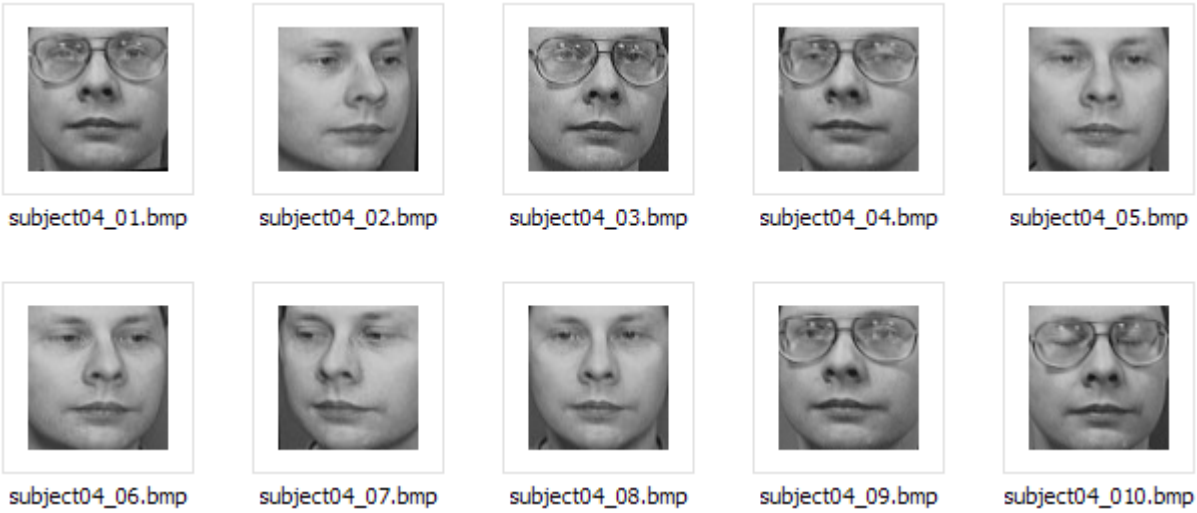


Figura 5.4: Exemplo de imagens normalizadas do banco ORL.

A base de imagens CMU PIE (*Pose, Ilumination, and Expression*) é composta por 68 indivíduos, totalizando 41.368. As imagens foram registradas por 13 câmeras em 21 condições diferentes de flashes. O subconjunto utilizado para o experimento foi formado

contendo 20 indivíduos, 24 imagens por pessoa, todas frontais e com expressão facial normal. Cada uma das 24 imagens corresponde a uma condição diferente de iluminação. O propósito de tal conjunto foi ter uma base em que se pudesse analisar o comportamento dos algoritmos dado a mudança apenas da iluminação, fixando as outras variáveis tais como expressão facial, rotação e orientação. Todo o processo de normalização é análogo ao aplicado às imagens do banco Yale, a diferença é que foram formados diversos conjuntos de treinamento e teste para verificar o desempenho dos métodos dado a quantidade de elementos nos conjuntos de treinamento e teste. Os pares treinamento e teste foram (12,12), (10,14), (8,16), (6,18) e (4,20), onde a primeira coordenada corresponde ao número de imagens por indivíduo usadas no conjunto de treinamento e a segunda coordenada é o equivalente para o conjunto de teste.

A figura 5.5 apresenta as imagens do banco PIE em seu formato original e a figura 5.6 apresenta as mesmas imagens após algumas etapas de pré-processamento.

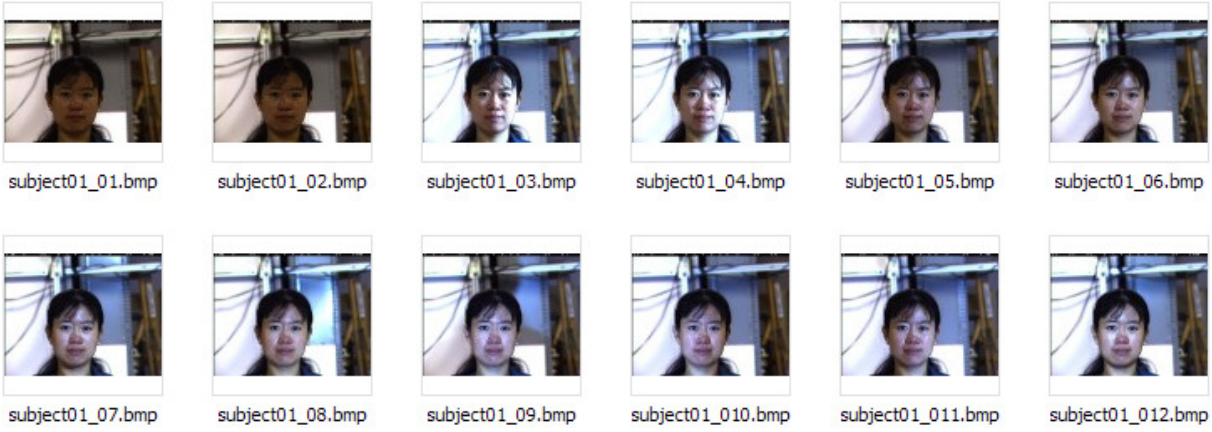


Figura 5.5: Exemplo de imagens originais do banco PIE.

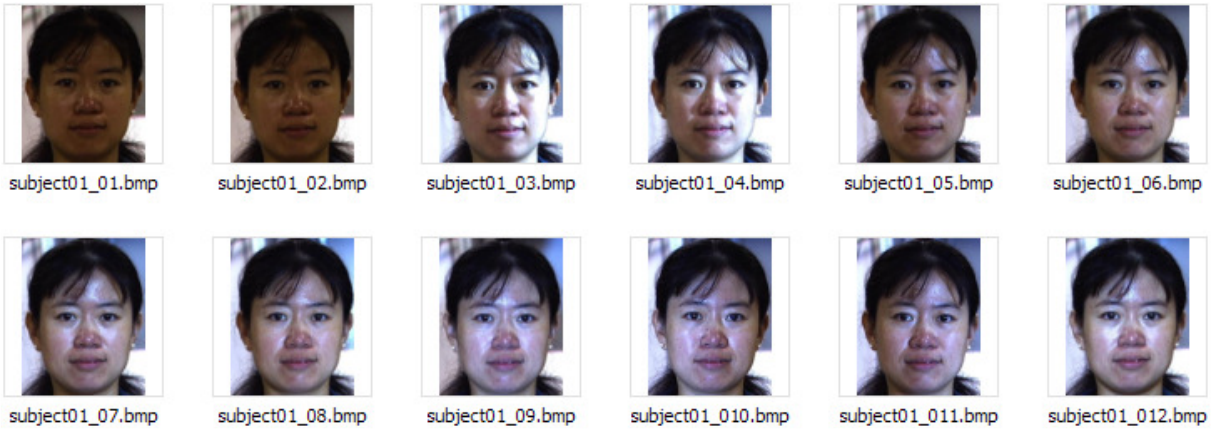


Figura 5.6: Exemplo de imagens normalizadas do banco PIE.

5.2 Experimentos

Ao longo da presente seção são apresentados os experimentos realizados visando avaliar os métodos holísticos Correlação, Eigenfaces, Fisherfaces e Laplacianfaces. Ao final de cada experimento apresentamos as devidas analises.

Experimento 01

- **Base Utilizada: Yale**
- Conjunto de treinamento com 7 imagens por pessoa;
- Conjunto de teste com 4 imagens por pessoa;
- Imagens escolhidas de maneira aleatória;
- Imagens com variações de iluminação e expressão facial;
- Formação dos conjuntos: Yale Original e Yale Recortado;
    - As imagens no conjunto Yale Original são compostas da face e de parte do fundo;
    - As imagens do conjunto Yale Recortado são compostas apenas da parte interna da face, procurando sempre que possível excluir o máximo do fundo.
- Imagens redimensionadas para as resoluções 64x64, 32x32 e 16x16.

Métodos testados nas condições de:

- Variações de iluminação e expressão facial;
- Redução de dimensionalidade;
- Variação na resolução;
- Imagem com e sem fundo.

Resultados alcançados

A figura 5.7 apresenta os resultados dos experimentos com o conjunto Yale Original tomado em três condições de resolução diferentes.

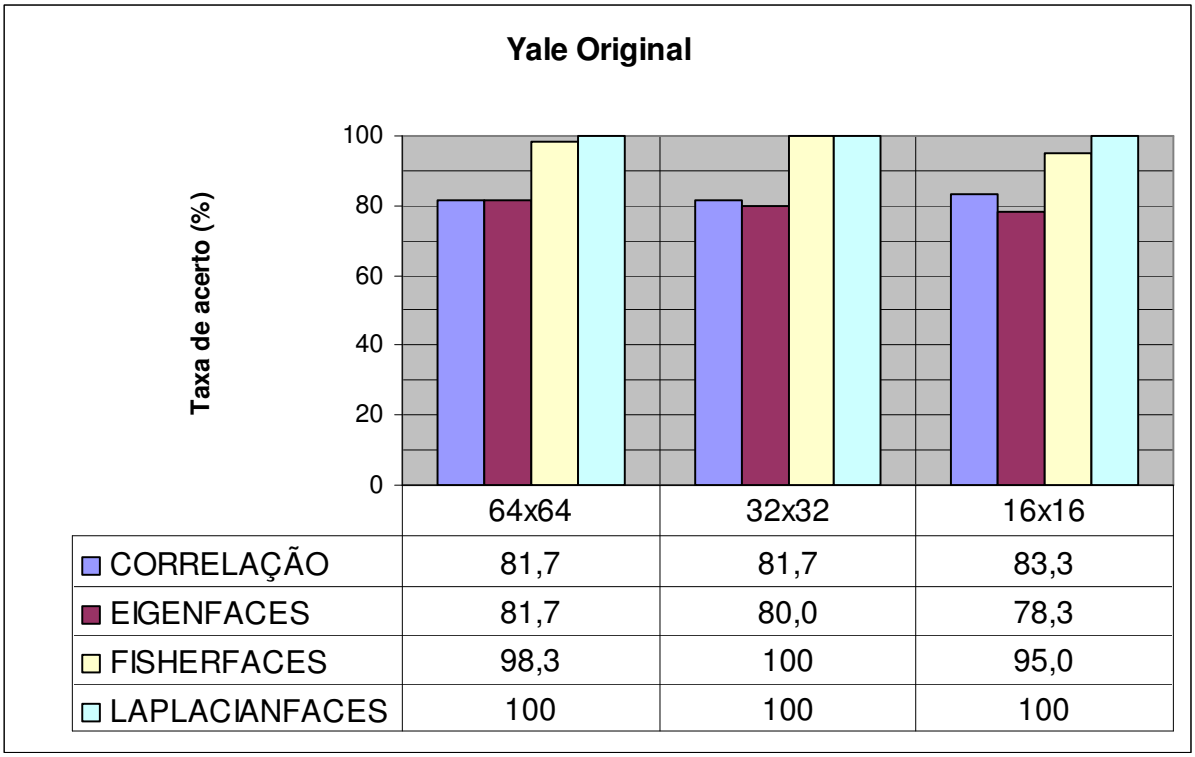


Figura 5.7: Conjunto Yale Original tomado em três resoluções diferentes.

A figura 5.8 apresenta a taxa de erro versus o aumento da dimensionalidade para o conjunto Yale Original tomado na resolução 64x64. Através da mesma é possível verificar o comportamento dos métodos Eigenfaces, Fisherfaces e Laplacianfaces dado o aumento da dimensionalidade do espaço de faces.

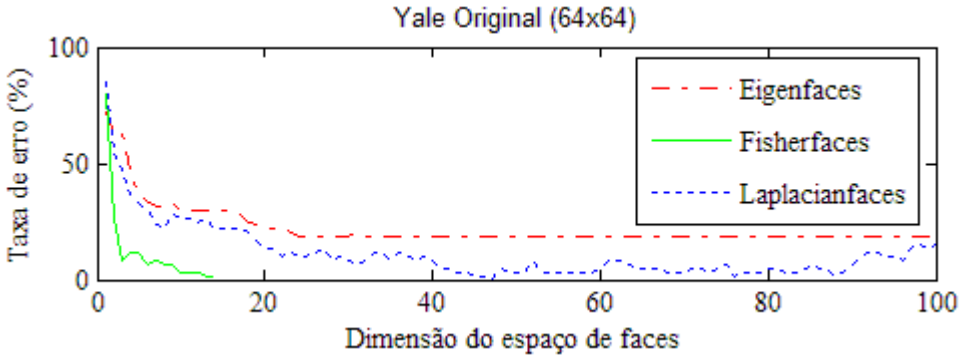
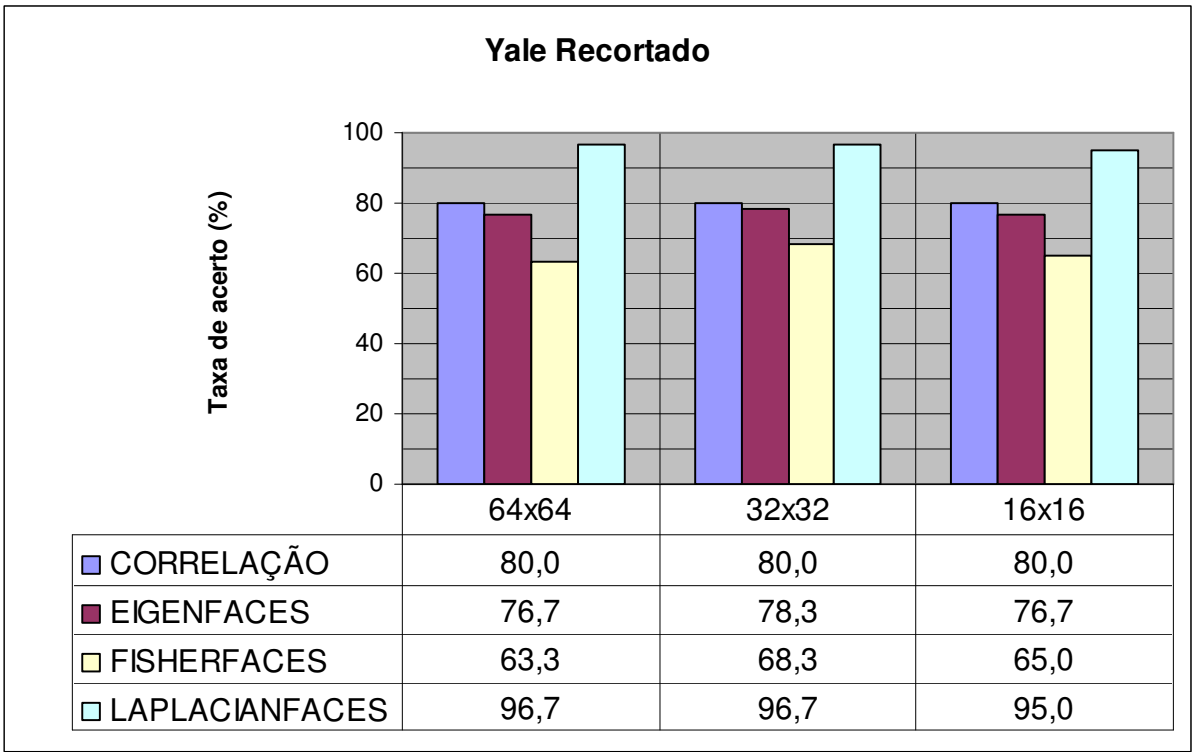


Figura 5.8: Taxa de erro versus aumento da dimensionalidade.

A figura 5.9 apresenta os resultados dos experimentos com o conjunto Yale Recortado tomado em três condições de resolução diferentes.



	64x64	32x32	16x16
CORRELAÇÃO	80,0	80,0	80,0
EIGENFACES	76,7	78,3	76,7
FISHERFACES	63,3	68,3	65,0
LAPLACIANFACES	96,7	96,7	95,0

Figura 5.9: Conjunto Yale Recortado tomado em três resoluções diferentes.

A figura 5.10 apresenta a taxa de erro versus o aumento da dimensionalidade para o conjunto Yale Recortado tomado na resolução 64x64. Através da mesma é possível verificar o comportamento dos métodos Eigenfaces, Fisherfaces e Laplacianfaces dado o aumento da dimensionalidade do espaço de faces.

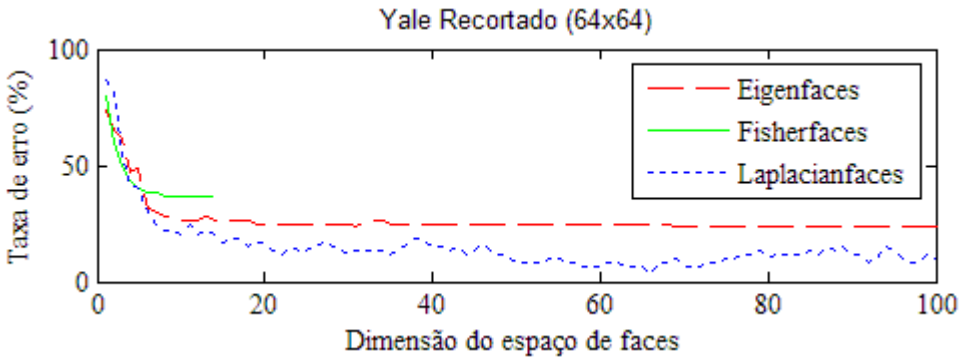


Figura 5.10: Taxa de erro versus aumento da dimensionalidade.

Análise dos resultados

- Se a imagem apresentar fundo homogêneo, o percentual de reconhecimento é maior quando parte do fundo for incluído;
- O método Laplacianfaces se manteve mais estável do que os demais;
- O método Fisherfaces não é necessariamente melhor do que o Eigenfaces (depende do experimento);
- O percentual de reconhecimento não reduz necessariamente com a diminuição da resolução das imagens;
- Existe um ponto ótimo para a redução de dimensionalidade.

Experimento 02

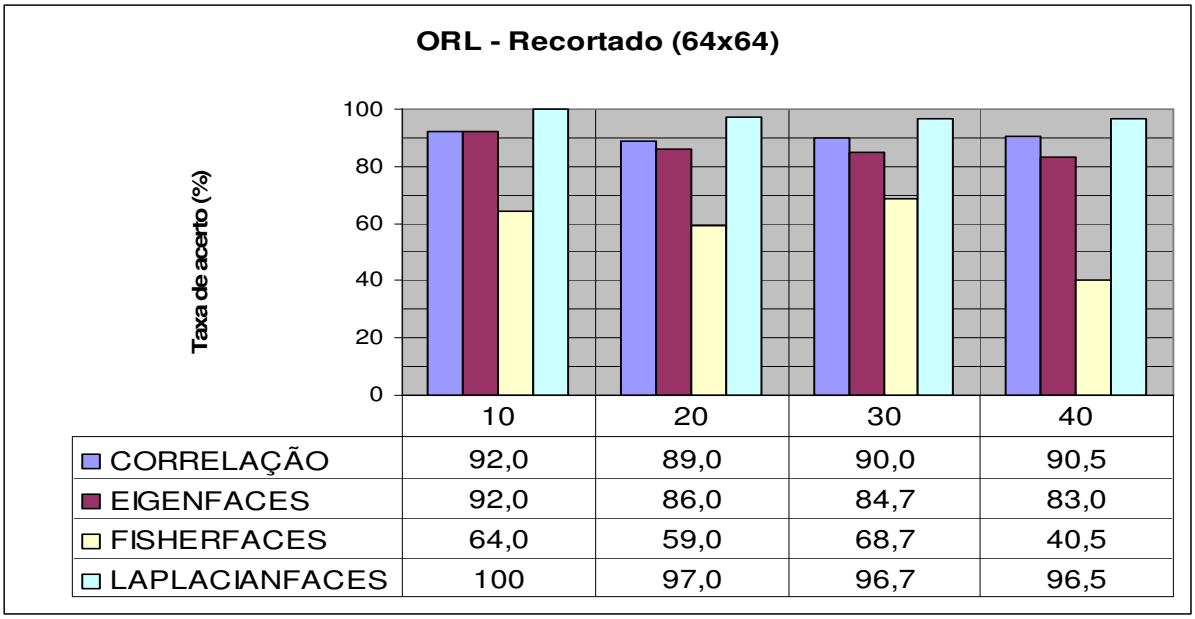
- **Base Utilizada: ORL**
- Conjunto de treinamento com 5 imagens por pessoa;
- Conjunto de teste com 5 imagens por pessoa;
- Imagens escolhidas de maneira aleatória;
- Imagens com variações de pose, iluminação e expressão facial;
- Formação do conjunto: ORL Recortado;
    - As imagens do conjunto ORL Recortado são compostas apenas da parte interna da face, procurando sempre que possível excluir o máximo do fundo.
- Imagens redimensionadas para as resoluções 64x64, 32x32 e 16x16.

Métodos testados nas condições de:

- Aumento do número de indivíduos a serem reconhecidos. Grupos contendo 10, 20, 30 ou 40 pessoas a serem reconhecidas.

Resultados alcançados

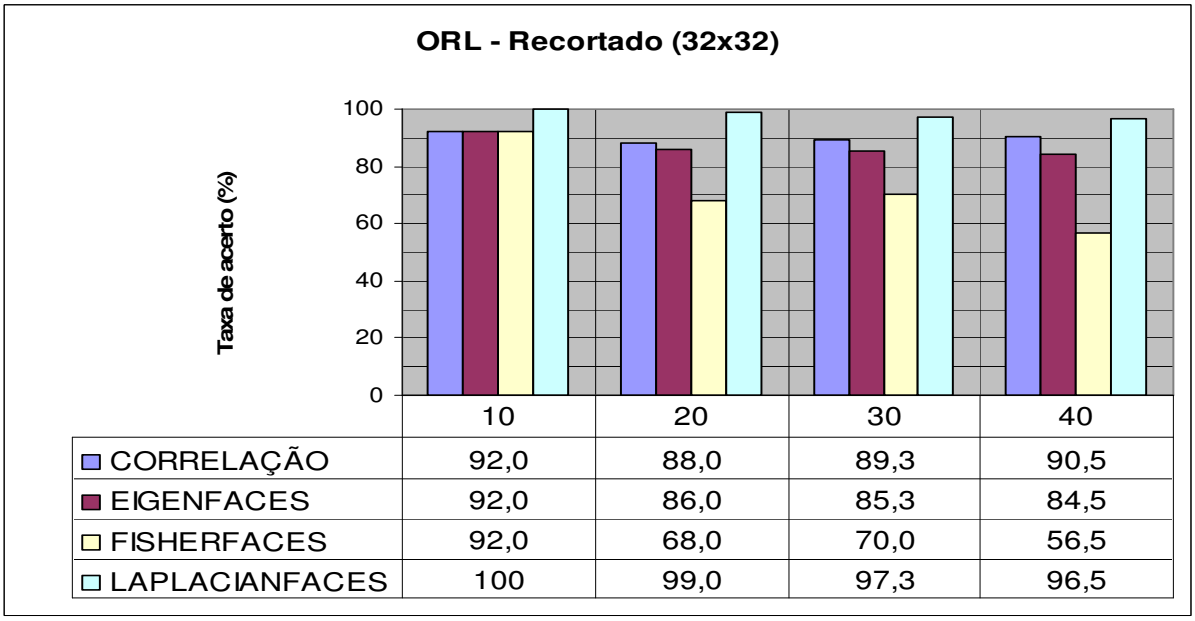
A figura 5.11 apresenta os resultados dos experimentos com o conjunto ORL Recortado. Subconjuntos contendo 10, 20, 30 ou 40 indivíduos a serem reconhecidos.



	10	20	30	40
CORRELAÇÃO	92,0	89,0	90,0	90,5
EIGENFACES	92,0	86,0	84,7	83,0
FISHERFACES	64,0	59,0	68,7	40,5
LAPLACIANFACES	100	97,0	96,7	96,5

Figura 5.11: Conjunto ORL Recortado dividido nos subconjuntos contendo 10, 20, 30 ou 40 indivíduos a serem reconhecidos.

A figura 5.12 apresenta os resultados dos experimentos com o conjunto ORL Recortado. Subconjuntos contendo 10, 20, 30 ou 40 indivíduos a serem reconhecidos.



	10	20	30	40
CORRELAÇÃO	92,0	88,0	89,3	90,5
EIGENFACES	92,0	86,0	85,3	84,5
FISHERFACES	92,0	68,0	70,0	56,5
LAPLACIANFACES	100	99,0	97,3	96,5

Figura 5.12: Conjunto ORL Recortado dividido nos subconjuntos contendo 10. 20. 30 ou 40 indivíduos a serem reconhecidos.

A figura 5.13 apresenta a taxa de erro versus o aumento da dimensionalidade para o subconjunto ORL Recortado tomado na resolução 32x32 e contendo 20 indivíduos a serem reconhecidos. Através da mesma é possível verificar o comportamento dos métodos Eigenfaces, Fisherfaces e Laplacianfaces dado o aumento da dimensionalidade do espaço de faces.

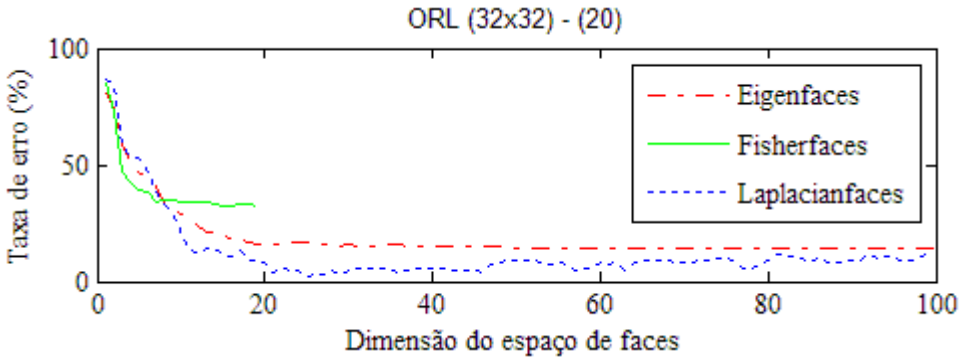


Figura 5.13: Taxa de erro versus o aumento da dimensionalidade.

Análise dos resultados

- No geral, ocorre queda no reconhecimento dado o aumento do número de indivíduos a serem reconhecidos;
- Fisherfaces é o método que mais degenera;
- Laplacianfaces continua sendo o método mais estável;
- O percentual de reconhecimento, necessariamente não reduz com a diminuição da resolução.

Experimento 03

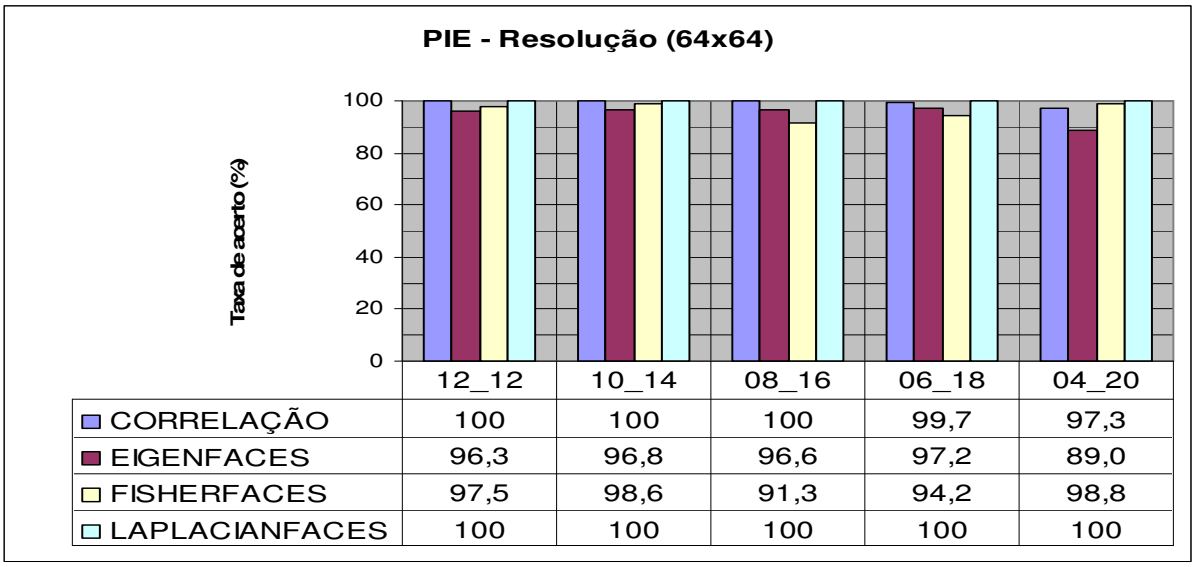
- **Base Utilizada: PIE**
- Conjuntos de (treinamento,teste): (12,12), (10,14), (8,16), (6,18) e (4,20);
- Imagens escolhidas de maneira aleatória;
- Imagens com variação de iluminação;
- Formação do conjunto: PIE Recortado;
  - As imagens do conjunto PIE Recortado incluem toda a face, contudo procura excluir o máximo do fundo.
- Imagens redimensionadas para as resoluções 64x42, 32x32 e 16x16.

Métodos testados nas condições de:

- Variação na iluminação;
- Variação do número de amostras nos conjuntos de treinamento e teste.

Resultados alcançados

A figura 5.14 apresenta os resultados dos experimentos com o conjunto PIE Recortado tomado na resolução 64x64. Cada coluna da tabela de resultados corresponde a uma divisão entre elementos no conjunto de treinamento e elementos no conjunto de testes.



	12_12	10_14	08_16	06_18	04_20
CORRELAÇÃO	100	100	100	99,7	97,3
EIGENFACES	96,3	96,8	96,6	97,2	89,0
FISHERFACES	97,5	98,6	91,3	94,2	98,8
LAPLACIANFACES	100	100	100	100	100

Figura 5.14: Conjunto PIE Recortado dividido em diferentes conjuntos de treinamento e teste.

A figura 5.15 apresenta a taxa de erro versus o aumento da dimensionalidade para o subconjunto PIE Recortado tomado na resolução 64x64 e conjunto de treinamento e teste (4,20). Através da mesma é possível verificar o comportamento dos métodos Eigenfaces, Fisherfaces e Laplacianfaces dado o aumento da dimensionalidade do espaço de faces.

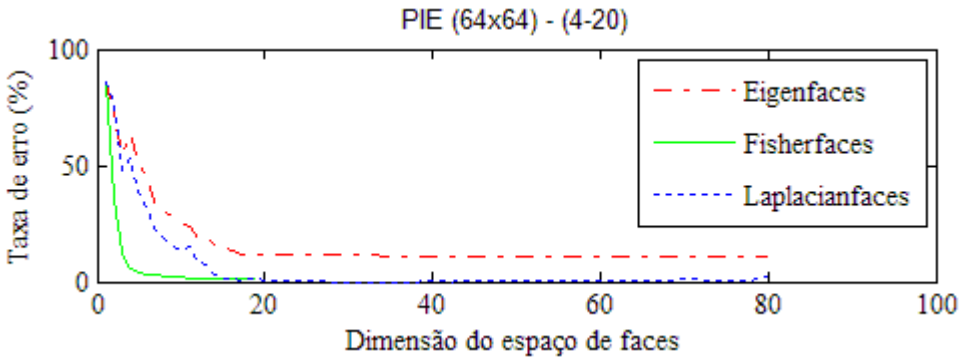
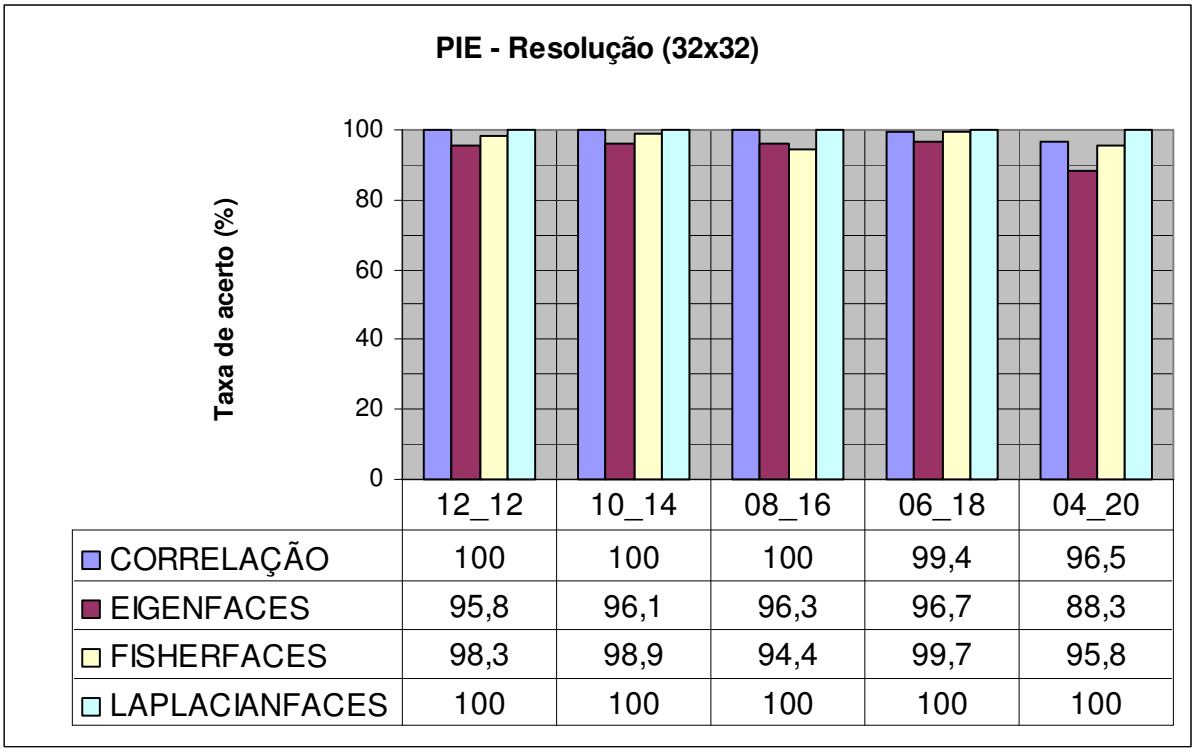


Figura 5.15: Taxa de erro versus o aumento da dimensionalidade.

A figura 5.16 apresenta os resultados dos experimentos com o conjunto PIE Recortado tomado na resolução 32x32. Cada coluna da tabela de resultados corresponde a uma divisão entre elementos no conjunto de treinamento e elementos no conjunto de testes.



	12_12	10_14	08_16	06_18	04_20
CORRELAÇÃO	100	100	100	99,4	96,5
EIGENFACES	95,8	96,1	96,3	96,7	88,3
FISHERFACES	98,3	98,9	94,4	99,7	95,8
LAPLACIANFACES	100	100	100	100	100

Figura 5.16: Conjunto PIE Recortado dividido em diferentes conjuntos de treinamento e teste.

Análise dos resultados

- A iluminação influencia menos do que a variação de expressão facial, visto que todos os métodos apresentaram elevados percentuais de reconhecimento;
- No geral, reduzir o número de amostras no conjunto de treinamento provoca queda na taxa de reconhecimento.

5.3 Análise comparativa

O Fisherfaces introduziu a transformação LDA na intenção de produzir um método mais robusto do que o Eigenfaces. Contudo, os experimentos mostraram que nem sempre o Fisherfaces é superior ao Eigenfaces;

O Laplacianfaces introduziu a transformação LPP na intenção de produzir um método mais robusto do que o Eigenfaces. Em todos os experimentos realizados foi possível comprovar tal superioridade;

Com o foco no Eigenfaces, novas abordagens foram propostas na intenção de melhorar a robustez de tal método perante as seguintes situações:

- Variação de luminosidade;
- Imagens com partes da face encobertas;
- Conjunto de treinamento com poucas amostras por indivíduo.

5.3.1 Variantes sugeridas

- O **CT-Eigen** foi proposto para lidar com a questão da variação de luminosidade;
- O **CT-Eigenblocos** foi proposto para lidar com a questão de faces parcialmente encobertas;
- O **Expansão** foi proposto para lidar com a questão do número reduzido de amostras no conjunto de treinamento;
- O **Geométrico-LPP** foi proposto para testar a combinação entre o método Características Geométricas e o Laplacianfaces.

5.4 CT-Eigen

No Eigenfaces o percentual de acerto tende a diminuir com o aumento da não linearidade do espaço de faces. Em vista disso é comum a aplicação de técnicas tal como Equalização do Histograma para minimizar os efeitos da variação de luminosidade dado que esta variável exerce forte influência na não linearidade do espaço de faces.

O método aqui proposto aplica a Transformada Census sobre as imagens antes da etapa do Eigenfaces, figura 5.17. Como pode ser visto na figura 5.19, dado duas imagens de um mesmo indivíduo diferentes apenas nas condições de iluminação na qual foram registradas, após serem submetidas à Transformada Census darão origem a duas novas imagens visualmente mais similares do que as anteriores. Portanto, o objetivo do CT-Eigen é ser um método de reconhecimento facial mais robusto do que Eigenfaces pelo fato de trabalhar em um subespaço de faces mais linear.

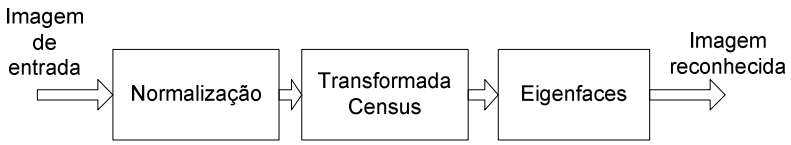


Figura 5.17: Etapas do CT – Eigen.

A hipótese é que quanto mais similares forem as imagens de um determinado conjunto, mais linear será o espaço de faces representado por tal conjunto. Por conseguinte mais representativas serão as matrizes transformação geradas pelos métodos Eigenfaces e Fisherfaces, traduzindo-se assim em um aumento no percentual de acerto na etapa de classificação. Para a classificação foi utilizada a distância euclidiana.

5.4.1 Transformada Census

A Transformada Census é uma transformação local não paramétrica; ela foi inicialmente proposta por (ZABIH E WOODFILL, 1996). O cálculo é realizado por intermédio de comparações entre os pixels de uma dada região e o elemento central da região em questão. De maneira formal, a Transformada Census define regiões quadradas $N(x)$, as mais utilizadas são da ordem 3x3 e 5x5, onde $N(x)$ corresponde à vizinhança em torno do pixel x, sendo que $x \notin N(x)$. A saída da CT é uma string de bits, onde cada bit é calculado pela equação 5.1. Por fim o novo valor do pixel x é dado pela conversão da string de bits em decimal.

$$C(x) = \bigotimes_{y \in N} \zeta(I(x), I(y)) \tag{5.1}$$

- $I(x)$: Pixel central da região quadrada;
- $I(y)$: Pixel da vizinhança;
- $\zeta(I(x), I(y))$: Função comparação, $\zeta = 1$ quando $I(y) > I(x)$ e $\zeta = 0$ caso contrário;
- $\otimes$: No caso acima é visto como um operador de concatenação;
- $C(x)$: É visto como o Kernel, de centro nulo, definido sobre a região $N(x)$.

A figura 5.18 mostra um subconjunto de todas as características locais que a transformada Census é capaz de gerar a partir de regiões 3x3. Tais matrizes são chamadas de Kernels da Transformada Census.

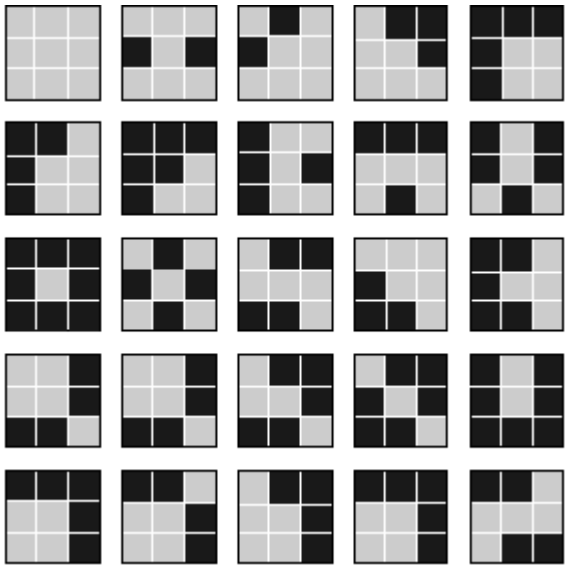


Figura 5.18: Subconjunto de possíveis Kernels.

A figura 5.19 mostra o resultado da aplicação da CT sobre duas imagens de um mesmo indivíduo, as quais foram tomadas em situações distintas de iluminação.

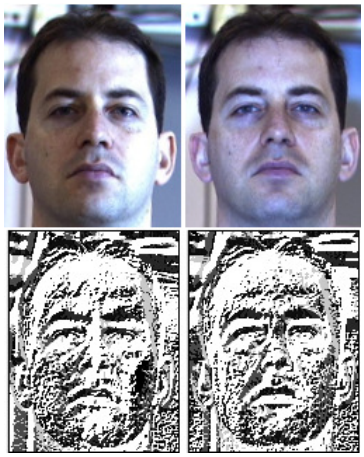


Figura 5.19: Exemplo de invariância à iluminação. Embora a iluminação varie significativamente nas imagens originais, as imagens resultantes da Transformada Census permanecem bastante similares.

5.5 CT-Eigenblocos

A técnica CT-Eigenblocos é uma combinação entre a Transformada Census e Eigenfaces. A figura 5.20 apresenta as etapas do algoritmo.

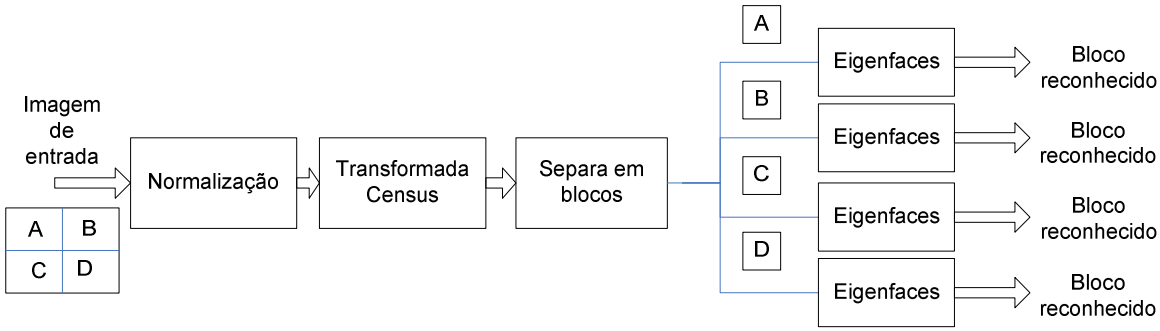


Figura 5.20: Etapas do CT-Eigenblocos.

Após as imagens terem sido normalizadas e submetidas à Transformada Census elas são divididas em quatro blocos, cada qual forma uma nova base de imagens. O objetivo aqui é conseguir o reconhecimento mesmo com oclusão parcial ou mudanças significativas em partes da face.

Aplicando o reconhecimento por bloco é possível reconhecer partes da face para encontrar a pessoa correta na base de referência. Também é possível definir uma escala de confiança que vai de 1 a 4, onde 4 indica que todos os quadrantes foram reconhecidos e por conseguinte a certeza é máxima.

Se o sistema for configurado para aceitar como acerto todos os casos em que ao menos 1 quadrante tenha sido reconhecido, o desempenho tende a ser mais robusto e superior às demais técnicas.

No final das etapas do CT-Eigenblocos, cada quadrante da imagem de entrada estará associado a um quadrante de uma imagem do conjunto de treinamento; como critério de escolha da classe de saída, indivíduo reconhecido, basta selecionar no conjunto de treinamento a classe que ocorreu um maior número de vezes nas saídas do CT-Eigenblocos, em caso de empate, basta selecionar aquela que apresentou o quadrante de menor distância em relação ao respectivo quadrante na imagem de entrada, o resultado desta consideração é apresentado no capítulo 6.

5.6 Expansão

A técnica de expansão foi concebida para o caso em que a base de treinamento seja pequena, apresentando poucas imagens por pessoa. A figura 5.21 apresenta as etapas do algoritmo.

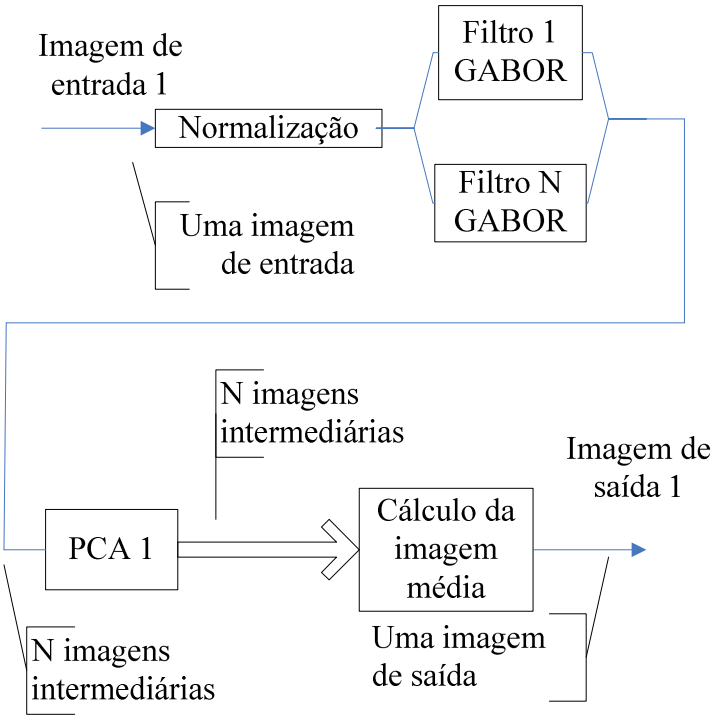


Figura 5.21: Etapas para construção da base de referência.

Após a normalização (rotação, recorte e escala) cada imagem é convoluída com os filtros de Gabor, gerando um conjunto de N imagens. Para cada conjunto intermediário é calculada a matriz de transformação do PCA e aplicada sobre o mesmo. Do conjunto resultante é calculada a imagem média que por sua vez será armazenada para formar a base de referência. As matrizes de transformação do PCA também são guardadas, pois são utilizadas na etapa do reconhecimento.

Para o reconhecimento, figura 5.22, cada imagem de entrada é submetida à normalização e aos N filtros de Gabor gerando N imagens. Sobre as imagens resultantes aplicam-se as matrizes de transformação do PCA previamente calculadas. Após cada etapa de PCA é calculada a imagem média que será comparada com as imagens médias armazenadas na base de referência. As comparações podem ser realizadas por diversos classificadores como Euclidiano e Nearest-Neighbor. No trabalho em questão foi usado o classificador euclidiano.

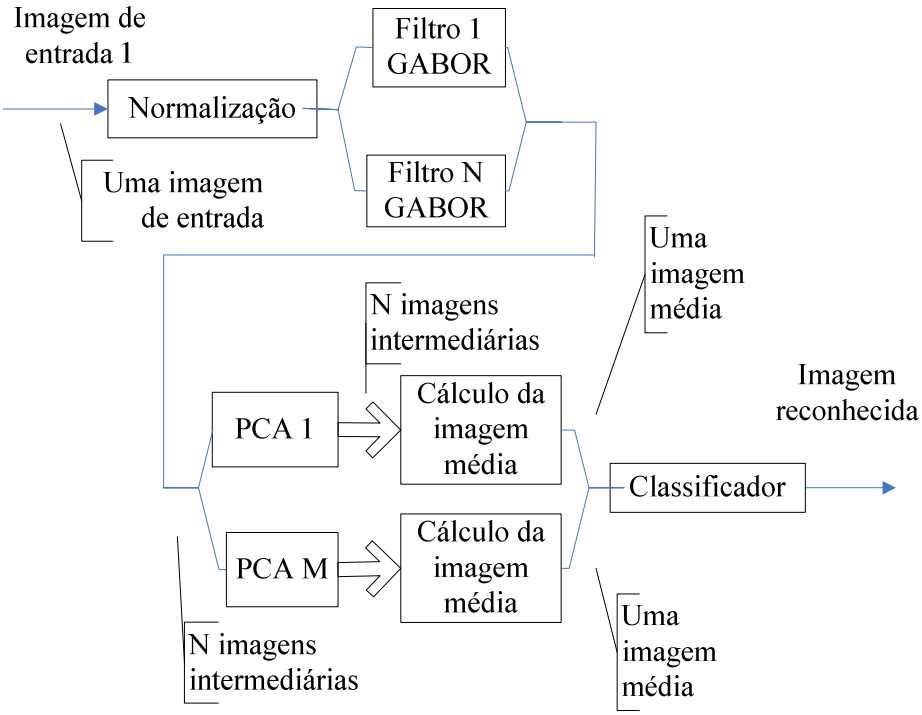


Figura 5.22: Etapas do processo de reconhecimento.

5.6.1 Transformada de Gabor

Em processamento de imagem a transformada de Gabor é vista como um operador $h(x, y)$ utilizado com a finalidade de realçar a qualidade da imagem (MARQUES, 2004). Em trabalhos de reconhecimento facial é comum encontrá-la sendo usada basicamente em duas situações: na primeira a mesma é aplicada como um filtro direcional para realce de bordas (ERSI e HAJEBI, 2003) e (QIAN, SU e XU, 2008). Neste caso $h(x, y)$ é convoluído com a imagem de entrada. A segunda abordagem é o uso de Gabor como descritor de características (EHSAN e TSOTSOS, 2009) aplicando-o sobre os pontos escolhidos para o vetor de saída $O(x)$ após o processo de esparcificação.

A equação 5.2 descreve um filtro de Gabor bidimensional direcional e real cuja representação gráfica é dada pela figura 5.23.

$$G(x, y, f, \theta) = \exp\left(-0.5 \cdot \left(\frac{x_r^2}{gx^2} + \frac{y_r^2}{gy^2}\right)\right) \cdot \cos(2 \cdot \pi \cdot f \cdot y_r) \tag{5.2}$$

A função de Gabor bidimensional direcional e real pode ser vista como o produto entre uma gaussiana, $\exp\left(-0.5\cdot\left(\frac{x_r^2}{gx^2}+\frac{y_r^2}{gy^2}\right)\right)$, e a função cosseno, $\cos(2\cdot\pi\cdot f\cdot y)$.

Onde

- (x_r, y_r): Coordenadas do pico do envelope Gaussiano;
- $\begin{array}{l} x_r = x\cdot\cos(\theta) - y\cdot sen(\theta) \\ y_r = y\cdot\cos(\theta) - x\cdot sen(\theta) \end{array}$: Sistema de equações que modelam a rotação do envelope Gaussiano;
- (gx, gy): Largura da base do envelope Gaussiano em relação aos eixos x e y;
- f : Freqüência da portadora cossenoidal ao longo do eixo y;

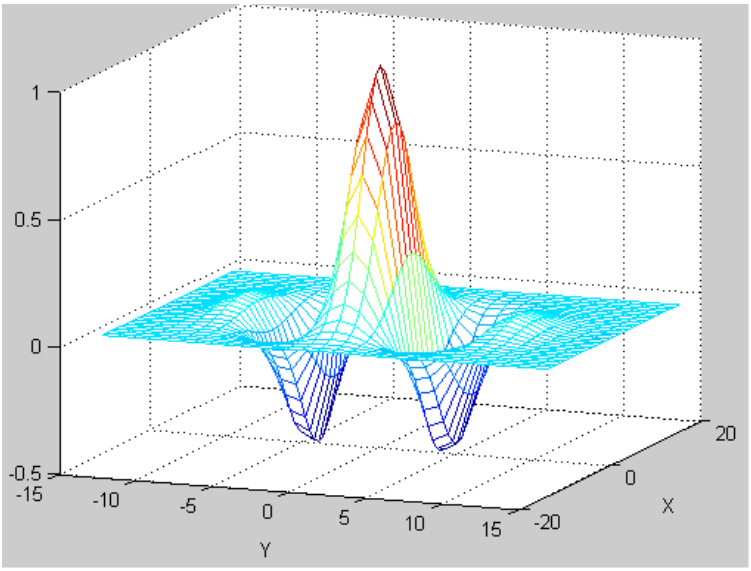


Figura 5.23: Resposta impulsional do filtro de Gabor bidimensional direcional e real.

Em (ERSI e HAJEBI, 2003) e (QIAN, SU e XU, 2008) foi feito o uso da forma complexa do filtro de Gabor. Em tal situação a imagem de interesse é dada pelo valor absoluto de $g(x, y)$.

A equação 5.3 descreve um filtro de Gabor bidimensional direcional e complexo cuja representação gráfica é dada pelas figuras 5.24 e 5.25.

$$G_c(x, y, f, \theta) = G_r(x, y, f, \theta) + j\cdot G_i(x, y, f, \theta) \tag{5.3}$$

Onde

- $G_r(x, y, f, \theta) = \exp\left(-0.5 \cdot \left(\frac{x_r^2}{gx^2} + \frac{y_r^2}{gy^2}\right)\right) \cdot \cos(2 \cdot \pi \cdot f \cdot y_r)$: Parte real do filtro de Gabor;
- $j \cdot G_i(x, y, f, \theta) = j \cdot \exp\left(-0.5 \cdot \left(\frac{x_r^2}{gx^2} + \frac{y_r^2}{gy^2}\right)\right) \cdot sen(2 \cdot \pi \cdot f \cdot y_r)$: Parte imaginária do filtro de Gabor;
- $\begin{matrix} x_r = x \cdot \cos(\theta) - y \cdot sen(\theta) \\ y_r = y \cdot \cos(\theta) - x \cdot sen(\theta) \end{matrix}$: Sistema de equações que modelam a rotação do envelope Gaussiano;
- $G_r(x, y, f, \theta)$, $G_i(x, y, f, \theta) \in \Re$ e $G_c(x, y, f, \theta) \in C$

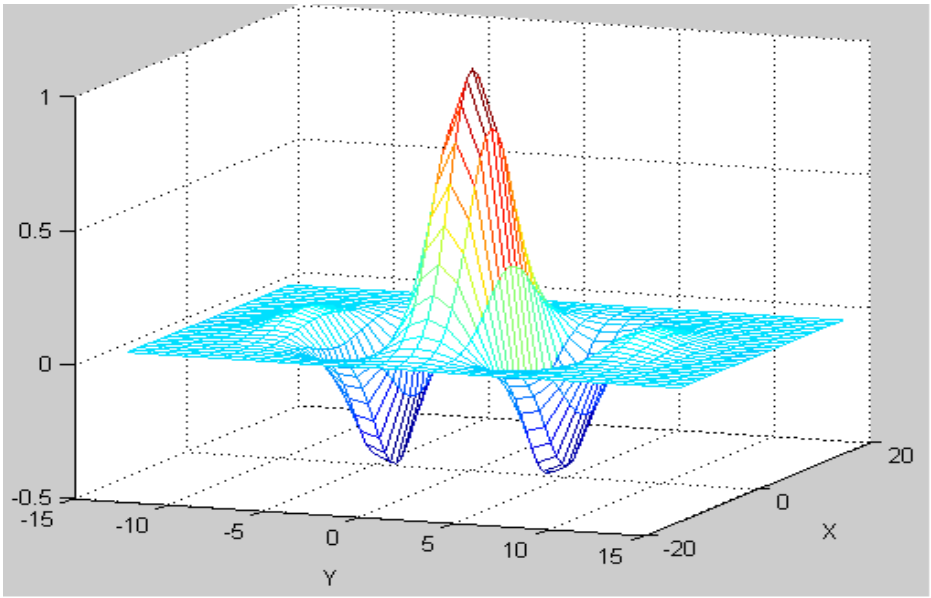


Figura 5.24: Parte real da resposta impulsional do filtro de Gabor bidimensional direcional e imaginário.

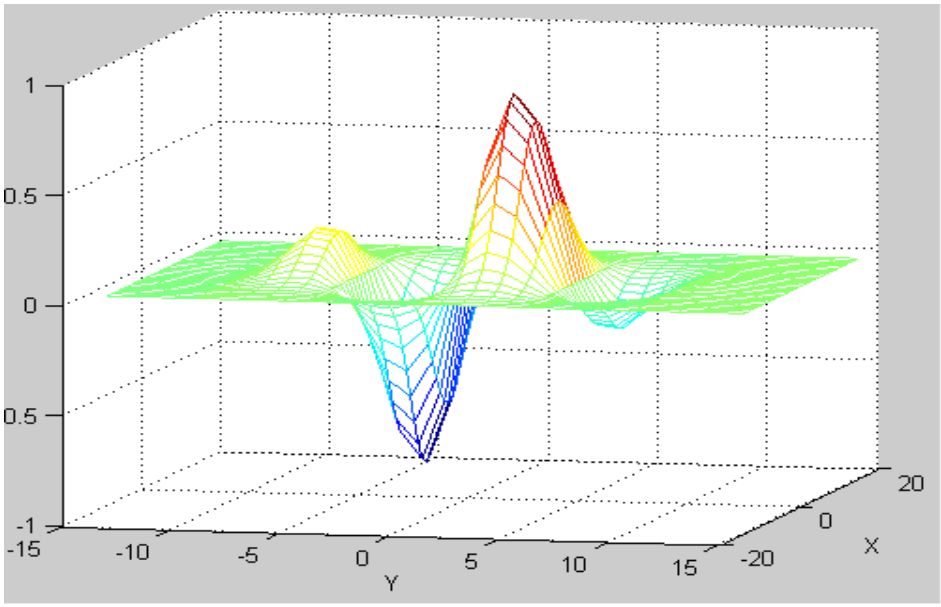


Figura 5.25: Parte imaginária da resposta impulsional do filtro de Gabor bidimensional direcional e imaginário.

O processo de convolução é apresentado na figura 5.6 e modelado pela equação 5.4.

$$g(x,y) = h(x,y) \otimes f(x,y) \tag{5.4}$$

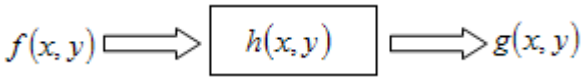


Figura 5.26: Diagrama de bloco de um sistema.

- $f(x,y)$: Corresponde à imagem de entrada;
- $g(x,y)$: Corresponde à imagem de saída;
- $h(x,y)$: Corresponde à máscara de convolução espacial;
- $\otimes$: Neste caso o símbolo apresentado é interpretado como operador de convolução.

Onde $f(x,y)$ e $g(x,y)$ são imagens com as mesmas dimensões e $h(x,y)$ é uma matriz quadrada com dimensões variadas. A escolha das dimensões de $h(x,y)$ depende do tamanho da máscara que se desejar criar. No caso de $h(x,y)$ ser um número complexo, a imagem de interesse será dada pelo módulo de $g(x,y)$.

5.7 Geométrico – LPP

O método proposto é um híbrido entre Características Geométricas e Laplacianfaces. Dado que o escopo da dissertação não trata de sistemas de localização e marcação automática das características geométricas, foram desenvolvidos, em Matlab, dois programas para normalização da imagem e marcação manual das diferentes características geométricas da face, figuras 5.27 e 5.28.

Figura 5.27: Interface para rotação e recorte da face.

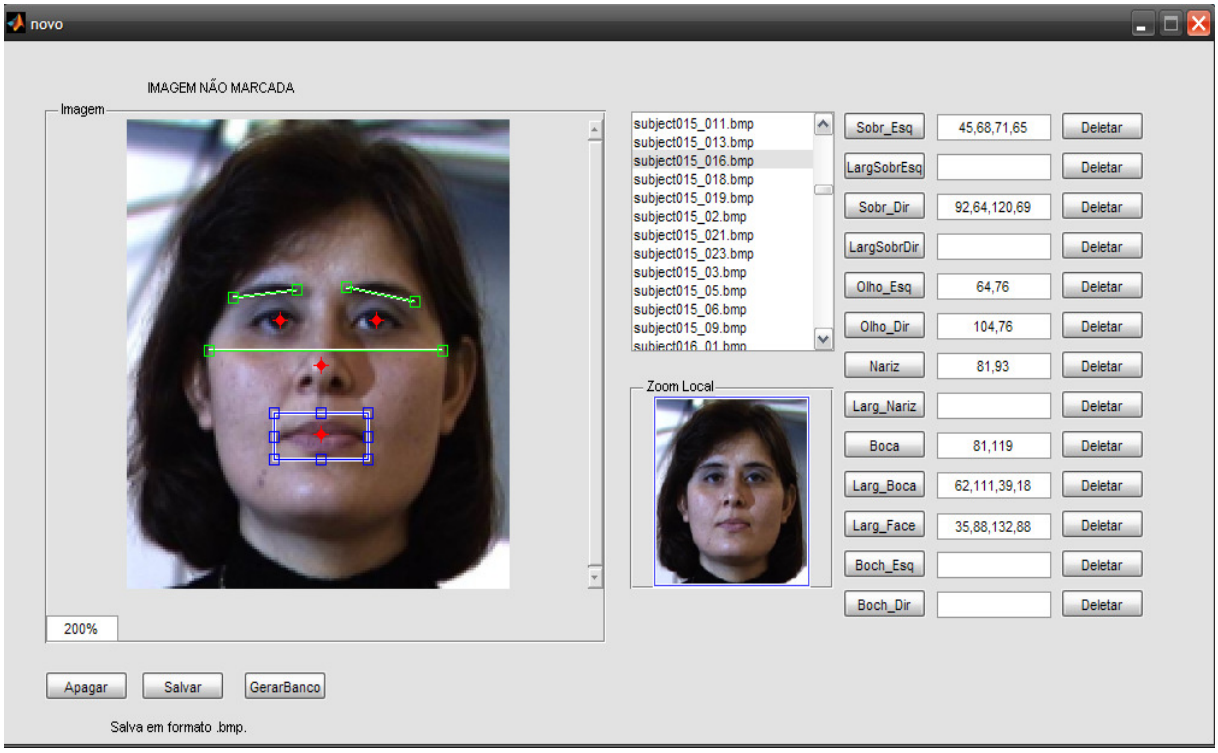


Figura 5.28: Interface para marcação da características geométricas.

A seguir é apresentado o conjunto de características de interesse:

- A = Largura da sobrancelha esquerda;
- B = Largura da sobrancelha direita;
- C = Distância da sobrancelha esquerda ao olho esquerdo;
- D = Distância da sobrancelha direita ao olho direito;
- E = Largura da face;
- F = Largura da bochecha esquerda;
- G = Largura da bochecha direita;
- H = Largura do Nariz;
- I = Largura da boca;
- J = Distância entre olhos;
- K = Distância do nariz ao olho esquerdo;
- L = Distância do nariz ao olho direito;
- M = Distância do nariz à boca;
- N = Distância do olho esquerdo à boca;
- O = Distância do olho direito à boca.

De posse de tais valores o vetor característico foi construído com base em todas as possíveis razões entre as características previamente marcadas, dando origem a um vetor contendo 105 coordenadas, figura 5.29. O uso das razões entre os valores foi aplicado para que o vetor característico apresentasse a propriedade de ser invariante à resolução, visto que as características geométricas já são invariantes à iluminação.

```
V = [A/B A/C A/D A/E A/F A/G A/H A/I A/J A/K A/L A/M A/N A/O ...
B/C B/D B/E B/F B/G B/H B/I B/J B/K B/L B/M B/N B/O ...
C/D C/E C/F C/G C/H C/I C/J C/K C/L C/M C/N C/O ...
D/E D/F D/G D/H D/I D/J D/K D/L D/M D/N D/O ...
E/F E/G E/H E/I E/J E/K E/L E/M E/N E/O ...
F/G F/H F/I F/J F/K F/L F/M F/N F/O ...
G/H G/I G/J G/K G/L G/M G/N G/O ...
H/I H/J H/K H/L H/M H/N H/O ...
I/J I/K I/L I/M I/N I/O ...
J/K J/L J/M J/N J/O ...
K/L K/M K/N K/O ...
L/M L/N L/O ...
M/N M/O ...
N/O]';
```

Figura 5.29: Vetor de características geométricas.

Conforme a Seção 4.1.3, o método Laplacianfaces recebe como entrada um vetor coluna originário de uma imagem que foi rotacionada, recortada, passada para tons de cinza e por fim transformada em um vetor coluna. No método proposto o vetor característico é apresentado como entrada ao Laplacianfaces e a saída, já com dimensão reduzida, é submetida ao classificador Nearest-Neighbor, figura 5.30.

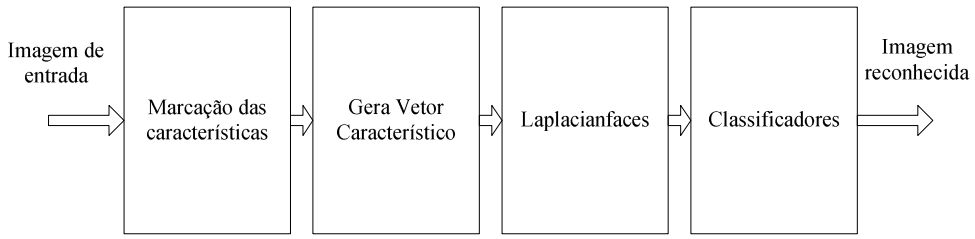


Figura 5.30: Etapas do algoritmo Geométrico – LPP.

6 Experimentos e resultados

Neste capítulo são apresentados os experimentos 4, 5, 6 e 7, cuja finalidade é validar os métodos propostos CT-Eigen, CT-Eigenblocos, Expansão e Geométrico-LPP. Para cada experimento são apresentados os resultados alcançados seguidos das devidas análises. Ao final dos testes é feita uma análise comparativa.

A figura 6.1 apresenta o Mapa de Experimentos dividido nas etapas de Pré-processamento, Transformações e Classificadores.

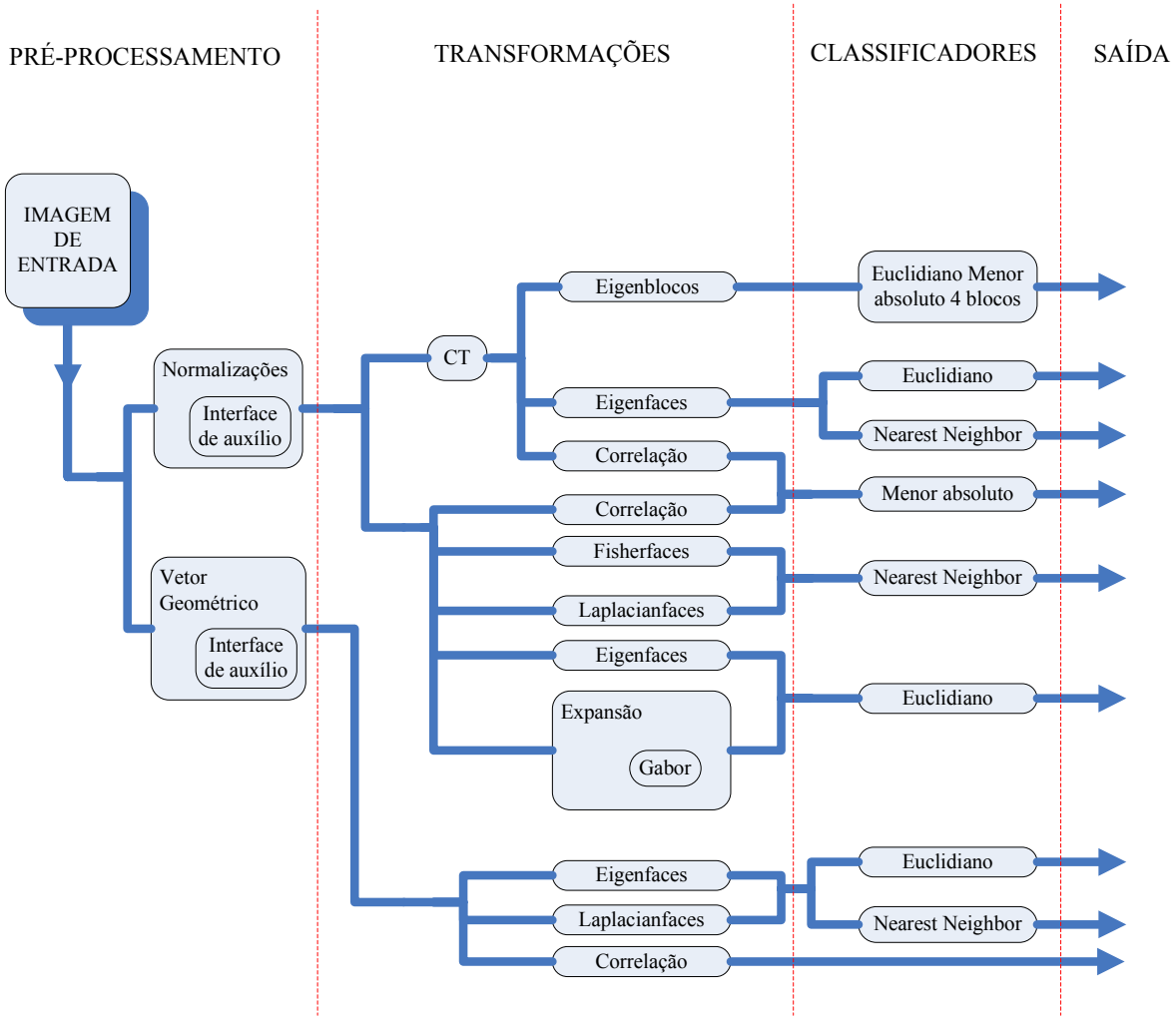


Figura 6.1: Mapa de experimentos.

6.1 Experimentos

Experimento 04

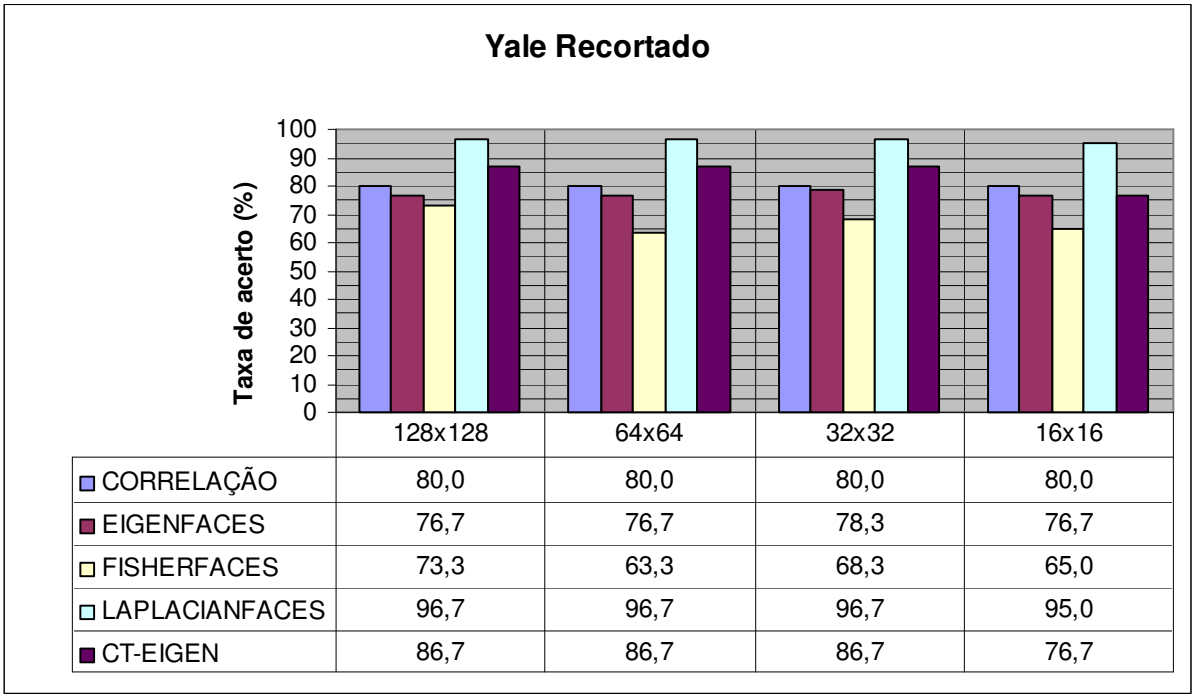
- **Motivo do experimento:** Verificar se o CT-Eigen é mais robusto do que o Eigenfaces baseado na hipótese de que a Transformada Census é capaz de diminuir a não linearidade do especo de faces.

Bases Utilizadas

- Yale, ORL e PIE

Resultados alcançados

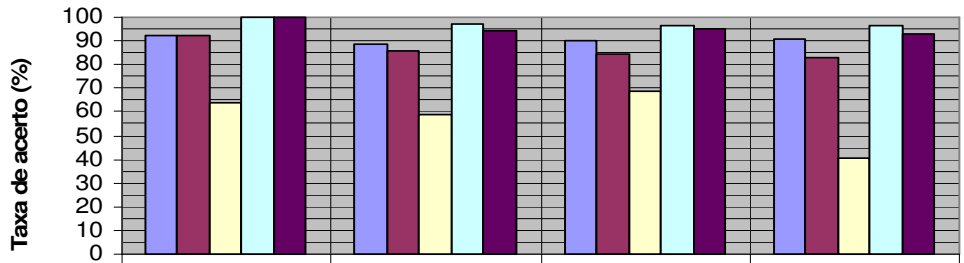
A figura 6.2 apresenta os resultados dos experimentos com o conjunto Yale Recortado tomado em quatro condições de resolução diferentes.



	128x128	64x64	32x32	16x16
CORRELAÇÃO	80,0	80,0	80,0	80,0
EIGENFACES	76,7	76,7	78,3	76,7
FISHERFACES	73,3	63,3	68,3	65,0
LAPLACIANFACES	96,7	96,7	96,7	95,0
CT-EIGEN	86,7	86,7	86,7	76,7

Figura 6.2: Conjunto Yale Recortado tomado em quatro resoluções diferentes.

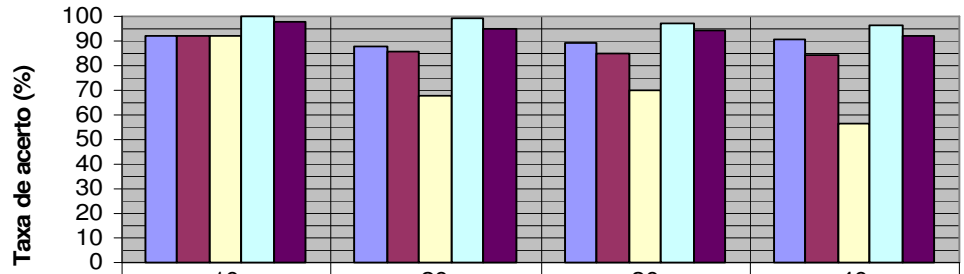
A figura 6.3 apresenta os resultados dos experimentos com o conjunto ORL Recortado tomado na resolução 64x64. Subconjuntos contendo 10, 20, 30 ou 40 indivíduos a serem reconhecidos.



	10	20	30	40
CORRELAÇÃO	92,0	89,0	90,0	90,5
EIGENFACES	92,0	86,0	84,7	83,0
FISHERFACES	64,0	59,0	68,7	40,5
LAPLACIANFACES	100	97,0	96,7	96,5
CT-EIGEN	100	94,0	94,7	93,0

Figura 6.3: Conjunto ORL Recortado dividido nos subconjuntos contendo 10. 20. 30 ou 40 indivíduos a serem reconhecidos.

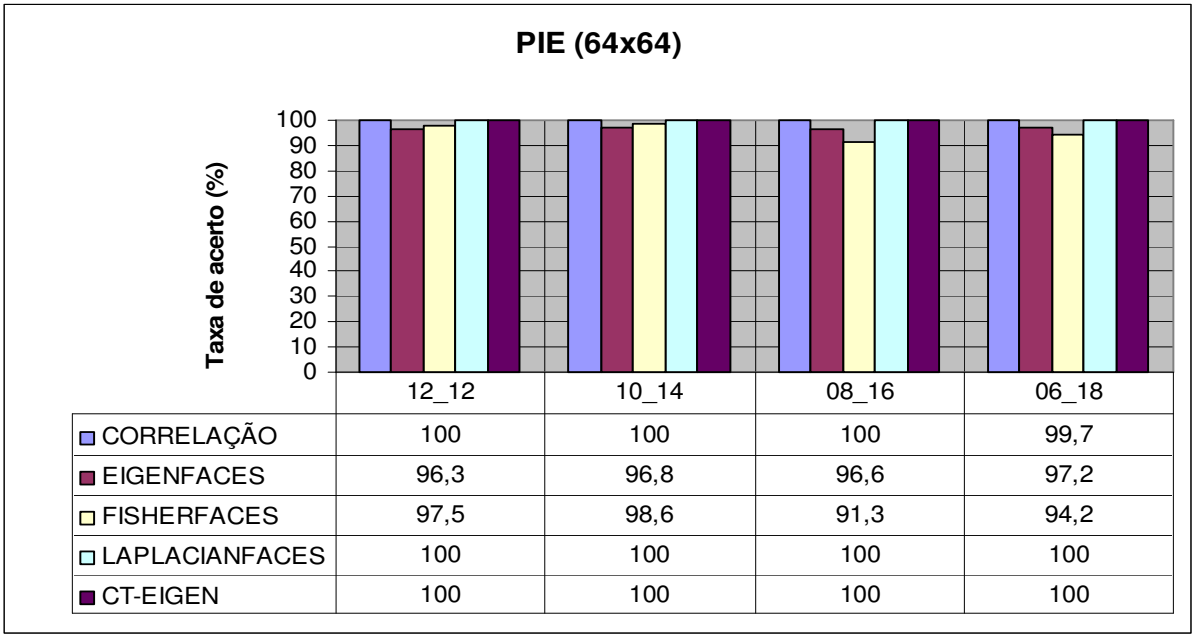
A figura 6.4 apresenta os resultados dos experimentos com o conjunto ORL Recortado tomado na resolução 64x64. Subconjuntos contendo 10, 20, 30 ou 40 indivíduos a serem reconhecidos.



	10	20	30	40
CORRELAÇÃO	92,0	88,0	89,3	90,5
EIGENFACES	92,0	86,0	85,3	84,5
FISHERFACES	92,0	68,0	70,0	56,5
LAPLACIANFACES	100	99,0	97,3	96,5
CT-EIGEN	98,0	95,0	94,0	92,0

Figura 6.4: Conjunto ORL Recortado dividido nos subconjuntos contendo 10. 20. 30 ou 40 indivíduos a serem reconhecidos.

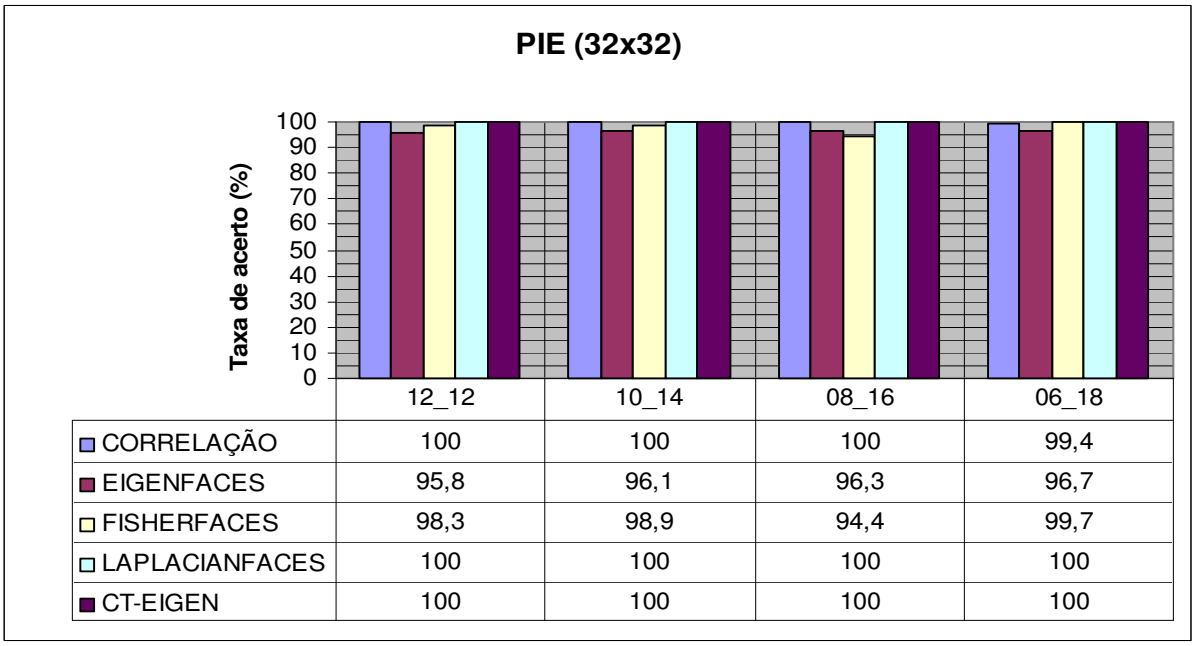
A figura 6.5 apresenta os resultados dos experimentos com o conjunto PIE Recortado tomado na resolução 64x64. Cada coluna da tabela de resultados corresponde a uma divisão entre elementos no conjunto de treinamento e elementos no conjunto de testes.



	12_12	10_14	08_16	06_18
CORRELAÇÃO	100	100	100	99,7
EIGENFACES	96,3	96,8	96,6	97,2
FISHERFACES	97,5	98,6	91,3	94,2
LAPLACIANFACES	100	100	100	100
CT-EIGEN	100	100	100	100

Figura 6.5: Conjunto PIE Recortado dividido em diferentes conjuntos de treinamento e teste.

A figura 6.6 apresenta os resultados dos experimentos com o conjunto PIE Recortado tomado na resolução 32x32. Cada coluna da tabela de resultados corresponde a uma divisão entre elementos no conjunto de treinamento e elementos no conjunto de testes.



	12_12	10_14	08_16	06_18
CORRELAÇÃO	100	100	100	99,4
EIGENFACES	95,8	96,1	96,3	96,7
FISHERFACES	98,3	98,9	94,4	99,7
LAPLACIANFACES	100	100	100	100
CT-EIGEN	100	100	100	100

Figura 6.6: Conjunto PIE Recortado dividido em diferentes conjuntos de treinamento e teste.

Análise dos resultados

- CT-Eigen foi mais robusto do que o Eigenfaces nas condições de:
  - Variação de luminosidade;
  - Crescimento no número de indivíduos a serem reconhecidos;
  - Variação no número de indivíduos nos conjuntos de treinamento e teste;
  - Variação na resolução das imagens.
- Em alguns casos o CT-Eigen foi similar ao Laplacianfaces.

Experimento 05

- **Motivo do experimento:** Verificar se o método CT-Eigenblocos é de fato mais robusto do que o Eigenfaces, no quesito imagens com partes da face parcialmente encobertas.

Aqui apresentamos o desempenho dos métodos quando a imagem a ser reconhecida está parcialmente encoberta. Para realização do experimento, cada imagem dos conjuntos de treinamento foi dividida em quadrantes, figura 6.7, e para cada ciclo de experimentos um dos quadrantes foi pintado de preto. O quadrante em preto modela a parte encoberta da face.

(Original) (Q1) (Q2) (Q3) (Q4)

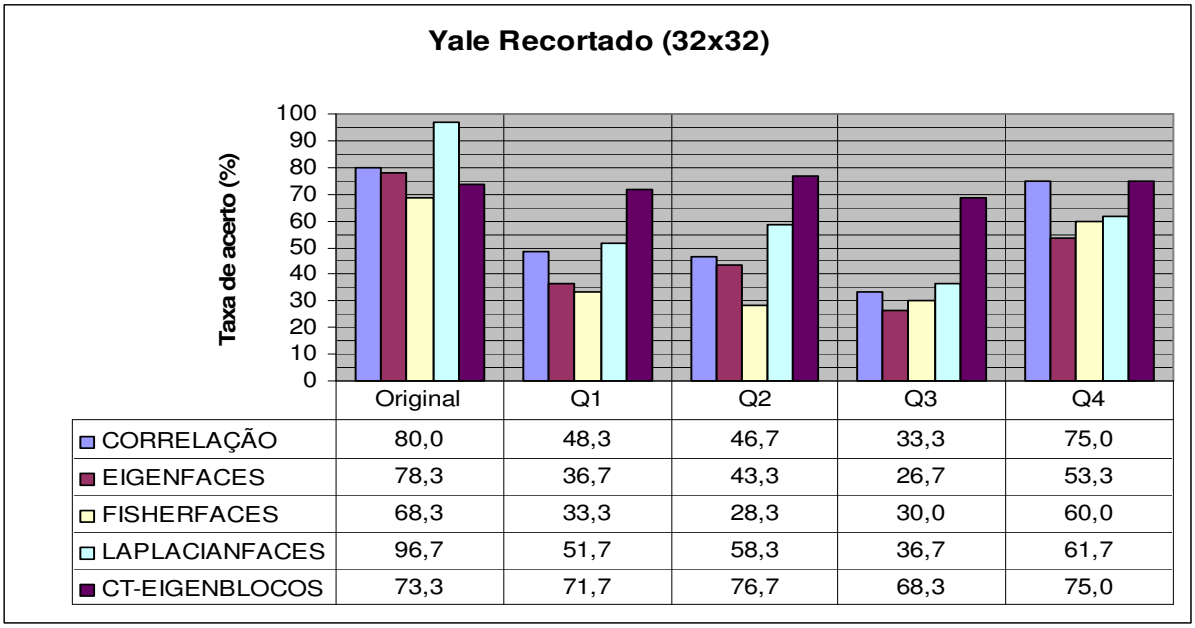
Figura 6.7: Imagem original e seus quadrantes encobertos.

Bases Utilizadas

- Yale, ORL e PIE

Resultados alcançados

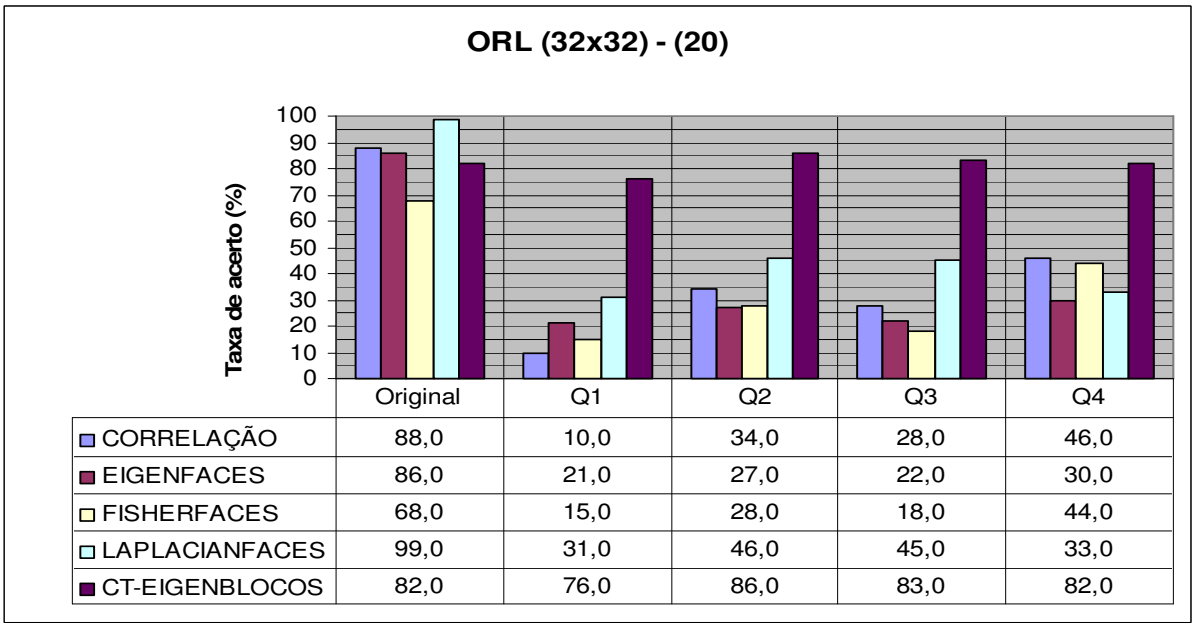
A figura 6.8 apresenta os resultados dos experimentos com o conjunto Yale Recortado. A coluna "Original" corresponde ao caso em que as imagens não apresentam parte encoberta; já as colunas Q1, Q2, Q3 e Q4 indicam o quadrante que se encontrava encoberto.



	Original	Q1	Q2	Q3	Q4
CORRELAÇÃO	80,0	48,3	46,7	33,3	75,0
EIGENFACES	78,3	36,7	43,3	26,7	53,3
FISHERFACES	68,3	33,3	28,3	30,0	60,0
LAPLACIANFACES	96,7	51,7	58,3	36,7	61,7
CT-EIGENBLOCOS	73,3	71,7	76,7	68,3	75,0

Figura 6.8: Conjunto Yale Recortado. Experimento considerando quadrantes encobertos.

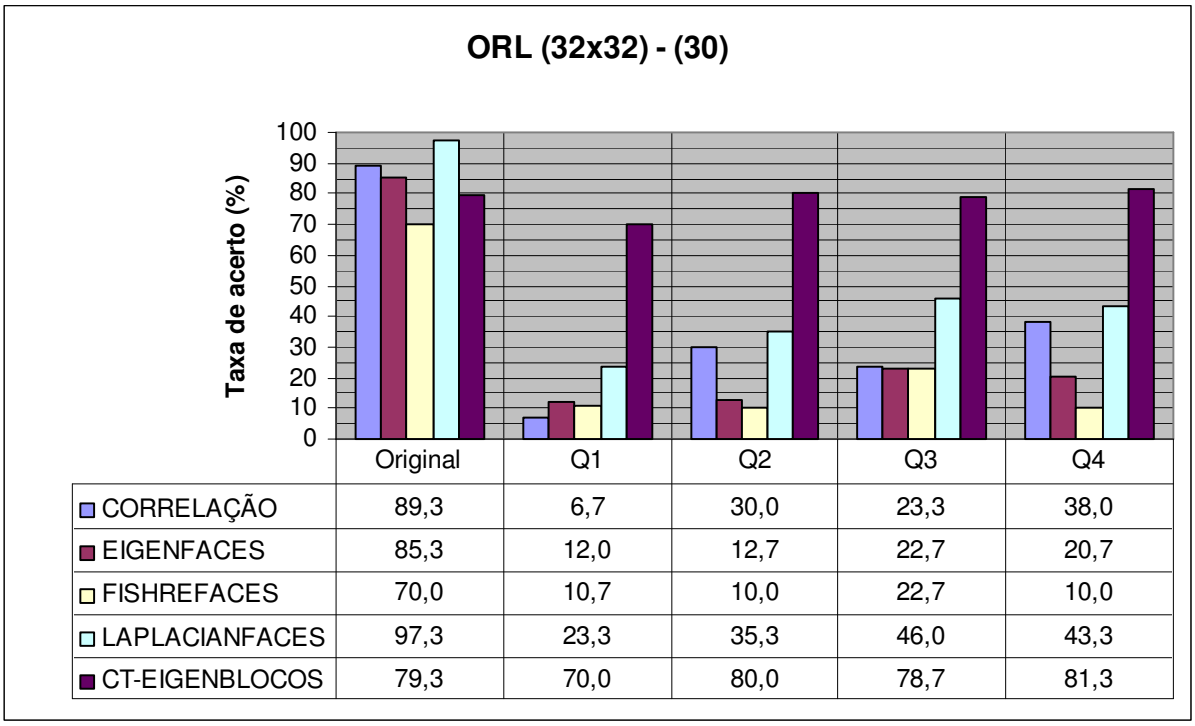
A figura 6.9 apresenta os resultados dos experimentos com o conjunto ORL Recortado.



	Original	Q1	Q2	Q3	Q4
CORRELAÇÃO	88,0	10,0	34,0	28,0	46,0
EIGENFACES	86,0	21,0	27,0	22,0	30,0
FISHERFACES	68,0	15,0	28,0	18,0	44,0
LAPLACIANFACES	99,0	31,0	46,0	45,0	33,0
CT-EIGENBLOCOS	82,0	76,0	86,0	83,0	82,0

Figura 6.9: Subconjunto ORL Recortado, contendo 20 indivíduos a serem reconhecidos. Experimento considerando quadrantes encobertos.

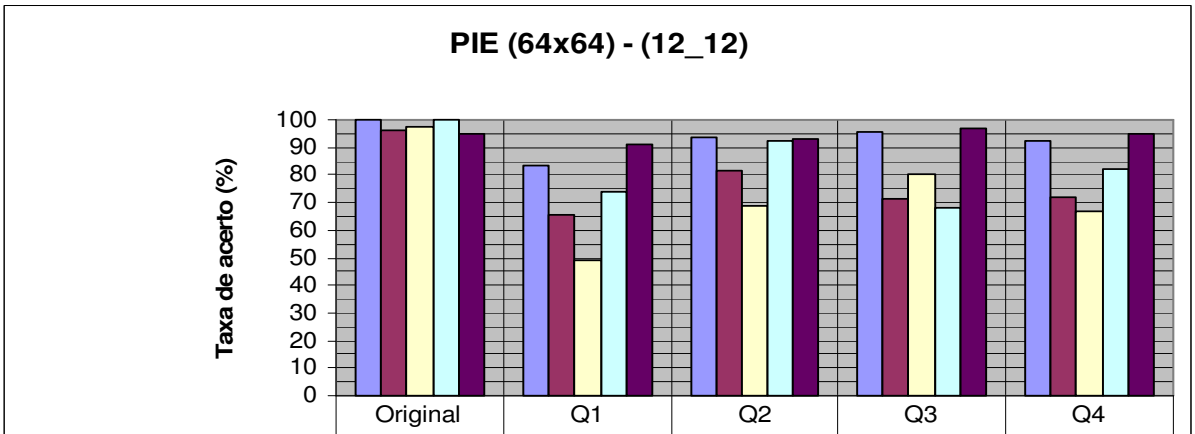
A figura 6.10 apresenta os resultados dos experimentos com o conjunto ORL Recortado.



	Original	Q1	Q2	Q3	Q4
CORRELAÇÃO	89,3	6,7	30,0	23,3	38,0
EIGENFACES	85,3	12,0	12,7	22,7	20,7
FISHREFACES	70,0	10,7	10,0	22,7	10,0
LAPLACIANFACES	97,3	23,3	35,3	46,0	43,3
CT-EIGENBLOCOS	79,3	70,0	80,0	78,7	81,3

Figura 6.10: Subconjunto ORL Recortado, contendo 30 indivíduos a serem reconhecidos. Experimento considerando quadrantes encobertos.

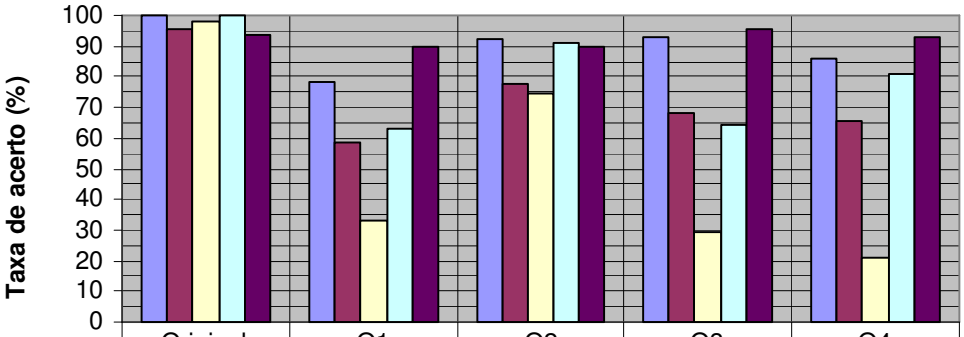
A figura 6.11 apresenta os resultados dos experimentos com o conjunto PIE Recortado tomado na resolução 64x64 e conjunto de treinamento e teste (12,12).



	Original	Q1	Q2	Q3	Q4
CORRELAÇÃO	100	83,7	93,7	95,8	92,5
EIGENFACES	96,3	65,4	81,7	71,2	72,0
FISHERFACES	97,5	48,8	68,8	80,0	67,1
LAPLACIANFACES	100	74,2	92,1	67,9	82,1
CT-EIGENBLOCOS	94,6	91,3	92,9	96,7	94,6

Figura 6.11: Conjunto PIE Recortado, conjunto de treinamento e teste (12,12). Experimento considerando quadrantes encobertos.

A figura 6.12 apresenta os resultados dos experimentos com o conjunto PIE Recortado tomado na resolução 32x32 e conjunto de treinamento e teste (12,12).



	Original	Q1	Q2	Q3	Q4
CORRELAÇÃO	100	78,3	92,1	92,9	85,8
EIGENFACES	95,8	58,8	77,9	67,9	65,8
FISHERFACES	98,3	32,9	74,6	29,6	21,3
LAPLACIANFACES	100	62,9	90,8	64,6	80,8
CT-EIGENBLOCOS	93,8	90,0	90,0	95,4	93,3

Figura 6.12: Conjunto PIE Recortado, conjunto de treinamento e teste (12,12). Experimento considerando quadrantes encobertos.

Análise dos resultados

- Os métodos Correlação, Eigenfaces, Fisherfaces, Laplacianfaces e Expansão apresentaram queda brusca no percentual de reconhecimento para o caso de faces parcialmente encobertas. Segundo (ZHAO, CHELLAPPA et al., 2003) para que eles reconheçam corretamente é necessário que a imagem da face esteja completa ou que a parte encoberta seja pequena.
- O CT-Eigenblocos apresentou apenas uma pequena queda no percentual de reconhecimento, comprovando ser o mais robusto para o caso de imagens com faces parcialmente encobertas, superando até mesmo o método Laplacianfaces.

Experimento 06

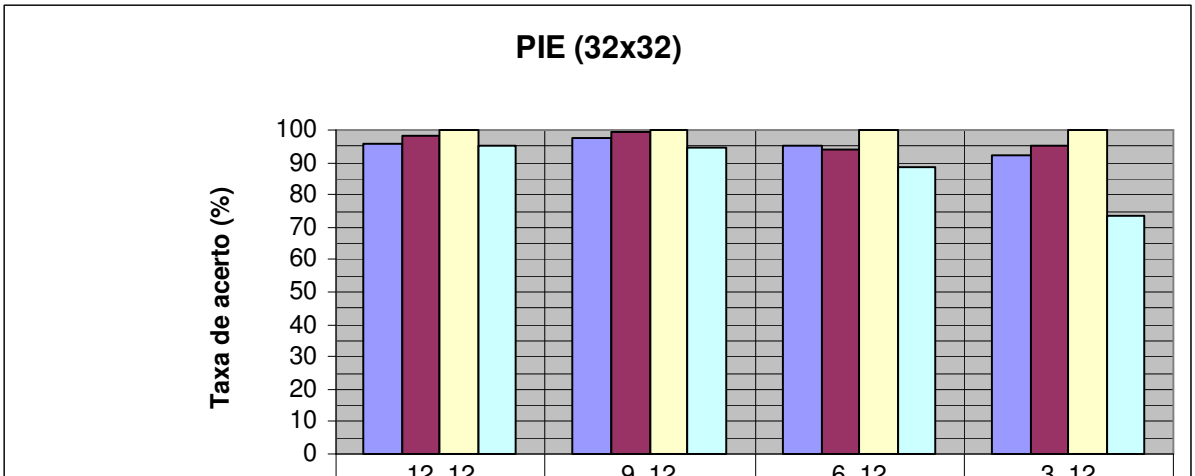
- **Motivo do experimento:** Verificar se o método **Expansão** é mais robusto que o **Eigenfaces** para conjunto de treinamento com número reduzido de amostras por indivíduo.

Base Utilizada

- PIE

Resultados alcançados

A figura 6.13 apresenta os resultados dos experimentos com o conjunto PIE Recortado tomado na resolução 32x32. Cada uma das colunas da tabela de resultados representa uma composição diferente dos conjuntos de treinamento e teste.



	12_12	9_12	6_12	3_12
EIGENFACES	95,8	97,5	95,4	92,1
FISHERFACES	98,3	99,6	93,8	95,4
LAPLACIANFACES	100	100	100	100
EXPANSÃO	95	94,58	88,3	73,5

Figura 6.13: Conjunto PIE recortado, dividido em diferentes grupos de treinamento e teste. Experimento considerando quadrantes encobertos.

Análise dos resultados

- Foi comparável em alguns casos, mas degenerou com a redução do número de amostras no conjunto de treinamento;
- Não atingiu as expectativas;
- Precisa ser mais explorado dado a variação da família de Gabor Wavelets, suas freqüências, orientações e resoluções.

Experimento 07

- **Motivo do experimento:** Testar a combinação entre os métodos Características Geométricas e o Laplacianfaces. Tal combinação é chamada de Geométrico-LPP.

Base Utilizada

- ORL

Resultados alcançados

A figura 6.14 apresenta os resultados dos experimentos com o conjunto ORL Recortado resolução 32x32.

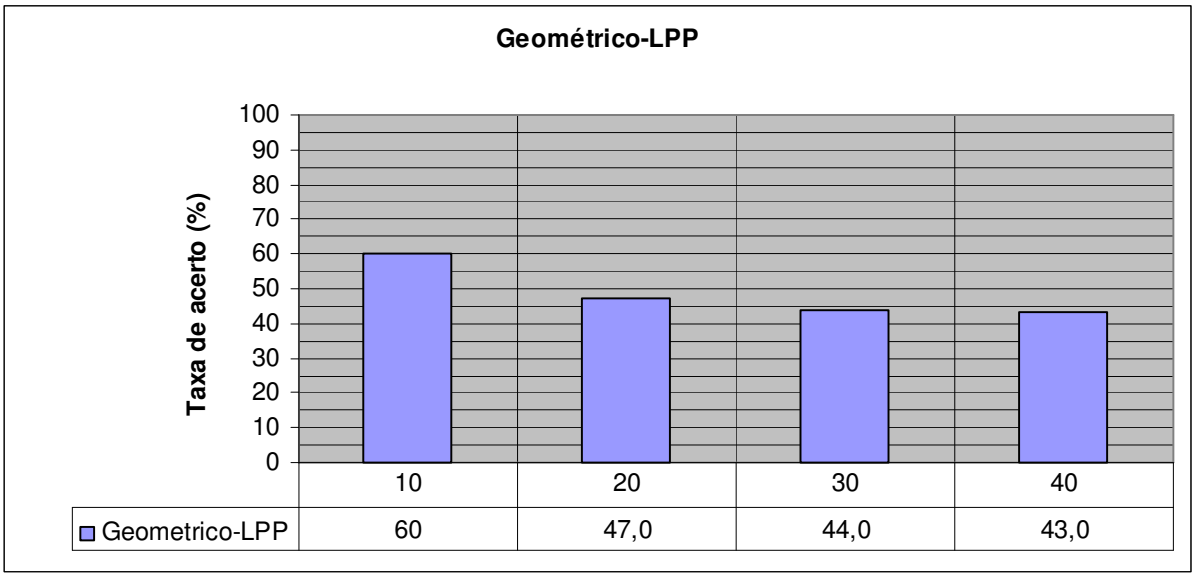


Figura 6.14: Conjunto ORL Recortado, considerando conjuntos contendo 10, 20, 30 ou 40 indivíduos a serem reconhecidos.

Análise dos resultados

- O método em questão não apresentou resultados satisfatórios.
  - **Possíveis causas**
    1. O número de características selecionadas pode ter sido insuficiente;
    2. A quantidade de imagens por indivíduo no conjunto de treinamento pode ter sido insuficiente para o vetor de características construído;
    3. Características puramente geométricas não agregarem, por si só, informação suficiente.

6.2 Análise comparativa

O CT-Eigen mostrou ser mais robusto que o Eigenfaces frente a conjuntos em que as imagens apresentam variação de iluminação. Tal fato comprova a hipótese de que a Tranformada Census é capaz de diminuir a não linearidade do espaço de faces.

Para testes com imagens de faces parcialmente encobertas o método CT-Eigenblocos mostrou ser mais robustos que os métodos tomados como referência. Estes por sua vez apresentaram queda brusca na taxa de reconhecimento;

O método Expansão foi comparável em alguns casos, mas degenerou com a redução do número de amostras no conjunto de treinamento. Logo NÃO atingiu as expectativas;

O método Geométrico-LPP NÃO atingiu resultados satisfatórios (taxa de acerto inferior à 50%), para o conjunto de características selecionadas.

7 Conclusões e Trabalhos futuros

Esta dissertação apresentou um estudo comparativo entre métodos de reconhecimento facial do tipo holístico e também propôs e avaliou as abordagens CT-Eigen, CT-Eigenblocos, Expansão e Geométrico-LPP, dentre as quais CT-Eigen e CT-Eigenblocos atigiram os resultados esperados.

Cabe ressaltar que o desempenho dos sistemas de reconhecimento está associado a todas as etapas, ou seja, normalização, transformações de dimensionalidade e classificadores. É muito comum encontrar trabalhos comparativos submetendo a saída das transformações a apenas um classificador, normalmente o euclidiano por critério de simplicidade e afirmando que um determinado método é superior ao outro. Como contra-exemplo a saída do método CT-Eigen foi aplicada aos classificadores Euclidiano e Nearest-Neighbor, e mesmo o euclidiano sendo considerado mais simples e menos robusto, para este experimento ele gerou os melhores resultados. Portanto, o desempenho final dos sistemas de reconhecimento facial também é influenciado pelo classificador aplicado e não somente pelo pré-processamento e transformações da imagem.

Por fim esta dissertação também cumpriu com o objetivo de estender a pesquisa de (MOUTINHO, 2005), o qual tratou sobre a localização facial e propôs como trabalho futuro, o estudo de técnicas de reconhecimento facial.

7.1 Trabalhos Futuros

Como sugestão de trabalho futuro pode ser realizada uma análise de desempenho da técnica de Expansão dado a variação da família de Gabor Wavelets utilizada, suas freqüências, orientações e resoluções.

Outro trabalho futuro seria testar qual é o número mínimo de características geométricas necessárias para que tal abordagem Geométrico-LPP atinja resultados ao menos comparáveis com as técnicas clássicas de PCA e LDA. Verificar também se existe uma relação inversamente proporcional entre o comprimento do vetor de características geométricas e o número de amostras por indivíduo no conjunto de treinamento.

Uma terceira sugestão é o desenvolvimento de um módulo capaz de identificar se a imagem da face a ser reconhecida está completa ou parcialmente encoberta. No caso da face estar parcialmente encoberta, a imagem seria enviada para um módulo auxiliar de reconhecimento, treinado com o método CT-Eigenblocos, caso contrário a imagem seria enviada diretamente a um módulo treinado com o método Laplacianfaces.

Por fim um próximo passo de extrema importância na busca de máquinas inteligentes é o desenvolvimento de módulos capazes de reconhecer expressões faciais (sorrindo, piscando, triste, feliz, normal e cansado) e estimar gênero, raça e idade. Por exemplo, o reconhecimento de expressões faciais pode ser aplicado em empresas em que os funcionários trabalham em situação de forte estresse, pois seria possível traçar um mapa dinâmico correlacionando o desempenho da equipe e o nível de estresse que ela apresentava no ponto analisado.

8 REFERÊNCIAS BIBLIOGRÁFICAS

ANTON, H.; RORRES, C. **Elementary Linear Algebra**, 9th ed., New York: John Wiley & Sons, 2005. 624p.

BELHUMEUR, P. N.; HESPANHA, J. P.; KRIEGMAN, D.J. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, **IEEE Trans. Pattern Analysis and Machine Intelligence**, v.19, n.7, pp. 711-720, Julho 1997.

BERTILLON, A. La couleur de l'iris, Rec. Sci., v.36, n.3, pp. 65-73, 1885.

BLEDSOE,W. **Man-Machine Facial Recognition** [R]. Technical Report PRI 22, Panoramic Research Inc, Palo Alto, CA, 1966.

BOYD, S. P.; VANDENBERGHE, L. **Convex optimization**. New York: Cambridge University Press, 2004.

BRUNELLI, R.; POGGIO, T. **Face Recognition: Features vs. Templates**, *IEEE Trans. On Pattern Analysis and Machine Intelligence,* v.15, n.10, pp. 1042-1052, Oct. 1993.

COHEN, A.; KOVACEVIC, J. "**Wavelets: the mathematical background**," *Proc. IEEE*, vol. 84, no. 4, pp. 514–522, 1996.

CHANG, Y.; HU, C.; TURK, M. **Manifold of Facial Expression**. In: Proc. IEEE Int'l Workshop Analysis and Modeling of Faces and Gestures, Outubro 2003.

CHEN, J.; YUAN, B. **Face Manifold Analysis Based on LFA Features**. In: *Congress on Image and Signal Processing*, Maio. 2008, pp. 580-583.

DAUGMAN, J. G. High Confidence Visual Recognition of Person by a Test os Statistical Independence, **IEEE Trans. Pattern Analysis and Machine Intelligence**, v.15, n.11, pp.1148-1161, 1993.

DUDA, R. O.; Hart, P. E.; Stork, D. G. **Pattern Classification**. New York : Wiley Interscience, 2001.

EHSAN, F. E.; TSOTSOS, J. K.. **Local Feature Analysis for Robust Face Recognition**. In: Proceedings IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA 2009).

ERSI, E. F.; HAJEBI, K. **Face Recognition by fiducial point analysis**. In: Canadia Conference on ELectrical and Computer Engineering. Toward Caring and Humane Technology (Cat n.03ch37436), 2003, pp. 1187-1190.

FRANZINI, M. A.; WITBROCK, M. J.; LEE, K. F. **Speaker-Independent Recognition of Connected Utterances Using Recurrent and Non-recurrent Neural Networks**. In: IEEE International Joint Conference on Neural Networks (3rd IJCNN'89), 1989.

GONZALEZ, R.; WOODS, R. **Digital Image Processing**, São Paulo: E. Blücher, 2000. 509p.

HE, X.; NIYOGI, P. **Locality Preserving Projections**, Advances in Neural Information Processing Systems, 2003.

HOFFMANN, G. **CieLab Color Space**. Disponível em <http://www.fho emden.de/~hoffmann/cielab03022003.pdf>. Acesso em 03/02/2011.

JAIN, A. K.; ROSS, A.; PRABHAKAR, S. An Introduction to Biometric Recognition**. IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image-and Video-Based Biometrics**, v.14, n.1, pp. 4-20 Jan. 2004.

KANADE, T. **Picture Processing System by Computer Complex and Recognition of Human Faces**, doctoral dissertation, Kyoto University, 1973.

KAYA, Y.; KOBAYASHI, K. **A basic study on human face recognition**, in *Frontiers of Pattern Recognition* (S. Watanabe, Ed.). 1972, p. 265.

KIRBY, M.; SIROVICH, **L. Application of the Karhunen- Loeve procedure for the characterization of human faces**. IEEE Trans. Pattern Analysis and Machine Intelligence, v.12, n.1, pp. 103-108, Jan.1990.

LIU, C.; WECHSLER, H. **Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model For Face Recognition**. IEEE *Trans. Image Processing,* v.11, n.4, pp. 467-476, 2002.

MAATEN, L. V.; POSTMA, E. **Dimensionality Reduction: A Comparative Review**. Disponível em <http://www.uvt.nl/ticc>. Acesso em 04/05/2010.

MARQUES, A. C. P. B. **Extração de Minúcias em Imagem de Impressões Digitais utilizando Redes Neurais**. Disponível em < http://teses2.ufrj.br/Teses/NCE_M/AiramCarlosPaisBarretoMarques.pdf >. Acesso em 13/07/2009.

MARTINEZ, A. M.; KAK, A. **PCA versus LDA**. IEEE Transactions on Pattern Analysis and Machine Intelligence , 2001.

MOGHADDAM, B.; PENTLAND, A. P. **Probabilistic Visual Learning for Object Representation**, IEEE Trans. Pattern Analysis and Machine Intelligence, v. 9, pp. 696- 710, 1997.

MONTGOMERY, D. C.; GEORGE, C. R. **Applied statistics and probability for engineers**, 5th ed., New York: John Wiley & Sons, 2003. 706p.

MOUTINHO, A. M. **Identificação de padrões faciais usando redes neurais artificiais**. Dissertação (Mestrado em informática) – UFRJ/IM/NCE, Rio de Janeiro, 214f, 2005. Disponível em < http://teses.ufrj.br/NCE_M/AdrianoMartinsMoutinho.pdf>. Acesso em 04/03/2010.

MOSES, Y.; ADINI, Y.; ULLMAN, S. “Face Recognition: **The Problem of Compensating for Changes in Illumination Direction**,” European Conf. Com. Vision, 1994, pp. 286-296.

PENEV, P. S.; ATICK, J. J. **Local Feature Analysis: a general statistical theory for object representation** *Network: Comput. Neural Syst,* v.7, n.3, pp. 471- 500, 1996.

QIAN, Z. M.; SU, P.Y.; XU, D. **Face Recognition Based on Local Feature Analysis**. In: *2008 International Symposium on Computer Science and Computational Technology*, Dezembro 2008, pp. 264-267.

RABINER, L. R.; **A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition**, Proceedings of the IEEE, v.77, n.2, pp. 257-286, 1989.

ROWEIS, S. T.; SAUL, L. K. **Nonlinear Dimensionality Reduction by Locally Linear Embedding**, Science, 2000, v.290, pp. 2323-2326.

ROWEIS, S.; SAUL, L.; HINTON, G. **Global Coordination of Local Linear Models**. In: Proc. Conf. Advances in Neural Information Processing System 14, 2001.

RUDASI, L. **Text-independent automatic speaker identification using partitioned neural networks**, Tese de Ph.D., Old Dominion University, 1999.

SCHMIDT, M.; GOLDEN, J.; **GISH, H. GMM Sample Statistic Log-Likelihoods for Text-Independent Speaker Recognition**. In: Proc. Eurospeech'97, pp. 855-858, Rhodes, Greece, 1997.

SEUNG, H. S.; LEE, D. D. **The Manifold Ways of Perception**, Science, v.290, Dezembro 2000.

SHAHIN, I.; BOTROS, N. **Speaker identification using dynamic time warping with stress compensation technique**. In: Proceedings of Southeastcon’98, p.65-68, IEEE, 1998.

SIM, T.; BAKER, S. et al. **The CMU Pose, Illumination, and Expression (PIE) Database**. In: Proc. IEEE Int’l Conf. Automatic Face and Gesture Recognition, Maio 2002.

TAPPERT, C. C.; WAKAHARA, C. Y. S. The state of art in on-line handwriting recognition, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, pp.785-807, 1990.

TENENBAUM, J.B.; LANGFORD, J.C. **A global geometric framework for nonlinear dimensionality reduction**, Science, 2000, v.290, pp. 2319-2323.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**, 4th ed., Burlington: Elsevier, 2009. 961p.

TURK, M.; PENTLAND, A. **Face Recognition Using Eigenfaces**. In: *IEEE Conf. on Computer Vision and Pattern Recognition*, 1991. *Proc.* pp. 586-591.

WISKOTT, L.; FELLOUS, J. M.; KRUGER, N. Face Recognition by Elastic Bunch Graph Matching, **IEEE Trans. Pattern Analysis and Machine Intelligence**, v.19, p. 775-779, 1997.

XIAOFEI, H.; SHUICHENG, Y.; YUXIAO, H.; NIYOGI, P. et al. **Face Recognition Using Laplacianfaces**, IEEE Trans. Pattern Anal. and Mach. Intell., v.27, n.3, pp. 328-340, Mar.2005.

YANG, M. H. **Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods**. In: Proc. 5th Int'l Conf. Automatic face and Gesture Recognition, Maio 2002.

ZABIH, R.; WOODFILL, J. **A non-parametric approach to visual correspondence**. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996.

ZHAO, W.; CHELLAPPA, R. et al. Face recognition: A literature survey, **ACM Computing Surveys**, 2003, 35(4), 399-458.